

DISPONIBILIDAD LÉXICA
EN TRES CENTROS DE INTERÉS RELACIONALES
EN ESTUDIANTES NATIVOS DE ESPAÑOL Y
SINOHABLANTES DE ELE



**UNIVERSITAT
ROVIRA I VIRGILI**

Máster de Enseñanza de Lenguas. Español Lengua Extranjera.

Trabajo fin de máster

Autor: Blanca Corral Olivares

Tutor: José Antonio Moreno Villanueva

Fecha de entrega: 9 de junio



**Máster en Enseñanza de Lenguas Extranjeras:
Español como Lengua Extranjera**

Fecha: 06/06/2021

Trabajo de Fin de Máster

DECLARACIÓN JURADA DE AUTORÍA DE TRABAJO ACADÉMICO

El alumno / La alumna Blanca Corral Olivares con D. I.04239773-E

Hace constar que es el autor /la autora del trabajo:

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE.

Y manifiesta su responsabilidad en la realización del mismo, en la interpretación de datos y en la elaboración de conclusiones. Manifiesta asimismo que las aportaciones intelectuales de otros autores utilizados en el texto se han citado debidamente.

En este sentido,

DECLARA:

Que el trabajo remitido es un documento original y no ha sido publicado en español o en otro idioma con anterioridad, total o parcialmente, por otros autores.

Que el abajo firmante es públicamente responsable de sus contenidos y elaboración, y que no ha incurrido en fraude científico o plagio.

Que si se demostrara lo contrario, el abajo firmante aceptará las medidas disciplinarias o sancionadoras que correspondan.

Fdo.: _____

Resumen

La disponibilidad léxica nació en la década de 1950. Tras aplicarse en el ámbito de la sociolingüística y la psicolingüística, en los últimos años han surgido numerosos trabajos que exploran su aplicación en el ámbito del español como lengua extranjera. Este trabajo se enmarca en esta línea y sigue la metodología habitual en los estudios sobre el léxico disponible. Con el objetivo de profundizar en la evaluación de la competencia léxica de los estudiantes, se trabaja con tres centros de interés de carácter relacional (Salud, Relaciones familiares y sociales e Informática y nuevas tecnologías) a partir de una muestra de 40 alumnos universitarios, 20 nativos y 20 estudiantes de ELE sinohablantes de nivel B1-B2. Los datos obtenidos se analizan tanto cuantitativa como cualitativamente, atendiendo no solo al número de palabras o vocablos, sino también al análisis de las palabras más disponibles, su clasificación gramatical o su tratamiento en el PCIC.

Palabras clave

Disponibilidad léxica; estudiantes sinohablantes de ELE; salud; relaciones familiares y sociales; informática y nuevas tecnologías

Abstract:

Lexical availability was born in the 1950s. After being applied in the field of sociolinguistics and psycholinguistics, in recent years there have been numerous works that explore this application in the field of Spanish as a foreign language. This work is part of this line and follows the usual methodology in studies on the lexical available. With the aim of deepening the evaluation of students lexical competence, we work with three centers of relational interest (Health, Family and social relations and Informatics and new technologies) from a sample of 40 university students, 20 native speakers and 20 ELE students with B1- B2 level. The data obtained are analyzed quantitatively and qualitatively, taking into account not only the number of words, but also the analysis of the most available words, their grammatical classification or their treatment in the PCIC.

Keywords:

Lexical availability; health; family and social relationships; Informatics and new technologies

ÍNDICE

1. Introducción	8
1.1. Preguntas de investigación.	9
1.2. Hipótesis.	9
1.3. Objetivos	9
1.4. Metodología.	10
1.5. Estructura del trabajo.	10
2. Marco teórico	14
2.1 Competencia léxica	14
2.1.1. La competencia léxica en el MCER y PCIC	
2.1.2. Tratamiento del léxico según los distintos enfoques metodológicos.	
2.1.2.1. Método comunicativo	
2.1.2.2. Enseñanza de la lengua mediante tareas.	
2.1.2.3. Enfoque léxico.	
2.2 Selección léxica	22
2.2.1. La necesidad de una selección léxica.	
2.3. Disponibilidad léxica	24
2.3.1. Estudios de disponibilidad léxica en ELE.	
2.3.2. Centro de interés.	
2.3.3. Factores cognitivos de la disponibilidad léxica.	
2.4. La disponibilidad léxica como instrumento de evaluación.	29
2.5. Estado de la cuestión.	31
2.5.1. Sociolingüística.	
2.5.2. Psicolingüística.	
2.5.3. Dialectología	
2.5.4. Etnolingüística	
2.5.5. Didáctica de las lenguas.	
3. Encuesta de disponibilidad léxica en los ámbitos de Salud, Relaciones familiares y sociales e Informática y nuevas tecnologías	36
3.1. Perfil de los encuestados	37
3.2. Administración de la encuesta y tratamiento de datos.	38

4. Análisis de los datos	43
4.1 Análisis cuantitativo	43
4.2 Análisis cualitativo	50
4.2.1 Unidades léxicas con un índice superior a 0,1	
4.2.1.1 Salud	
4.2.1.2 Relaciones familiares y sociales	
4.2.1.3 Internet y nuevas tecnologías	
4.2.2 Análisis de las categorías gramaticales	
4.2.3 Comparativa y evaluación de las unidades léxicas con el PCIC	
5. Conclusiones	63
6. Bibliografía	67
7. Anexos.	71
Anexo 1. Documento principal para la explicación de encuestas	71
Anexo 2. Índices de disponibilidad completos	72

Relación de tablas

Tabla 1. Gradación de la riqueza del vocabulario según el MCER.
Tabla 2. Gradación del dominio del vocabulario según el MCER.
Tabla 3. Centros de interés del Proyecto Panhispánico.
Tabla 4. Información general de los encuestados.
Tabla 5. Ejemplo del formato para la edición de respuesta para el centro de interés Salud
Tabla 6. Exposición de los dígitos utilizados con <i>Dispogen</i>
Tabla 7. Resultados generales de los informantes nativos
Tabla 8. Resultados generales de los informantes no nativos
Tabla 9. Palabras del centro de interés Salud con un índice igual o superior a 0,1 en ambos casos.
Tabla 10. Palabras del centro de interés Relaciones con un índice igual o superior a 0,1 en ambos casos.

Tabla 11. Palabras del centro de interés Informática con un índice igual o superior a 0,1 en ambos casos.

Tabla 12. Análisis de las categorías gramaticales

Tabla 13. Distribución de UL utilizadas por al menos 3 informantes del grupo de los no nativos según los niveles de referencia del MCER.

Tabla 14. Índice de disponibilidad léxica centro de interés *Salud*, grupo nativo

Tabla 15. Índice de disponibilidad léxica centro de interés *Relaciones*, grupo nativo

Tabla 16. Índice de disponibilidad léxica centro de interés *Informática*, grupo nativo

Tabla 17. Índice de disponibilidad léxica centro de interés *Salud*, grupo no nativo

Tabla 18. Índice de disponibilidad léxica centro de interés *Relaciones*, grupo no nativo

Tabla 19. Índice de disponibilidad léxica centro de interés *Informática*, grupo no nativo

Relación de gráficas

Gráfico 1. Diferencias en el número de palabras por centro de interés

Gráfico 2. Diferencia en el número de vocablos por centro de interés

Gráfica 3. Promedio de palabras por informante

Gráfica 4. Índice de cohesión por centro de interés

1. Introducción

No han sido pocos los intentos para establecer el léxico que emplea una determinada comunidad de habla. Uno de los primeros lo encontramos en la década de 1950, cuando un grupo de investigadores franceses trataron de determinar cuál era el léxico fundamental de la lengua francesa, todo ello con una finalidad pedagógica, pues se preguntaban qué léxico debían incorporar al currículo para la enseñanza del francés en las colonias. Una de las lenguas que mejor han aprovechado este campo es el español, tanto en lengua materna como en ELE, gracias al impulso de Humberto López Morales, quien impulsó el desarrollo del *Proyecto Panhispánico de Disponibilidad Léxica* (PPHDL) con el propósito de establecer el léxico disponible general del mundo hispánico. Estos estudios cuentan con un amplio abanico de posibilidades de aplicación, y a pesar de que los principales trabajos se hayan enfocado al estudio del léxico en comunidades de hablantes nativos, en las últimas décadas se han llevado a cabo estudios para determinar o perfeccionar el léxico que deben aprender los estudiantes de una lengua extranjera.

Este trabajo se enmarca en los estudios de lexicoestadística, que busca mediante diversos métodos, índices estadísticos para las unidades léxicas. Nuestro trabajo es paralelo en cierto modo a este conjunto de estudios, ya que analiza el léxico que tienen los alumnos que estudian el español como lengua extranjera. En esta investigación, por otra parte, se trabaja con tres centros de interés distintos de los habituales en el *Proyecto Panhispánico*, en concreto se atiende a: *Salud; Relaciones familiares y sociales e Informática y Nuevas tecnologías*. Asimismo, se establecen comparaciones con el PCIC atendiendo a los diferentes niveles de competencia lingüística que tienen nuestros informantes no nativos, a fin de extraer posibles déficit léxico.

El corpus de este trabajo lo constituyen 160 encuestas que completaron 40 informantes de diferentes universidades, tanto nativos como no nativos, en este caso sinohablantes, durante el año 2021. Los datos obtenidos han sido analizados a través del programa *Dispogen*. Quiero destacar que la selección de los datos ha sido lo más abarcadora posible, es decir, se incluyen todas las palabras que tengan una relación con el centro de interés (aunque esta sea lejana), debido a que el objetivo principal es conocer y comparar el léxico disponible de los informantes.

1.1. Preguntas de investigación.

- ¿El índice de disponibilidad léxica es una herramienta adecuada para evaluar el léxico?
- ¿Se corresponde el vocabulario que han aprendido los estudiantes de español con el que manejan los nativos? ¿Cuáles son sus principales diferencias?
- ¿Los estudiantes de español muestran menor competencia léxica, medida en términos de disponibilidad, que los estudiantes nativos?
- ¿Los estudiantes reaccionan igual a los distintos centros de interés?
- ¿Una menor cantidad de unidades léxicas en un centro de interés específico es indicio de una menor competencia léxica?

1.2. Hipótesis

Seguidamente se exponen las hipótesis de las que se parte para la realización del trabajo y que pretenden dar respuesta a nuestros planteamientos iniciales.

- El índice de disponibilidad léxica es una herramienta adecuada para evaluar el léxico.
- No todo el vocabulario que aprenden los estudiantes se corresponde con el que realmente emplean los nativos.
- Los estudiantes no reaccionan por igual para todos los centros de interés, que dependerá de su nivel de estudio o necesidades.
- Un mayor número de unidades léxicas puede ser indicio de un mayor dominio de la competencia lingüística.

1.3. Objetivos

Para poder dar respuesta a nuestras preguntas de investigación, confirmar o no nuestras hipótesis y poder seguir una metodología adecuada, nos plantearemos los siguientes objetivos:

- Determinar la influencia de factores socioculturales (variables como sexo, nivel de estudios procedencia...) en el aprendizaje del léxico.
- Conocer las aplicaciones de los estudios de disponibilidad léxica en el ámbito de ELE.

- Conocer el vocabulario disponible de los estudiantes de español en los tres centros de interés analizados.
- Comparar el léxico disponible aprendido por los estudiantes de español con el léxico disponible de los nativos.
- Comprobar la eficacia de los estudios de disponibilidad léxica para evaluar el léxico.

1.4. Metodología

Este trabajo pretende analizar el léxico disponible de unos informantes específicos, en relación con tres centros de interés de carácter relacional. Con la finalidad de responder a los objetivos planteados, el estudio se divide en dos partes complementarias:

- Una primera parte teórica, en la que se ha realizado una revisión bibliográfica detallada de artículos relacionados con nuestro objetivo de estudio, especialmente aquellos basados en investigaciones de disponibilidad léxica relativas a diferentes centros de interés.
- Una parte empírica, en la que, en primer lugar, se ha administrado la encuesta de disponibilidad léxica mediante Survey Monkey y WeChat a los diferentes informantes nativos y no nativos respectivamente. Posteriormente, los resultados se han analizado con la ayuda de la herramienta *Dispogen*. El análisis de los resultados se divide en: análisis cuantitativo que consistirá en analizar y comparar los resultados obtenidos atendiendo al número total de palabras; nº de vocablos; promedio de palabras; índice de cohesión y la densidad del centro de interés. En cuanto al análisis cualitativo, se analiza el tipo de léxico que compone los distintos centros, especialmente aquel léxico con un índice superior o igual a 0,1 y se lleva a cabo un análisis y comparación de las unidades léxicas obtenidas y su aparición en PCIC en los diferentes niveles de competencia lingüística.

1.5. Estructura

El trabajo se estructura en cinco partes fundamentales, más anexos y bibliografía. La primera parte corresponde a la introducción, la cual se estructura en diferentes apartados donde se desarrollan las preguntas de investigación que guían el trabajo, el objetivo de la investigación, las hipótesis que nos hemos marcado y que serán desarrolladas a lo largo de este trabajo, la metodología utilizada y, por último, la estructura del trabajo. La

segunda parte corresponde al marco teórico y en ella se presentan la revisión bibliográfica de los aspectos principales de nuestra investigación:

- El apartado 2.1 establece la definición de competencia léxica atendiendo al MCER y PCIC. Se ofrece una breve explicación del tratamiento del léxico según los diferentes enfoques metodológicos.
- El apartado 2.2 se centra en la definición del léxico y las razones por las que es necesario realizar una selección.
- El apartado 2.3 define la disponibilidad léxica, detalla los diferentes centros de interés, tanto los tradicionales como los ajenos al Proyecto Panhispánico, y, por último, atiende a las aplicaciones de los estudios de disponibilidad léxica.
- El apartado 2.4 estudia la disponibilidad léxica como instrumento de evaluación.
- El apartado 2.5 profundiza en las aplicaciones actuales de los estudios de disponibilidad léxica en ELE, centrándose en la sociolingüística, la psicolingüística, la dialectología, la etnolingüística y la didáctica de las lenguas.

La tercera parte muestra la metodología seguida para la creación y administración de la encuesta a los diferentes encuestados atendiendo a tres centros de interés relacionales: *Relaciones sociales y familiares, Salud e Informática y nuevas tecnologías*. Además, se ofrece el perfil de los encuestados y la organización de los datos obtenidos.

La cuarta parte atiende al análisis cuantitativo y cualitativo de los resultados, a fin de ofrecer y exponer los datos obtenidos de la manera más detallada posible. Este apartado se estructura en dos partes principales:

- El apartado 4.1 ofrece el análisis cuantitativo de los datos, que consiste en analizar y comparar los resultados obtenidos atendiendo al número total de palabras, nº de vocablos, al promedio de palabras, al índice de cohesión y a la densidad del centro de interés de ambos grupos de informante. Todo ello se acompaña de tablas y gráficos que ayudan a interpretar los datos.
- El apartado 4.2 ofrece el análisis cualitativo de los datos, y en él se analiza el tipo de léxico que integra los distintos índices, especialmente aquel léxico con un índice de disponibilidad léxica superior o igual a 0,1. Asimismo, se hace un análisis y comparación de las unidades léxicas obtenidas y su aparición en PCIC en los diferentes niveles de competencia lingüística.

Tras el posterior análisis, se recogen en las últimas páginas las conclusiones, centrándonos, sobre todo, en los resultados más relevantes de nuestro estudio.

Finalmente, se detalla la bibliografía utilizada para la elaboración de este trabajo. En los anexos se ofrece el formulario ofrecido a los informantes (nativos) con la explicación y acceso a las diferentes encuestas, así como también se encuentra los índices de disponibilidad completos, tanto de informantes nativos como no nativos.

2. Marco teórico

2.1 Competencia léxica

2.1.1. La competencia léxica en el MCER y el PCIC

El *Marco común europeo de referencia* (MCER, 2001) establece qué tienen que aprender o adquirir los alumnos que aprenden una lengua e indica de forma explícita que el objetivo de la enseñanza-aprendizaje, además de centrarse en las competencias generales, ha de lograr el desarrollo de la competencia lingüística comunicativa, en la cual se distinguen seis subcompetencias: competencia léxica; competencia gramatical; competencia semántica; competencia fonológica; competencia ortográfica y competencia ortoépica. A su vez, cada una de estas competencias integra conocimientos declarativos, destrezas y habilidades para saber hacer, y estrategias tanto para comunicarse como para aprender (Pérez, 2015, p. 214). Esta clasificación de la competencia lingüística puede ofrecer una visión fragmentada de las competencias, como si fuera posible aprender los conocimientos léxicos de forma aislada, sin embargo, esto no es así. La competencia léxica se inserta de manera horizontal, vertical y transversal en los diferentes capítulos del MCER (Baralo, 2005, p. 28).

El tratamiento de la competencia léxica en el MCER es muy sencillo. La competencia léxica se define como “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo se compone de elementos léxicos y elementos gramaticales.” (MCER, 2001, p. 108). Los elementos léxicos que comprende son los siguientes: (Ibid., p. 108-109)

1. Expresiones hechas, compuestas por palabras que se utilizan y se aprenden como un todo. Las expresiones incluyen
 - Fórmulas fijas, que comprenden exponentes directos de funciones comunicativas; refranes, proverbios, etc. y arcaísmos residuales
 - Modismos, que comprenden metáforas lexicalizadas (*estiró la pata*) e intensificadores, ponderativos o epítetos (*blanco como la nieve*).
 - Estructuras fijas, aprendidas y utilizadas como conjuntos no analizados, en los que se insertan palabras o frases para formar oraciones (*por favor, ¿sería tan amable de + infinitivo?*).
 - Frases hechas como verbos con régimen preposicional (*atreverse a*) o locuciones prepositivas (*por medio de*).

- Régimen semántico, trata de expresiones que se componen de palabras que habitualmente se utilizan juntas (*cometer un crimen*).
2. Polisemia, palabras que pueden tener varios sentidos distintos (*banco*: lugar para sentarse o entidad financiera). Se incluyen miembros de otras clases abiertas de palabras, como sustantivos, verbos, adjetivos o adverbios. Además, se pueden establecer otros conjuntos léxicos con fines gramaticales.

Los elementos gramaticales pertenecen a clases cerradas de palabras como: artículos; cuantificadores; demostrativos; pronombres personales; pronombres relativos y adverbios interrogativos; posesivos; preposiciones; verbos auxiliares y conjunciones.

En el MCER se ofrecen escalas ilustrativas ¹para la gradación del conocimiento del vocabulario y para la capacidad de controlar ese conocimiento:

RIQUEZA DE VOCABULARIO	
C2	Tiene un buen dominio de un repertorio léxico muy amplio, que incluye expresiones idiomáticas y coloquiales; muestra que es capaz de apreciar los niveles connotativos del significado.
C1	Tiene un buen dominio de un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Buen dominio de expresiones idiomáticas y coloquiales.
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y sobre temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria, como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.
A2	Tiene suficiente vocabulario para desenvolverse en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos.
	Tiene suficiente vocabulario para expresar necesidades comunicativas básicas. Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia.
A1	Tiene un repertorio básico de palabras y frases aisladas relativas a situaciones concretas.

Tabla 1. Tabla de gradación de riqueza de vocabulario según el MCER (2001, p. 109).

DOMINIO DEL VOCABULARIO	
C2	Utiliza con consistencia un vocabulario correcto y apropiado.
C1	Pequeños y esporádicos deslices, pero sin errores importantes de vocabulario.
B2	Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación.
B1	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes.
A2	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.
A1	No hay descriptor disponible.

Tabla 2. Tabla de gradación del dominio de vocabulario según el MCER (2001, p. 109)

¹ Los ejemplos que se muestran han sido extraídos del MCER del capítulo 5. *Las competencias del usuario o alumno*.

El Plan Curricular del Instituto Cervantes (PCIC), establece las especificaciones de los descriptores correspondientes a los seis niveles de referencia. Indica de modo general las actividades comunicativas que el alumno es capaz de llevar a cabo en cada nivel y las competencias que podrá activar mediante el uso de las estrategias correspondientes (Pérez, 2015, p. 221). Propone un esquema basado en cinco componentes: gramatical, cultural, pragmático-discursivo, aprendizaje y nocional. Este último componente trabaja con el componente léxico.

El componente nocional está formado por los inventarios de *Nociones generales* y *Nociones específicas*. En un primer lugar, las nociones generales son aquellas que un hablante puede necesitar independientemente del contexto de comunicación, por lo que están referida en su mayoría a conceptos abstractos. Los apartados que presenta este inventario son:

- Nociones existenciales.
- Nociones cuantitativas.
- Nociones espaciales
- Nociones temporales
- Nociones cualitativas
- Nociones evaluativas
- Nociones mentales.

El inventario de Nociones específicas tiene que ver con los detalles más concretos del “aquí y ahora” y se relaciona con las interacciones, transacciones o temas determinados (PCIC, 2006). Asimismo, se afirma la idea de que es un inventario abierto a través de la siguiente afirmación (PCIC, 2006):

Las entradas o exponentes aislados que se presentan en los distintos epígrafes del inventario han sido seleccionados con el objetivo de dar una idea del tipo de unidades léxicas que el hablante debe conocer en cada uno de los niveles, aunque hay que dejar claro que no se trata de una lista cerrada.

La organización de inventario de nociones específicas es el siguiente:

- | | |
|---|---------------------------------------|
| -Individuo: dimensión física. | -Servicios |
| -Individuo: dimensión perceptiva y anímica. | - Compras, tiendas y establecimientos |

- | | |
|--|------------------------------------|
| - Identidad personal | - Salud e higiene |
| - Relaciones personales. | - Viajes, alojamiento y transporte |
| - Alimentación | - Economía e industria. |
| - Educación | - Ciencia y tecnología. |
| - Trabajo | - Gobierno, política y sociedad. |
| - Ocio | - Actividades artística |
| - Información y medios de comunicación | - Religión y filosofía. |
| - Vivienda | - Geografía y naturaleza. |

El inventario es de carácter abierto y modificable en función de las necesidades de los alumnos y de las características de cada situación particular de enseñanza.

2.1.2. Tratamiento del léxico según los distintos enfoques metodológicos.

El léxico de una lengua forma un conjunto abierto difícil de sistematizar. Además, en él convergen muchos aspectos que hacen difícil su enseñanza- aprendizaje, ya sea por su forma, su significado o su uso (Núñez, 2019, p.162).

La enseñanza- aprendizaje del léxico es uno de los aspectos esenciales de todo método de enseñanza, pero el tratamiento que se le da depende de la época y la metodología en que se basen. En este sentido, pueden distinguirse diferentes etapas en la evolución de los métodos de enseñanza de segundas lenguas. En este trabajo, sin embargo, nos hemos centrado en los desarrollos más recientes.

2.1.2.1. El método comunicativo

Tras la aparición del concepto de competencia gramatical en la década de 1950 y el de competencia comunicativo en la de 1970, el concepto de competencia se traspasó al ámbito comunicativo para configurar la progresión y los resultados de aprendizaje (Núñez, 2019, p, 169). Actualmente, y siguiendo las ideas de Canale y Swain, la enseñanza aprendizaje de una lengua extranjera debe atender a la enseñanza de la competencia comunicativa, entendiendo esta como una gran macrocompetencia compuesta por varias subcompetencias: competencia discursiva sociolingüística, estratégica y gramatical.

La subcompetencia gramatical estaría formada por el conocimiento del código lingüístico, del léxico, las reglas de morfología y sintaxis, pronunciación y ortografía. Para alcanzarlo los estudiantes deben dominar tres aspectos fundamentales: la forma, el uso y el significado. El objetivo de los enfoques comunicativos es la enseñanza de la lengua con fines comunicativos, es decir, que los alumnos alcancen una competencia comunicativa, y la gramática es fundamental para este objetivo, ya que como parte fundamental de la competencia comunicativa se encuentra la subcompetencia gramatical. (Sánchez, 2010, p. 149)

El aprendizaje de una lengua debe comprender la adquisición de habilidades que permitan interactuar con otros hablantes (tanto oralmente como por escrito). Para ello es necesario cuatro destrezas: expresión oral y escrita y comprensión oral y lectora. Integran el léxico en sus ejercicios como un instrumento y debe estar presentado junto con aspectos gramaticales, pues su objetivo último es la comunicación. Predomina la perspectiva de uso, por lo que es común un léxico coloquial frente a otros métodos más gramaticales que apuestan por un léxico más culto orientado a la escritura o comunicación formal (Núñez, 2019, p. 169).

El léxico se enseña según las necesidades específicas de los alumnos o los objetivos que quieran potenciarse en las unidades del curso, lo que favorece la organización de las unidades léxicas en áreas temáticas y campos nocional-funcionales. Se parte de simulaciones que reproducen situaciones reales, por lo que para reflejar la lengua en uso se necesitan materiales reales y el vocabulario siempre contextualizado, con una importancia similar a la gramática (Núñez, 2019, p. 169-170).

2.1.2.2. Enseñanza de la lengua mediante tareas

Surge en el ámbito anglosajón en la década de 1990. Promueve el aprendizaje real de la lengua mediante tareas que simulan procesos o situaciones que ocurren en la realidad y a los que deberá de enfrentarse el aula, por ejemplo, abrir una cuenta corriente, una presentación laboral, etc. En este sentido, las actividades serán las herramientas para construir esa gramática del alumno, por lo que el profesor deberá presentar los contenidos gramaticales supeditados a las funciones comunicativas correspondientes. Para construir la gramática del alumno, el profesor utilizará diferentes actividades con objetivos diversos, presentando a los alumnos actividades con el foco en la forma, con el foco en el uso y con el foco en la forma y en el uso al mismo tiempo (Martín, 2010, p. 150). Por lo

tanto, es a lo largo de su realización cuando aparece la gramática, vocabulario o los contenidos socioculturales, por lo que se realiza de modo inverso a como se hacía en el anterior método. La realización de actividades intermedias o subtareas secuenciadas favorece la interacción con otros compañeros y promueve el uso discursivo de la lengua. (Núñez, 2019, p.170).

Atendiendo a Martín (2010, p. 150) la secuencia de trabajo es flexible y depende de factores como el contexto de aprendizaje, la competencia lingüística del profesor, su experiencia, los objetivos o estilos de aprendizaje de los alumnos.

2.1.2.3. El enfoque léxico

Surge como una evolución de los enfoques que dotan a los aprendientes de herramientas lingüísticas para lograr una competencia comunicativa, en concreto, el enfoque comunicativo y el enfoque por tareas. Bajo esta premisa, diversos lingüistas estudian la manera más eficaz de enseñar el léxico en las aulas, sin abandonar la competencia comunicativa (Vidiella, 2011, p. 7). Michael Lewis (1993) fue quien impulsó este cambio de paradigma y cuyos cuatro principios básicos se van a englobar bajo la obra titulada *The Lexical Approach*.

El primer principio que nos encontramos es la importancia de una enseñanza explícita. Tras la enseñanza comunicativa había disminuido la enseñanza explícita del léxico, ya que se creía que los alumnos lo aprenderían implícitamente a partir del contexto; después se confirmó que es necesario un estudio explícito del léxico para dominar una lengua extranjera. Se estima que alrededor del 50 % del discurso nativo está basado en *chunks* o estructuras prefabricadas (Erman y Warren, 2000) y que la importancia del dominio por parte del estudiante de estas secuencias está reconocida. Por ello, las ventajas de su enseñanza son más que evidentes, si partimos de que son imprescindibles para conseguir un dominio nativo. Por un lado, al ser almacenadas y recuperadas como una sola unidad, su procesamiento será holístico, esto quiere decir, que contribuye al conocimiento basado en la memoria, se aplica para codificar y decodificar elementos predecibles o esperados. Por otro lado, contribuyen a la precisión, pues bloquean interferencias en L1 u otra lengua, al mismo tiempo que evitan las estructuras poco naturales resultado de la gramática abstracta o parafraseo (Sánchez, 2010, p. 90). Por tanto, el Enfoque Léxico defiende que los profesores deben estar atentos a los dos tipos de aprendizaje y programar los cursos

de forma que haya oportunidades tanto el aprendizaje implícito como explícito (Higueras, 2012, p. 7). Los dos tipos de aprendizaje son de naturaleza distinta y con objetivos diferentes. Por un lado, el aprendizaje explícito centra su atención en la información que hay que aprender y garantizar su aprendizaje, esto requiere una gran cantidad de tiempo, lo que haría imposible aprender así todas las palabras que necesitamos para hablar. En cambio, el aprendizaje implícito aparece cuando ya se está usando la lengua con fines comunicativos y el tiempo empleado en él es doblemente beneficioso, pero tiene el inconveniente de que su aprendizaje es más lento gradual. Se potencia mediante diferentes tipos de actividades. Sukmen (1997), enumera las diferentes técnicas para potenciar la enseñanza indirecta o explícita (Higueras, 2012, p. 7):

- Crear un gran lexicón.
- Integrar las nuevas palabras con las ya conocidas.
- Proporcionar varios encuentros con las palabras enseñadas.
- Trabajar con las palabras en tareas de distinta complejidad, es decir, no sólo memorizar sino conseguir un nivel de procesamiento profundo.
- Facilitar la creación de imágenes.
- Hacer que las palabras nuevas parezcan reales conectándolas con el mundo real.
- Enseñar los significados más frecuentes de cada palabra
- Ofrecer información sobre familias.
- El profesor debe saber cuáles son las palabras más problemáticas

Desde el enfoque léxico también se asume que los estudiantes no pueden recibir instrucción explícita sobre la inmensa cantidad de ítems léxicos necesaria para dominar la lengua meta. Por ello, se aboga por una instrucción mas bien estratégica, el profesor entrenaría al estudiante a descubrir por sí mismo los bloques prefabricados en el input que recibe para que este aplique estrategias de segmentación y tomar nota del input que recibe (Pérez, 2015, p. 76).

Otro de los principios básicos del enfoque léxico es saber concebir el léxico en unidades complejas con el fin de que se anoten las palabras que suelen coaparecer juntas (colocaciones y expresiones idiomáticas) o las que suelen ir asociadas a un contexto determinado (expresiones institucionalizadas). La adquisición del lenguaje tiene lugar de dos formas: aprendizaje de ítems (o unidades individuales) frente al aprendizaje del

sistema. Una vez que el alumno sabe un segmento léxico o *chunk*, puede analizarlo y aprender el significado de cada palabra (Higueras, 2012, p. 10). En definitiva, este principio se basa en prestar más atención por parte de los profesores y alumnos al reconocimiento y memorización de segmentos léxicos para mejorar la fluidez, para permitir que opere tanto el aprendizaje de ítems asilados, como el aprendizaje de sistema (Higueras, 2012, p. 12).

El siguiente principio se basa en la importancia de la organización sintagmática u horizontal de las palabras, es decir, del co-texto y del contexto, y por consiguiente de las colocaciones (Higueras, 2012, p. 12). La enseñanza cualitativa del léxico es otro de los principios básicos: no se trata solo de aprender palabras nuevas, sino de ampliar el conocimiento que tenemos sobre ellas, conocer sus colocaciones, los adjetivos para calificarlo. También se centra mucho en el registro, en los contextos de uso en los que pueden o deben usarse ciertas fórmulas o expresiones institucionalizadas, y en el significado contextual y cotextual.

Por último, Lewis afirmó que la gramática puede explicarse a partir del léxico y que la lengua es léxico gramaticalizado, y en consecuencia debe dedicarse más tiempo al léxico. El principio de idiomatidad defiende que el léxico está formado por unos amplios *chunks* que el hablante tiene memorizados y que puede combinar al hablar (esta afirmación supera el principio de selección libre). Esta idea hay que interpretarla de acuerdo con la Lingüística cognitiva, es decir, no hacer separaciones entre gramática y léxico (Higueras, 2012, p. 12).

Una de las críticas que recibió el Enfoque Léxico es que no desarrollaba una teoría del aprendizaje, pero en realidad, de los escritos de Lewis se puede entender que el Enfoque Léxico no supone una ruptura con el Enfoque comunicativo, sino un desarrollo de este, por lo que hereda sus presupuestos sobre la enseñanza y añade la importancia de la enseñanza de las unidades léxicas frente a las palabras y los otros principios que se acaban de desarrollar (Higueras, 2012, p. 12).

2.2 Selección léxica

2.2.1. La necesidad de la selección léxica en la enseñanza de lenguas extranjeras

El vocabulario juega un papel principal en el aprendizaje de una lengua pues es el componente que otorga sentido al sistema lingüístico; de ahí su complejidad, que se manifiesta no solo en el gran número de unidades, sino en su variación (diatópica, diastrática o diafásica). Estas razones justifican la necesidad de realizar una selección explícita y consciente de las unidades léxicas a trabajar en cada uno de los niveles de la lengua extranjera. El MCER apoya esta idea mediante la siguiente afirmación: “quienes elaboran materiales de exámenes y manuales están obligados a hacer una selección léxica” (Consejo de Europa, 2002, p. 149). A pesar de esto, en los programas curriculares pocas veces se detallan los criterios que orientaron la selección y la graduación por niveles, lo cual imposibilita su comprensión y análisis, dejando la interrogante de si en verdad hubo algún principio que rigiera el proceso (Paredes, 2015).

Una vez clara la necesidad de una selección léxica, nos centraremos en cómo realizar dicha selección, cómo limitar o establecer lo que es más apropiado enseñar. Gréve y Van Passel en 1971 (Grethel, 2017, p.5) proponen la primera catalogación de los criterios de selección léxica, que reúnen en torno a tres formas de delimitar el vocabulario: selección subjetiva; selección objetiva y selección objetiva corregida.

La selección subjetiva se basa en la valoración o intuición personal del autor sin un sustento científico para ello. Puede tener ciertos aciertos, ya que es el profesor quien mejor conoce el grupo, sus necesidades e intereses; sin embargo, esta actitud no es fiable ni generalizable, debido a que la intuición como criterio exclusivo o “fantasía del autor” puede llevar al fracaso en el aprendizaje del léxico, y por lo tanto, en su comunicación. Por otro lado, los recuentos léxicos en cada material tendrán diferencias significativas (Grethel, 2017, p. 5).

La preocupación por encontrar un orden de enseñanza del vocabulario llevó a realizar secuenciaciones basadas en la frecuencia estadística de los vocablos (Salazar, 2004, p.261), lo que dio lugar a la selección objetiva, que proporciona datos fiables basados en métodos científicos (en este caso léxico-estadísticos). El empleo de léxicos frecuentes equiparaba la frecuencia a la utilidad, lo que ha llevado a admitir que una palabra frecuente es útil, sin embargo, no se puede aceptar que las palabras poco frecuentes no

sean útiles (Bartol, 2010, p.88). Se pueden mencionar diferentes limitaciones derivadas del empleo de la frecuencia en la enseñanza (Salazar, 2004; Bartol, 2010, Tomé, 2015, p. 355):

- No proporciona un orden didácticamente adecuado ya que rompen series léxicas (como por ejemplo los días de la semana); las listas presentan muchas diferencias.
- Hay palabras infrecuentes pero útiles, encontramos palabras muy útiles para la comunicación que aparecen en niveles inferiores.
- En los niveles más altos la frecuencia puede dejar de ser un criterio de selección adecuado.
- Los textos que sirven de base son fundamentalmente escritos, lo que conlleva un marcado sesgo desde la variedad diafásico y diastrático, aunque se está intentando solventar esta limitación con las publicaciones más recientes.
- Falta de correspondencia entre el lema y la unidad léxica ya que se suele agrupar las distintas variantes de una palabra en una misma entrada; en otras palabras, no resuelve el problema de la polisemia.

Por último, la selección objetiva corregida combina las formas de selección anteriormente nombradas, es decir, se apoya en métodos léxico-estadísticos, pero matiza dichos resultados con juicios o criterios personales. Este método puede atender tanto a la eficacia (empleo de una palabra en diferentes contextos); como a la productividad (la palabra da lugar a otras palabras) y la universalidad (palabras utilizadas en todos los países hispanos).

No encontramos consenso sobre cuál es el método más adecuado para la selección léxica, pero autores como Paredes (2015, p. 19) creen que la selección objetiva con una complementación de criterios subjetivos es el proceder más adecuado para la selección de léxico. Con el primer método se atiende a la necesidad de enseñar un vocabulario básico (vocabulario que se utiliza con frecuencia), que depende de la intuición o la capacidad del profesor, por lo que suele producir diferencias notables entre los materiales, en cambio con el segundo, se pretende enseñar el vocabulario más eficaz, por aparecer en un mayor número de contextos y más productivo por permitir un mayor número de derivaciones (Sanz, 2000, p. 638).

Por otro lado, el MCER propone cuatro opciones (Consejo de Europa, 2002, p.149):

- a) elegir palabras y frases clave: en áreas temáticas necesarias para la consecución de tareas comunicativas adecuadas a las necesidades de los alumnos (...);
- b) seguir unos principios léxico-estadísticos que seleccionen las palabras más frecuentes (...) o las palabras que se utilizan para áreas temáticas delimitadas; c) elegir textos (auténticos) hablados y escritos y aprender o enseñar todas las palabras que contienen, y d) no realizar una planificación previa del desarrollo del vocabulario (...).

Una vez presentado las ventajas de selección léxica, y así como las valoraciones de cada uno de los métodos específicos de selección, nos centraremos en la disponibilidad léxica como el parámetro más adecuado para la selección del vocabulario.

2.3. Disponibilidad léxica

2.3.1. Estudios de disponibilidad léxica

El enfoque actual de la enseñanza de segundas lenguas postula la necesidad de utilizar materiales extraídos del contexto real de aprendizaje, entendiendo que estos son las herramientas más adecuadas para el desarrollo de la competencia comunicativa de los aprendices. Este tipo de materiales es, precisamente, el que se recoge en las encuestas sobre disponibilidad léxica (Paredes, 2015, p.2).

Los primeros estudios de disponibilidad léxica surgieron cuando los investigadores franceses se plantearon qué léxico debían incorporar al currículo para la enseñanza del francés en las colonias. Para ello utilizaron criterios objetivos que justificasen sus elecciones léxicas y se recurrió a la frecuencia, partiendo de que las palabras más habituales deben ser las que se enseñen en un primer momento. Pronto se dieron cuenta de que este criterio no resultaba útil, ya que en los listados de frecuencia no figuraban las palabras esenciales para comunicarse, o bien que cuando aparecían, lo hacían en posiciones muy retrasadas, lejos del número razonable de palabras que llevar al aula. En el caso del español² las primeras palabras que aparecerían fueron gramaticales (preposiciones, conjunciones, artículos...), verbos copulativos y auxiliares de escaso o nulo contenido semántico. No fue hasta la posición 47 del listado de las palabras más frecuentes en español según el CREA, cuando apareció la primera palabra con contenido léxico (Paredes, 2015, p. 3).

² Los datos proceden del Corpus de Referencia del Español Actual (CREA).

Las palabras en el discurso se pueden diferenciar en dos tipos: palabras que aparecen siempre independientemente del tema, y aquellas que solo aparecen cuando la conversación trata sobre un determinado tema. La primera característica hace referencia a las *atemáticas*, mientras que la otra característica hace referencia a las *temáticas*. El objetivo de la DL será identificar las palabras *temáticas*, que es más difícil que aparezcan en los listados de frecuencia, pero son aquellas conocidas y usadas por los hablantes en una comunicación cotidiana, y, por lo tanto, el objetivo a conseguir (Paredes, 2015, p.4).

Es en este contexto en el que surge la disponibilidad léxica, que utiliza técnicas de la psicología asociativa para acceder a la competencia de léxica de los estudiantes. Para ello a los encuestados se les pide que actualicen todas las palabras que conocen sobre un determinado tema o *centro de interés*. Asume una doble perspectiva, individual y social: desde el punto de vista individual, las palabras más disponibles serán aquellas que más rápidamente acudan a la mente del sujeto, desde el punto de vista social, serán las que colectivamente más hablantes hayan sido capaces de utilizar. De este modo es posible determinar para cada unidad léxica su índice de disponibilidad, que vendrá dado por la combinación de su frecuencia de aparición (variable social) y de su posición relativa en cada listado (variable individual) (Paredes, 2015, p.4).

El español, es sin duda, la lengua que mejor ha sabido aprovechar las ventajas de este nuevo método. La primera investigación en incluir el español fue un trabajo de naturaleza bilingüe (inglés-español) llevado a cabo por Victory (1971), aunque no hay que esperar muchos años para encontrar los primeros estudios dedicados exclusivamente al español. Los primeros estudios comienzan con Humberto López Morales (1973) en Puerto Rico, en el caso de España; no fue hasta 1983 con Axurmendi Ayerbe que se pretende describir sociológicamente el bilingüismo vasco-español. Tras ella se han llevado a cabo estudios que presentan el léxico disponible y lo analizan en las comunidades de Andalucía, Aragón, Asturias, Cantabria, Castilla-La Mancha, Castilla y León, Cataluña, Comunidad de Madrid, Comunidad Valenciana, Extremadura, Galicia, Islas Canarias, La Rioja, Navarra, País Vasco y en las ciudades autonómicas de Ceuta y Melilla (Hidalgo, 2019, p. 18-19).

Como hemos podido ver, los primeros trabajos sobre léxico disponible estuvieron enfocados hacia una rápida y eficaz difusión de la lengua extranjera, pero después se puso el foco en la lengua materna, desarrollando estudios sobre la descripción de la norma

léxica de una comunidad, el análisis de la organización del lexicón mental, el desarrollo de teorías cognitivas, la comparación de sintopías para determinar posibles carencias o deficiencias léxicas y mejorar los materiales didácticos, la selección y planificación lingüística o el estudio de comunidades bilingües (Hidalgo, 2019, p.20). En los últimos años, desde el año 2000, se ha reorientado hacia la lengua extranjera con estudios para el análisis de errores y la interlengua, el estudio del proceso de transferencia e interferencia, la presencia de extranjerismos, etc. La DL es flexible en tanto que puede abrirse a ámbitos más profesionales o técnicos para satisfacer las necesidades de estudiantes con fines específicos o analizar las consideraciones sociales de determinados sectores. En definitiva, la disponibilidad léxica supone el inicio del cambio para la mejora en la enseñanza de lenguas (Hidalgo, 2019, 20).

Por último y a modo de resumen se exponen las diferentes ventajas de los estudios de disponibilidad léxica (Bartol, 2010, p.92-102):

- La prueba está basada en un corpus, formado por las respuestas dadas a la encuesta. Este vocabulario disponible ayuda a reducir la arbitrariedad en la selección léxica.
- La disponibilidad léxica es medible a través de una fórmula matemática en la que se conjugan la frecuencia y el orden de activación de la palabra. Las palabras son medidas con el índice de disponibilidad, que permite establecer cortes en el grado de disponibilidad y que a su vez puede tenerse en cuenta para fijar el léxico de los diferentes niveles.
- La disponibilidad léxica ayuda a conocer la organización de la memoria semántica y el lexicón mental, de gran utilidad para la enseñanza del vocabulario.
- El léxico está recogido en campos nocionales (o también llamado centros de interés), que más o menos equivale a los temas y subtemas que desarrollan los ámbitos del MCER. Gracias a esta organización en campos se eliminan los problemas relacionados con la homonimia y polisemia, una palabra relacionada con su campo queda claro el significado en ese contexto. Además, permite valorar la integración social de las palabras comparándolas unas con otras.
- Las palabras más disponibles son muy estables en la lengua. La disponibilidad de una palabra se relaciona con tres parámetros: la tipicidad, la edad de adquisición y la familiaridad. Los dos primeros parámetros son los responsables de la unidad del léxico, de las semejanzas que se establecen entre el léxico de diferentes zonas,

ambos parámetros inciden en lo compartido por los miembros de la comunidad. En cambio, el grado de familiaridad puede variar de unas zonas a otras, de la época e incluso de unos individuos a otros.

- La utilización de una misma metodología permite la comparación diatópica.
- Los listados de léxico disponible aportan datos interesantes sobre la influencia de las variables sociales en el dominio léxico de los informantes. La variable que más influencia puede tener sobre el dominio léxico es el nivel cultural de la familia.
- Esta prueba es muy útil para medir el dominio del léxico activo de los estudiantes de ELE.

2.3.2. Centro de interés

Las investigaciones del *Proyecto Panhispánico* han contemplado 16 centros de interés:

01. Partes del cuerpo	09. Calefacción e iluminación
02. La ropa	10. La ciudad
03. Partes de la casa (sin los muebles)	11. El campo
04. Los muebles de la casa	12. Medios de transporte
05. Alimentos y bebidas	13. Trabajos del campo y del jardín.
06. Objetos colocados en la mesa para la comida	14. Los animales
07. La cocina y sus utensilios	15. Juegos y distracciones
08. La escuela: muebles y materiales	16. Profesiones y oficios.

Tabla 3. Centros de interés del Proyecto Panhispánico. Creación propia

Los dieciséis centros de interés del *Proyecto Panhispánico* son comunes a los que establecieron los estudios franceses en los primeros trabajos de disponibilidad. Actualmente son ya muchos los estudios que han añadido a estos centros otros campos, como por ejemplo los trabajos que ocupa la zona de Andalucía han incorporado el centro de interés de “*La mar*”. (López, 2008, p.34) o el trabajo dirigido por Pedro Salcedo Lagos que estudia el léxico matemático, introduciendo cuatro centros específicos: *datos y azar*, *números*, *álgebra* y *Geometría* (Salcedo, Ferreira y del Valle, 2014).

Los centros de interés presentan problemas teóricos relacionados con la forma en que se enuncian, con su número, su alcance y delimitación y con los resultados cuantitativos y cualitativos que se proporcionan (Tomé, 2015, p. 317). Los títulos que engloban dos categorías semánticas favorecen a un efecto de *priming*, que consiste en que el informante

tiende a reproducir primero la última área temática que ha leído u oído, de manera que, si se alterara el orden de categorías, cambiaría también el lugar y la frecuencia de emisión de los vocablos (Tomé, 2015, p. 31). Los centros de interés no son estructuras léxicas basadas en oposiciones entre los significados, sino que se fundamentan en relaciones de tipo asociativo y, como tales, vinculadas muy estrechamente a la realidad y a la experiencia de los informantes (Sánchez- Saus, 2011, p. 342)

Por otro lado, los centros de interés tradicionales aluden a realidades pocas conocidas o en ocasiones confusas para los informantes. Además, no consiguen abarcar todas las parcelas temáticas en que se divide nuestro mundo, por lo que han sido tachados de universalistas, aunque esta última característica puede tener alguna ventaja como la obtención de un vocabulario interesante para cualquier hablante independientemente de su lengua y cultura. Por el contrario, los centros más particulares, son capaces de reflejar las características de la comunidad de habla estudiada, siendo muy interesantes para los estudios etnolingüísticos y sociolingüísticos y para los de didáctica de lenguas extranjeras (Tomé, 2015, p. 317-318). Asimismo, un mismo vocablo puede aparecer en diferentes ámbitos léxicos, y esto refleja una característica inherente al sistema léxico: una palabra, ya sea con el mismo significado o con otro (palabras polisémicas) pertenecen simultáneamente a varios dominios léxicos y en cada uno presente un índice de disponibilidad (Tomé, 2015, p. 318). En definitiva, han sido varios los investigadores que han tratado de resolver o mitigar los problemas anteriormente planteados, ya sea con la introducción de cambios en sus títulos; con la adición de nuevos campos asociativos; supresión de CI; con la redefinición del propio concepto o con la precisión del centro considerado excesivamente amplio (Tomé, 2015, p. 318).

Para terminar con este apartado quiero destacar una cuestión fundamental que se ha mantenido a lo largo de todo el trabajo es que “los informantes no se restringen a aquellos elementos que parecen estrictamente a cada una de las categorías, sino que muy a menudo se dejan llevar por vínculos más laxos” (Sánchez-Saus, 2011, p. 196).

2.3.3. Factores cognitivos de la disponibilidad léxica

Con el fin de establecer qué variables cognitivas determinan la disponibilidad de una palabra se ha tomado en consideración los siguientes factores: edad de adquisición de la

palabra; familiaridad; facilidad con que la palabra evoca una imagen; tipicidad: frecuencia y su longitud (Tomé, 2015, p.50), y tras esto se afirma:

las palabras que eran ejemplos típicos de una categoría, que se correspondían con conceptos familiares y que habían sido adquiridas tempranamente eran significativamente más disponibles que las que resultaban ejemplos atípicos, poco familiares y se habían adquirido más tarde. La frecuencia, la longitud de la palabra y la imaginabilidad, por el contrario, no resultaron significativas en la predicción de la disponibilidad léxica (Tomé, 2015, p. 50).

La investigadora Tomé Cornejo (2015) en su trabajo pretende comprobar si la disponibilidad léxica en una segunda lengua se ve afectada por las mismas variables que la disponibilidad en la lengua materna, y a continuación exponemos sus resultados. En cuanto a la edad de adquisición, cuando se aprende L2 después de la infancia, la secuenciación del vocabulario no se corresponde exactamente con el orden en que se adquiere el léxico de la L1. Los términos aprendidos antes de la infancia se corresponden con juguetes, cuentos o rutinas diarias, con los años estos términos pierden relevancia. En cambio, los adultos que aprenden una lengua extranjera aprenden primero palabras relacionadas con la supervivencia en un país extranjero, tramitación de documentos, etc. Asimismo, el hecho de que ambas palabras sean cognados (semejanza entre la palabra en L1 y su traducción en L2) favorecerá a una mayor disponibilidad de la palabra en L2. En su investigación se deduce que la facilitación de las palabras cognadas puede verse afectada por el nivel de competencia, por el tiempo de exposición a la L2 y por último, por el contexto de enseñanza (Tomé, 2015, p. 51-53).

2.4. La disponibilidad léxica como instrumento de evaluación

Las pruebas de disponibilidad léxica aplicadas a hablantes extranjeros pueden utilizarse como un instrumento de medida de la competencia léxica en diversos centros de interés. Se entiende por competencia léxica “el conocimiento que se debe poseer para poder utilizar la palabra con propiedad” o “la capacidad de reconocer, aprender, recuperar, y relacionar las distintas palabras a nivel oral o escrito” (Jiménez, 2002, p. 152 y Tomé, 2015, 341). Según el objetivo que persiga la evaluación se utilizará un tipo de prueba (Tomé, 2015, p. 341- 342):

- Pruebas de dominio para evaluar el nivel de vocabulario de los estudiantes.

- Pruebas de aprovechamiento para determinar si los alumnos han asimilado los contenidos o han alcanzado los objetivos previstos.
- Pruebas de diagnóstico para identificar en qué áreas temáticas los alumnos presentan más carencias y necesitan una explicación más detallada.

Dentro de las pruebas de evaluación de la competencia léxica podemos distinguir aquellas que miden el tamaño del vocabulario (nº de palabras que el hablante conoce). El número de palabras se considera como un buen indicador para evaluar el nivel de competencia lingüística. Según López- Mezquita (2007, p. 28) “un mínimo de 2000 o 3000 familias de palabras serían necesarias para alcanzar unos niveles aceptables de comprensión”. Este tipo de prueba se divide en pruebas con ítems independientes y descontextualizados. Ambas miden el vocabulario como un elemento independiente y aislado de la competencia lingüística, pero las primeras (pruebas con ítems independientes y descontextualizados) miden el vocabulario sin relacionarlo con ningún contexto, en cambio, las pruebas con ítems independientes y descontextualizados se proporciona un texto párrafo o frase que sirve como contexto (Tomé, 2015, p. 344).

Por otro lado, encontramos aquellas pruebas que miden lo que el hablante sabe sobre esas palabras, es decir, el conjunto de aspectos que conoce de una palabra (miden la profundidad). Ser capaz de conocer y usar una unidad léxica implica tener diferentes tipos de conocimientos, que no se adquieren simultáneamente, sino de forma gradual y que pueden desarrollarse a un ritmo diferente (Izquierdo, 2003, p. 631). En cuanto a los ítems que miden la profundidad todos son considerados independientes y descontextualizados, por lo que miden el vocabulario sin relacionarlo con el contexto.

Las pruebas de disponibilidad léxica también pueden clasificarse según el tipo de actividad que se requiera del participante. Pueden ser productivas, lo que implica respuestas activas en la lengua extranjera, o bien receptivas, donde solo se necesita el reconocimiento de la opción correcta (Milton, 2009, p. 20 y Tomé, 2015, p. 342). En cuanto al método de corrección se distingue entre pruebas subjetivas, cuando el profesor debe juzgar las respuestas según su criterio o pruebas objetivas, cuando la corrección no atiende a criterios personales (Tomé, 2015, p. 344).

En definitiva, las pruebas de disponibilidad léxica podrían considerarse como un instrumento de medida de la extensión del vocabulario (nº de palabras) como de la profundidad (que sabe sobre la palabra), con ítems independientes y descontextualizadas

y mayoritariamente objetiva en cuanto a su corrección. Como herramienta de evaluación ofrece una estimación del tamaño del vocabulario temático al que pueden accederse de una manera inmediata y proporcionar unos indicadores que permiten llevar a cabo comparaciones significativas. Sin embargo, la utilidad para medir la profundidad del vocabulario es más limitada en un primer momento, ya que los únicos aspectos que se pueden observar es el conocimiento de la forma (son capaces de escribir o no una determinada palabra) y la vinculación con el campo semántico. En cambio, si se realiza un análisis más detallado de las respuestas, estas pueden proporcionar desde criterios subjetivos, a otros datos interesantes de las unidades léxicas como relaciones de hiperonimia o sinonimia. También pueden poner de manifiesto la presencia o ausencia de elementos sociales y culturales diferentes de la lengua meta (Tomé, 2015, p. 344-346).

Para concluir y que sirva de enlace al siguiente punto, podemos decir que la disponibilidad léxica se caracteriza por ser un indicador de gran utilidad para la evaluación del componente léxico, a pesar de que por sí sola no pueda medir todos los aspectos que intervienen en el conocimiento de una unidad léxica, además no parece posible encontrar una manera efectiva de evaluar todas las facetas del léxico a través de una única prueba (Tomé, 2015, p. 346). Esta aproximación a la disponibilidad léxica nos lleva a plantearnos cómo se encuentra actualmente su estudio y hacia qué dirección se sitúan sus investigaciones.

2.5. Estado de la cuestión.

Como hemos visto anteriormente, la disponibilidad léxica ha surgido como un criterio alternativo para la selección de las palabras adecuadas que deben enseñarse, y se basa en estudios léxico-estadísticos, por lo que se puede afirmar que el valor de estos datos es incuestionable. En el marco de la enseñanza de segundas lenguas, pueden distinguirse dos grupos de trabajos de léxico disponible en español. Por un lado, los estudios que analizan la disponibilidad léxica en español como lengua extranjera, por otro lado, los trabajos que se sirven del léxico disponible de hablantes nativos para compararlo con el vocabulario que aparece en los manuales de español para extranjeros o para hacer una selección del vocabulario que ha de enseñarse (Tomé, 2015, p. 328). En definitiva, en las investigaciones en las que han trabajado con hablantes ELE, tanto de la misma lengua

materna como distintas, junto a la descripción propia de los estudios de léxico disponible, se ha realizado el siguiente análisis (Tomé, 2015, p. 338):

- Cotejos del léxico disponible de hablantes nativos con el de otros estudiantes de español; con el vocabulario producido en textos por los propios participantes o con manuales de español como lengua extranjera (Fernández-Merino, 2011; Jiménez Berrio, 2012)
- Evolución de la disponibilidad en función del nivel de español. (Samper Hernández, 2001; Sánchez- Saus, 2011)
- Clasificación de las respuestas en función de su categoría gramatical, su registro, composición, condición de extranjerismos.... (Sánchez Gómez, 2006; Frey Pereyra, 2007)
- Análisis de las relaciones semánticas que se forman entre palabras más frecuentes mediante grafos (Sánchez- Saus, 2011)

Por otro lado, los análisis léxico- estadísticos pueden aportar datos de gran interés a disciplinas como la sociolingüística, la psicolingüística, la etnolingüística y la lingüística aplicada a la enseñanza de la lengua, además de ofrecer comparaciones interdialectales. Además, a las nombradas anteriormente se puede añadir *Lenguas en contacto*, *Lexicografía* y *Aplicaciones informáticas* (Gómez, 2004, p. 45).

2.5.1 Sociolingüística.

Nació alrededor de los años 60 en Estados Unidos y se centraba en el estudio de las relaciones entre la sociedad y la lengua. Es la principal perspectiva desde la que se aborda el estudio de disponibilidad léxica y a pesar de que la sociolingüística se desarrollase en la década de los 60, y que por lo tanto los creadores de la disponibilidad léxica no pudieran conocerla como tal, sus bases están presentes en los primeros estudios de léxico disponible (Serrano 2014, p. 73).

Según Gómez (2004, p. 47) el análisis cuantitativo de la disponibilidad léxica de cualquier comunidad dialectal nos ofrecerá una descripción de su norma léxica reflejada en el matiz social de dicha comunidad, y aunque estos estudios ofrecen su mejor y más valiosa contribución a los estudios, a las comparaciones dialectales, no hay que olvidar la incidencia de los distintos factores sociales pueden tener en el léxico disponible de una comunidad de habla. De ahí que, junto a las variables extralingüísticas tradicionales como el sexo, edad, lengua materna o nivel sociocultural, sea necesario introducir otras como el tipo de escolarización o el lugar de residencia, dado que existe una superioridad del

léxico disponible en el medio urbano frente al rural y de la enseñanza privada sobre la pública (Gómez, 2004, p. 47).

En la actualidad, los estudios de disponibilidad léxica se enmarcan en la variación léxica y estudian la relación que hay entre las respuestas de los informantes y unos factores lingüísticos (centro de interés) y unos factores sociales (sexo, nivel sociocultural, tipo de centro, etc.). Hasta la fecha, los factores más comúnmente analizados han sugerido la existencia de tendencias comunes, por ejemplo, la variable sexo, por lo general, se constata que las mujeres alcanzan mejores promedios que los varones (aunque sin grandes diferencias). En relación con el nivel sociocultural, la mejor producción léxica se relaciona con los estratos más altos de la sociedad, definido habitualmente en función del nivel de instrucción y profesión de los progenitores. Otras de las variables es la zona geográfica, normalmente clasificada en la dicotomía urbana/ rural según la ubicación del centro educativo, en este caso encontramos investigaciones donde se constata una mayor producción léxica en las zonas urbanas y estudios donde se corrobore lo contrario. Por último, la variable que más tarde se ha incorporado a los estudios del Proyecto, es el tipo de centro educativo, y hasta ahora predominan las investigaciones donde, al contrastar los datos, los centros privados o concertados alcanzan mejores resultados (Serrano, 2014, p.71 - 75).

2.5.2. Psicolingüística

La Psicolingüística se ha interesado siempre por la estructura del lexicón mental y las relaciones asociativas que se establecen, y para poder profundizar en ella ha llevado a cabo diversos experimentos. Los trabajos de disponibilidad léxica plasman de modo claro las relaciones que los hablantes establecen entre los vocablos; además según López Morales (1999), ofrecen la posibilidad de establecer relaciones entre un gran número de unidades, no solo entre dos, como ocurría con las pruebas clásicas de la psicolingüística que pedían una sola respuesta al estímulo verbal que ofrecía (Samper, 2002, p. 12).

2.5.3. Dialectología

La comparación entre las variedades diatópicas es uno de los dominios de mayor interés, permite destacar la aproximación o diferencia entre variedades diatópicas mediante el análisis cuantitativo de datos y su comparación (Samper, 2002, p. 10).

El primer precedente lo tenemos en el amplio trabajo llevado a cabo por W. Mackey (1971), que coteja el léxico disponible de Francia con el de la zona francófona de Canadá, pero cuando realmente se realizan estas comparaciones con herramientas científicas es a

partir de las fórmulas desarrolladas por J. López Chávez (1992-1995). También hay que citar a J.A. Samper, quien comparó distintos dialectos con los léxicos de Puerto Rico, y Gran Canaria.

2.5.4. Etnolingüística

Un área de aplicación de gran interés, aunque se ha desarrollado muy poco en el ámbito hispánico, es la etnolingüística. Mediante las pruebas de disponibilidad léxica y su posterior análisis se puede comprobar la aparición de unas determinadas unidades léxicas y la ausencia de otras, que pretenden reflejar las costumbres y culturas de los diversos países o comunidades (Samper, 2002 p. 12).

2.5.5 Didáctica de las lenguas

La enseñanza del francés como lengua extranjera fue la primera motivación que tuvieron R. Michéa y su equipo para investigar en el ámbito de la disponibilidad léxica. Por esta razón los trabajos que han seguido a este han tenido también un objetivo educativo (Samper, 2002, p. 13).

Dentro de este ámbito de trabajo podemos encontrar aquellas investigaciones centradas en la lengua materna. A pesar de que los primeros estudios no habían contemplado este objetivo pronto se aprovechó los resultados para mejorar la calidad de la enseñanza de la lengua materna (Samper, 2002, p. 13). Por otro lado, nos encontramos con las investigaciones centrada en la lengua extranjera, en el mundo hispanohablante son varios los autores que han descrito las ventajas que aporta este método de estudio, se resalta la importancia de contar con una base científica a la hora de seleccionar el vocabulario adecuado en la confección de los manuales de ELE y considera que los estudios de disponibilidad léxica pueden resultar muy provechosos, ya que aportan palabras que aparecen en determinados discursos y permiten discernir su utilidad según el orden de aparición (Samper, 2002, p. 14).

Hay que distinguir entre aquellos investigadores que han resaltado la importancia de los estudios de disponibilidad léxica para la enseñanza del español a partir del análisis de los resultados obtenidos por los hablantes nativos y a aquellas lingüísticas que han aplicado la metodología de estos trabajos a informantes extranjeros estudiantes de español, y es en esta última vertiente donde podemos destacar a Alberto Carcedo. Señala que las pruebas permiten examinar las distintas fases de aprendizaje del léxico; descubrir los errores más usuales, así como las tendencias de los hablantes no nativos a lo largo del

aprendizaje; planificar la enseñanza del léxico o realizar comparaciones con los listados de palabras de distintas zonas hispanohablantes (Samper, 2002, p. 16).

Asimismo, el estudio de disponibilidad léxica puede servir de ayuda para averiguar el efecto y la eficacia del método pedagógico usado en las clases, es decir, conocer los puntos débiles de la metodología que se usa (Palapinidi, 2012, p. 6). Gracias a la disponibilidad léxica y a sus encuestas, el profesor podrá encontrar posibles problemas de déficit léxico. El procedimiento para ello es tomar el número de palabras que propone el PCIC como las que deben conocer los aprendices en cada nivel y se contrasta con lo que realmente saben, demostrado por su aparición en las encuestas.

Por último, las encuestas permiten conocer aspectos relativos a la interlengua de los aprendices. Algunas de las líneas de análisis que se están explorando son: identificar los errores que se han producido y analizar sus causas; formular hipótesis sobre el lexicón mental de los hablantes bilingües; detectar posibles actitudes hacia la comunidad meta o lengua meta (Paredes, 2015, p. 25).

3. Encuesta de disponibilidad léxica en los ámbitos de Salud, Relaciones familiares y sociales e Informática y nuevas tecnologías

En el presente trabajo se estudia el léxico disponible atendiendo a tres centros de interés específicos ajenos al *Proyecto Panhispánico*: Relaciones familiares y sociales, Informática y nuevas tecnologías y Salud. Han sido seleccionados por su proximidad a los informantes y a sus características. Los informantes son alumnos universitarios de diferentes culturas, por lo que creemos que *Relaciones sociales y familiares* puede ser un medio de correlación de entre culturas. Lo mismo ocurre con *Salud*; debido a la situación sanitaria, la pandemia provocada por el COVID-19, creemos que será muy significativo analizarlo y por su aplicabilidad a las necesidades personales. Por último, *Informática y nuevas tecnologías*, debido a su impacto y cercanía en la sociedad.

La recogida de datos se ha centrado en alumnos con un perfil académico similar, todos ellos son estudiantes universitarios. Los informantes se encuentran divididos en estudiantes de español como lengua extranjera y estudiantes nativos. Los estudiantes de ELE, en este caso sinohablantes, pertenecen a dos universidades distintas: la Universidad de Beijing (10 informantes) y la Universidad de Nanjing (10 informantes). En cuanto a los informantes nativos, se distribuyen entre alumnos del Máster Enseñanza de Español como Lengua Extranjera de la Universidad Rovira i Virgili (Tarragona), y estudiantes del grado de Educación Primaria de la UCLM, con 10 informantes cada universidad. Se mantendrá la homogeneidad en la edad y nivel cultural de los informantes.

La encuesta consta de un cuestionario inicial basado en datos sociales, geográficos y culturales imprescindibles para el análisis de datos posterior, como la nacionalidad, sexo, lengua materna, nivel de español o el conocimiento de otras lenguas. La definición de las variables y de sus distintas opciones es una fase muy importante, debido a que esta información debe ser sometida a un proceso de codificación numérica que después se introducirá en el programa informático de análisis de disponibilidad léxica (Hidalgo, 2017, p. 52). Posteriormente se procede a la realización de los diferentes cuestionarios atendiendo a los tres centros de interés específicos. El mecanismo de obtención de los datos fue a través de encuestas mediante la plataforma *Survey Monkey* para los alumnos nativos; en cambio para los alumnos chinos se utilizó la aplicación de *WeChat*, para

facilitar su acceso a la encuesta debido a los numerosos problemas para acceder a otras páginas.

Las instrucciones eran claras y sencillas y se encontraban en el correo electrónico y dentro de la encuesta. Los alumnos tendrían un tiempo limitado de dos minutos para hacer la prueba, es decir, para escribir todos los vocablos que se les ocurrieran en relación con el centro de interés, además se hizo hincapié en el número ilimitado de palabras que podían escribir y en la libertad de las palabras ya que podían escribir todas aquellas que le vinieran a la mente, aunque no supieran muy bien cómo se escribe, ya que la ortografía no se tendría en cuenta. Con la limitación del tiempo se pretende que las respuestas no sean el resultado de actos reflexivos, sino de producciones instantáneas, pues se trata de captar el léxico que usarían (léxico activo) y no el léxico que conocen (Saralegui, y Taberner, 2008, p. 749).

El tiempo disponible según se ha indicado será de dos minutos, aunque algunas investigaciones han adaptado este parámetro, como es el ejemplo de López Chávez (2003), que establece un tiempo de respuesta de tres minutos. El principal problema que me encontré fue la dificultad de evaluar el tiempo tardado en realizar la encuesta, especialmente para aquellos que utilizaron WeChat, ya que ahí no podía ver el tiempo.

Los centros de interés están relacionados con el inventario de Nociones Específicas del PCIC. El centro de interés de *Relaciones familiares y sociales* se relaciona con el apartado 4. *Relaciones Sociales*, donde se encuentran los subapartados 4. 1. *Relaciones familiares* y 4.2. *Relaciones sociales*, que incluyen las palabras específicas para este centro de interés, en ellos figuran términos como *marido*, *vecino*, o *fan*. Después, podemos el apartado 13. *Salud e higiene* con diferentes subapartados relacionados con nuestro centro de interés, como 13.1. *Salud y enfermedades*; 13.2 *Heridas y traumatismos*; 13.3 *Síntomas* y 13.4 *Centros de asistencia sanitaria*. Por último, *Informática y nuevas tecnologías* se relaciona con el apartado 9.6 *Internet* y con el apartado 16.4. *Informática y nuevas tecnologías*.

3.1. Perfil de los encuestados

Los encuestados nativos son estudiantes del Máster Enseñanza de Lenguas Extranjeras: español como lengua extranjera de la universidad Rovira i Virgili (Tarragona), y estudiantes de la universidad Castilla La Mancha (UCLM), de la facultad de Educación. Los encuestados no nativos son estudiantes chinos de español de la

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

universidad de Pekín y universidad de Nanjing (China) y cuentan con diferente nivel de competencia, 13 alumnos de B1 y 7 de B2.

A continuación, se muestra una relación entre el número de estudiantes y sus respectivas universidades o centros de estudio:

País	Universidad	Número de alumnos
España	Universidad Rovira i Virgili (URV)	10
	Universidad Castilla La Mancha (UCLM)	10
China	Universidad de Beijing	10
	Universidad Nanjing	10

Tabla 4. Información general de los encuestados. Elaboración propia

En un principio, se pretendía optar por la homogeneidad de los estudiantes, tanto extranjeros como nativos, ya que hubiese sido más adecuado que todos estuviesen el mismo nivel de español, que estudiaran el mismo grado o que estudiaran en la misma situación. Sin embargo, no ha sido posible, debido a la baja participación del alumnado en este tipo de encuestas. Por esta razón se ha recurrido a otras universidades y a otros grupos.

Hay un alto porcentaje de mujeres en ambos casos, en concreto, los no nativos fueron todas mujeres. En el caso de los nativos, hay un 85% de mujeres frente un 15 % de hombres.

3.2. Administración de la encuesta y tratamiento de datos

La encuesta se administró a través de Internet, pero se utilizaron programas diferentes para la creación de la encuesta. En concreto, para los alumnos nativos se utilizó *Survey Monkey*, y para los alumnos no nativos se utilizó la aplicación de *WeChat* y se creó mediante las herramientas que esta misma aplicación propone, debido a que muchos estudiantes no podían utilizar los servicios y aplicaciones utilizados habitualmente en España. En ambos casos se adjuntaba un documento (se ofrece en el anexo 1 el documento enviado a los informantes nativos; para los no nativos es muy similar sólo cambia los enlaces) que recogía la explicación y el funcionamiento de esta investigación y los

diferentes enlaces. Esto se realizó a finales de marzo, pero, debido a la escasa participación de los estudiantes (tanto nativos como no nativos), se estuvo enviando a diferentes estudiantes durante todo el mes de abril e incluso principios de mayo.

A pesar de ser estudiantes nativos con diferente perfil todos se caracterizan por ser estudiantes universitarios y del área de educación o del máster de Enseñanza de Español como Lengua Extranjera. Una vez conseguidas todas las respuestas necesarias para el estudio, se empezaron a analizar los resultados.

El proceso de edición de los datos es una de las fases más sensibles de todo el proceso, ya que hay que uniformar las respuestas de los candidatos bajo una misma forma para que así el programa informático³ pueda actuar. Un trabajo de referencia para esta tarea es el de Sánchez Saus Laserna (2011, p. 227- 244). La autora sigue los criterios comunes a todos los trabajos de disponibilidad léxica y añade criterios propios de la investigación en español como lengua extranjera. A continuación, expondré los criterios comunes (Sánchez- Saus 2011, p. 227- 235):

- La eliminación de los términos repetidos. Aunque se haya dado sido en casos aislados, algunos informantes han repetido una misma palabra dentro de un centro de interés. Cuando esto ha ocurrido se ha eliminado la palabra que ocupa la posición más baja de la lista.
- La corrección ortográfica, sobre todo como resultado del desconocimiento (aparentemente) del uso de la tilde ya que apenas se emplea, aunque encontramos errores como *colegua* (colega); *yendo* (yerno) o *tiritios* (tiritas).
- Unificación ortográfica. Encontramos una gran cantidad de extranjerismos, que podemos dividirlo en aquellos que acepta el DRAE y el DEA, que se han introducido con la grafía que en ellos se consigna (como por ejemplo *router* o *software*). En cambio, aquellos extranjerismos que no aparecen en los diccionarios habituales, pero que son muy usados en nuestra lengua, se han mantenido con la grafía original (como es el caso de *Ipad* o *mouse*). Los informantes no han presentado variación ortográfica.
- Unificación de variantes meramente flexibles. Se han unificado bajo la forma considerada como no marcada: el masculino en los sustantivos, y adjetivos, y el

³ Existen varios programas para el análisis de datos: *Dispolex*, *Dispogen*, *Lexidisp*. Todos ellos presentan un funcionamiento parecido y unos requisitos de edición similares.

infinitivo en los verbos. Se ha mantenido en plural y diferenciado respecto a la forma en singular aquellas acepciones de una palabra que sólo pueden aparecen plural: *padres, hermanos, tíos*, aunque se ha mantenido en el caso de los heterónimos *padre- madre*. Hay ciertos sustantivos que no es tan fácil elegir el género como *computador- computadora*, en este caso se elige la forma femenina, pues es el que prefiere el DRAE.

- Tratamiento sintagmático. Se han aceptado aquellos compuestos sintagmáticos que recoge el DRAE (*lápiz de memoria*), y aquellos de uso habitual, ya que es así como se denomina a los objetos a los que hacen referencia, aunque no se recoja en los diccionarios (*familia política*).
- Tratamiento de las marcas comerciales y nombres de establecimientos o empresas, se ha decidido incorporar todos, y de hecho la mayoría aparece en *Informática y Nuevas tecnologías* y se refieren a marcas o programas habituales y su uso es muy frecuente, en lo referente a su ortografía, todas aparecen con la primera letra en mayúscula (*Facebook, Twitter...*).
- Tratamiento de los acortamientos. En esta investigación se han unificado de acuerdo con las formas plenas mediante el uso del paréntesis, por ejemplo, *tele* y *televisión* aparecerá *tele (visión)*. Se han respetado las abreviaturas complejas o siglas de uso habitual en español, en nuestra investigación podemos encontrar *VIH* o *MIR*

En cuanto a los criterios propios que emplea esta autora vamos a seguir los siguientes (Sánchez- Saus 2011, p. 235- 244):

- Las encuestas de disponibilidad léxica recogen una gran cantidad de errores⁴, de tipos muy diferentes, y ante la inexistencia de un criterio claro para determinar el margen de error por el cual considerar que una palabra no se conoce, he decidido incluir toda palabra que se entienda sin dificultad. No incluiré aquellas palabras que se trate de palabras de lengua diferente o aquellas que no entienda.
- Presentadores y modificadores, algunos alumnos escribieron palabras acompañadas de un artículo, de un adjetivo o de un adverbio. Los artículos fueron todos eliminados con el fin de ajustarnos a otros trabajos, y debido a que en ningún caso fue fundamental para entender a qué se refería el encuestado. En el caso de

⁴ En el artículo de Sánchez- Saus Laserna se puede encontrar una clasificación de los principales errores encontrados. Dicha clasificación se ha tenido en cuenta en este estudio.

los adjetivos o adverbios que a menudo acompañan a sustantivos o verbos se han eliminado aquellos que no suponían una modificación del significado indispensable para que esa palabra esté relacionado con el centro de interés propuesto.

- Amplitud de las relaciones asociativas. He intentado mantener todas las palabras de manera que los datos fueran un reflejo fidedigno de los conocimientos léxicos que tienen. Sin embargo, había elementos que se alejaban demasiado de los temas propuestos; por ejemplo, en el centro de interés Salud hemos encontrado palabras como *pan* o *belleza*, que he decidido mantener, pues se relaciona con ciertos términos que incluye, aunque el tema principal llegue a alejarse, en concreto, este fenómeno hace referencia a una mayor o menor prototipicidad.

Para la edición de los datos he utilizado el programa léxico- estadístico *Dispogen*, que permite generar de forma automática el índice de disponibilidad léxica. El programa fue creado por Max Sergio Echeverría, doctor en Lingüística y uno de los fundadores de la *Revista de Lingüística Teórica y Aplicada*. Desarrolló *Dispogen* y *Dispografo*, dos programas con los que se puede realizar un estudio cuantitativo y cualitativo del lexicon mental de los informantes. Para ello, se han introducido los datos procedentes de las encuestas, y tras introducir los criterios de edición habituales en los estudios de disponibilidad léxica, se han incorporado al programa informático. A través de *Dispogen* se obtiene información general sobre los promedios de respuesta, el número de palabras y vocablos por cada centro de interés, el índice de cohesión y el número de respuestas por informantes junto con la frecuencia, el porcentaje de aparición y la frecuencia acumulada.

En total, de las 50 muestras con las que contábamos al principio, 22 procedentes de informantes nativos y 28 procedentes de informantes no nativos, se terminó trabajando con el léxico disponible de 40 informantes en total; las respuestas restantes no reunían los requisitos para su edición: una escasa producción de unidades (hay algún caso en el que han escrito uno o dos términos), han escrito términos que no se corresponden con ese centro de interés o bien porque no han realizado todas las encuestas, dejando el estudio sin terminar (hay casos en el que solo han realizado la encuesta de datos sociolingüísticos dejando el resto sin hacer, por lo que no se tendrá en cuenta su participación).

Una vez que los datos se han editado, se someten al tratamiento estadístico con la ayuda de herramienta *Dispogen*, que permite utilizar determinadas variables mediante el

uso de unos códigos específicos. En esta investigación se ha tenido en cuenta las variables del número de alumno, el centro de interés y la variable nativo o no nativo. El primer paso es introducir los códigos correctos, en total hay 9 dígitos antes de las unidades léxicas, y a su vez se encuentran divididos en secciones: las primeras cinco cifras (11111) corresponden a la codificación de las variables (5 es lo máximo que permite el programa). Tras un espacio aparecen otros tres dígitos (001) que corresponde con el número asignado al informante (podrá ser hasta un máximo de 999), y por último, también separado por un espacio encontramos dos cifras que hace referencia al centro de interés o campo semántico (01 hace referencia a *Salud*; 02 hace referencia a *Relaciones familiares y sociales* y 03 hace referencia a *Informática y Nuevas tecnologías*). A continuación, se muestra la exposición de los dígitos y variables empleadas con *Dispogen*:

Variable nativo/no nativo	Variable número de alumno	Variable centro de interés
11111: variable nativo	001: informante número 1	01: centro de interés Salud
11112: variable no nativo	002: informante número 2	02: centro de interés Relaciones familiares y sociales
		03: centro de interés Informática y nuevas tecnologías.

Ejemplo

11111 001 01: informante nativo, primera encuesta analizada del centro de interés Salud

Tabla 5. Exposición de los dígitos utilizados con *Dispogen*

Aquí se muestra un ejemplo del formato utilizado e incorporado a *Dispogen* para su análisis, en este caso hace referencia al centro de interés de Salud.

11111 001 01 hospital, enfermero, médico, técnico de rayos, celador, jeringuilla, sala de espera, centro de salud, medicina, tratamiento, pastilla, receta, quirófano, saludable, bisturí
 11111 002 01 hospital, vacuna, enfermero, médico, COVID, quirófano, sala de espera, celador.
 11111 003 01 pierna, medicamentos, pastillas, jarabe, medico, enfermero, hospital, apendicitis, gastroenteritis, dentista, muela, oftalmólogo, uretra, ovarios, útero, mamografía, laboratorio, ojos

Tabla 6. Ejemplo del formato de edición de respuesta para el centro de interés *Salud*

4. Análisis de los datos

A partir de este momento, se ofrece una descripción y análisis del léxico disponible de informantes nativos y de informantes sinohablantes. En esta primera parte se ofrece los datos relativos a la muestra y a los informantes.

4.1. Análisis cuantitativo

El proceso de edición de los datos totales recogidos en las tres encuestas ha dejado un corpus final de 987 palabras repartida en los tres centros de interés trabajados, con una media de 16.45 palabras por informante y de 329 palabras escritas por centro. A diferencia de otros trabajos se ha incluido la densidad del centro, se calcula dividiendo el número de palabras totales por el número vocablos (Sánchez- Saus, 2011, p. 263).

En los trabajos de disponibilidad léxica se diferencia entre palabras y vocablos, de manera que el número total de palabra hace referencia a todas las unidades que han aparecido, incluyendo las repeticiones, en cambio, el número de vocablo solo cuenta cada significante una vez por centro de interés (Sánchez- Saus, 2011, p.261).

En primer lugar, nos centraremos en el análisis de los resultados generales. Por eso en la tabla 7, destinada a los resultados en los alumnos nativos, se recoge el total de las palabras, número total de palabras diferentes, media de las respuestas, el índice de cohesión y densidad del centro. Después la tabla 8 está destinada al análisis de los mismos parámetros desde los informantes no nativos, para así después proceder a la comparación.

Resultados generales de informantes nativos				
Centro de interés	Salud	Relaciones familiares y sociales	Informática y Nuevas tecnologías	Media
Total de palabras	310	333	344	329
N.º de vocablos	196	190	214	200
Promedio de palabras	15.5	16.65	17.2	16.45
Índice de cohesión	0.079081	0.0876315	0.08003738	0.08224996
Densidad del centro	1.5816326531	1.7434554974	1.6074766355	1.6441876667

Tabla 7. Resultados generales de informantes nativos. Elaboración propia

A través de esta tabla se puede comprobar que el centro de interés con mayor número de palabras es *Informática y Nuevas tecnologías*, pues como ya se ha dicho anteriormente es un tema muy cercano y de uso diario, seguido de *Relaciones familiares y sociales* con 333 palabras y por último *Salud* con 310 palabras.

Como ya he explicado anteriormente, los estudios de disponibilidad léxica diferencian entre palabras y vocablos. En números absolutos, en este grupo de informantes se recogen 600 vocablos, sin embargo, aquí se encuentran vocablos repetidos en los distintos centros de interés, ya que nuestro programa *Dispogen* no lo diferencia. Por consiguiente, la media de vocablos por centro de interés es 200 vocablos. Si nos fijamos en el número de palabras diferentes (vocablos), y hablando ahora, por tanto, de riqueza léxica en vez de productividad, encontramos que el centro de interés más rico es *Informática y Nuevas tecnologías* con 214 vocablos, aunque sin diferencias significativas respecto a los otros, ya que tienen 196 vocablos *Salud* y 190 vocablos de *Relaciones familiares y sociales*. Además, es el centro de interés más productivo, pues se recogen mayor cantidad de vocablos. El promedio de palabras también es superior en *Informática* (17.2) frente a *Relaciones* (16.65) y *Salud* (15.5).

En cuanto al análisis del índice de cohesión, se obtiene al dividir la media de las palabras por informante en cada centro por el número de vocablos que ese centro recoge. Nos permitirá averiguar si el centro de interés es compacto, es decir, si las respuestas de los informantes coinciden (en su mayoría), o por el contrario, si las palabras difieren significativamente, en este caso estaríamos hablando de un centro de interés difuso. La escala de cohesión comienza en el 0, e indica que todas las palabras son diferentes unas de otras, y termina en el 1, si todos los informantes han escrito exactamente las mismas palabras (Sánchez- Saus, 2011, p. 264). En cambio, la densidad del centro se obtiene a dividir el número de palabras totales por el número de sus palabras diferentes. Así pues, los centros más cohesionados son aquellos que presentan menos divergencia en las respuestas y es *Relaciones familiares y sociales*, en cambio el centro de interés más difuso es *Salud* tal y como afirman otros investigadores (Moreno y Madrona. 2019).

Estos datos afirman las diferencias existentes entre productividad de cada centro (relacionada con el número total de palabras) y su riqueza léxica (relacionada con el número de vocablos). La relación entre productividad y riqueza no es estable, pues depende de la cohesión y densidad (Sánchez- Saus, 2011, p. 266). Los centros de interés

más ricos desde el punto de vista del léxico son aquellos en los que el número de vocablos es mayor (*Informática y Nuevas Tecnologías*) y el más productivo es también *Informática*. Sin embargo, el centro de interés menos rico es *Salud* (310 palabras), pero no es el menos productivo, ya que se recogen 196 vocablos.

En cuanto los resultados obtenidos de las encuestas destinadas a informantes no nativos se sintetizan en la siguiente tabla:

Resultados generales de informantes no nativos				
Centro de interés	Salud	Relaciones familiares y sociales	Informática y Nuevas tecnologías	Media
Total de palabras	181	208	136	175
N.º de vocablos	117	80	84	94
Promedio de palabras	9.05	10.4	6.8	8.75
Índice de cohesión	0.077350	0.13	0.080952	0.095882
Densidad del centro	1.54700085	2.567	1.6190	1.911

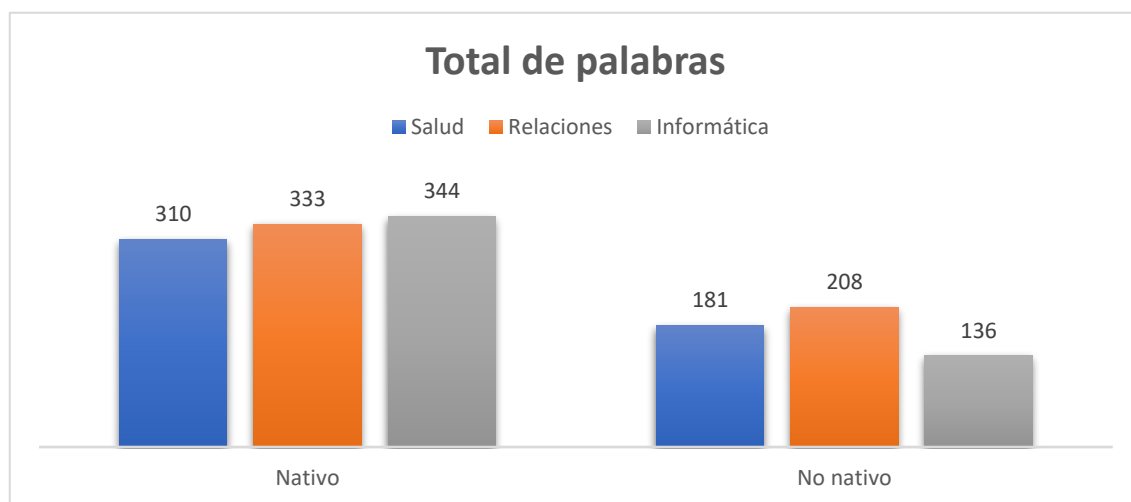
Tabla 8. Resultados generales de informantes no nativos. Creación propia.

El proceso de edición de los datos totales recogidos en las tres encuestas ha dejado un corpus final de 525 palabras repartidas en los tres centros de interés trabajados, con una media de 8.75 palabras por informantes y de 175 palabras escritas por centro, datos significativamente más bajos respecto al grupo de nativos. A través de estos resultados se puede comprobar que la categoría con mayor número de palabras es *Relaciones* (208), seguido de *Salud* (181), y por último *Informática* (136). En números absolutos, se recogen 281 vocablos, siendo el centro de interés *Salud* el que más tiene, por lo tanto, es el que mayor riqueza léxica tiene, en cambio, si hablamos de productividad (relacionado con el total de palabras), es *Relaciones*.

En cuanto al análisis del índice de cohesión o mejor dicho, los centros más cohesionados, en los que hay menos divergencia en sus respuestas, estaríamos hablando de *Relaciones* (0.13). En cambio, el centro el centro más difuso es *Salud* (0.0766).

A continuación, se analizará los resultados obtenidos en las diferentes encuestas correspondientes a cada centro de interés comparándolo entre los informantes nativos y no nativos.

En primer lugar, encontramos la gráfica 1 con el número total de palabras por centro de interés y dividido en los dos grupos diferentes. Como es de esperar, los nativos han producido un mayor número de palabras con 987 frente a 525 del grupo no nativo. Si se analiza cada centro de interés se puede ver que tienen un mayor número de palabras en todos los centros de interés con grandes diferencias, especialmente el hecho de que el centro de interés con mayor número de vocablos en los informantes nativos es *Informática y nuevas tecnologías* (344), y a la vez es el centro de interés con menos palabras para el grupo de los no nativos (136). En el resto de los centros se aprecia la misma tendencia, una mayor producción en el grupo nativos. Es abrumadora la diferencia en los totales, el grupo de nativos sobrepasa en 151 en el promedio de palabras por centro de interés.



Gráfica 1. Diferencias en el total de palabras por centro de interés

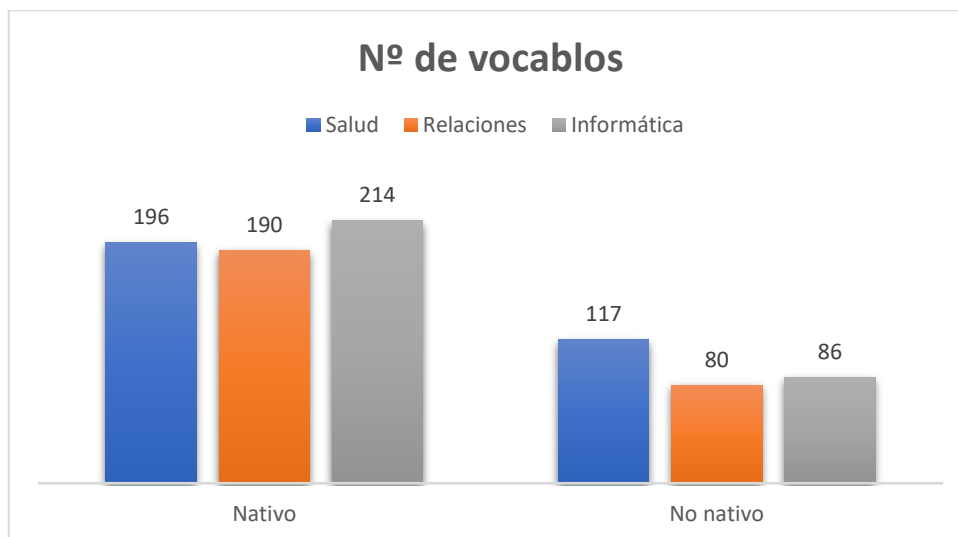
En cuanto al número de vocablos encontramos la comparación en la gráfica 2. Según Bartol (2004, p. 47), este parámetro depende de la naturaleza de cada centro de interés, pero también del número de informantes, ya que, a más informantes, mayor número de vocablos. En nuestra investigación esto no es un problema, ya que hemos trabajado con muestras idénticas (en cuanto al número de informantes), de modo que un mayor número de vocablos nos informará de la riqueza del grupo o centro de interés. Además, en la

investigación de Serrano (2014, p. 346) se afirma que este índice no es tan fiable como el promedio de palabras, ya que en este caso un solo informante atípico puede cambiarlo todo si realiza asociaciones distintas a las demás o porque tenga un mayor bagaje léxico, esto se puede ver reflejado en *Salud*, en el que los informantes nativos han contestado con dos tipos de asociaciones.⁵ En esta misma categoría aparecen numerosos tecnicismos en posiciones más retrasadas del listado que los nativos incluyen en este campo como: *endocrinología, glucosa o insulina*. Evidentemente, en un plan de estudios de ELE este léxico tan específico no es lo que se pretende conseguir, por lo que buscan un léxico general y es normal que no incluyan estos términos.

A través de nuestra investigación se puede concluir que el informante nativo produce una mayor cantidad de vocablos respecto al informante no nativo (600 frente a 281 respectivamente; esto es 319 vocablos más en el grupo de nativos). Al igual que ocurría con el total de palabras, el número de vocablos es mayor en los informantes nativos. Esto quiere decir, que tienen una mayor riqueza léxica en estas categorías. En definitiva y tras analizar estos dos parámetros se puede concluir que el grupo de los nativos no sólo tiene una mayor producción (mayor número de palabras actualizadas bajo ese centro), sino que también tienen una mayor riqueza léxica (mayor número de vocablos o asociaciones actualizados).

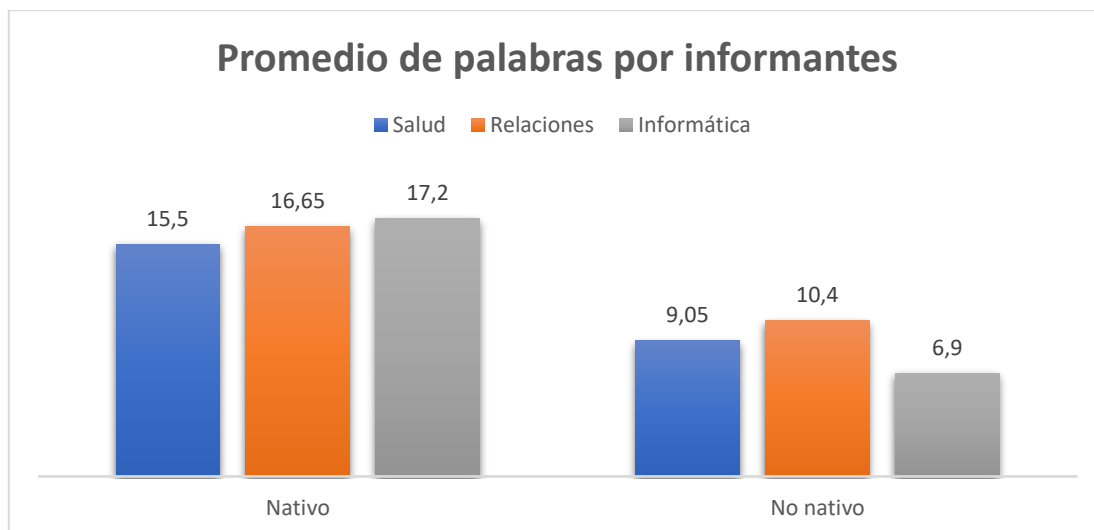
En la introducción de este trabajo se partía de varias hipótesis, entre ellas destacamos que los estudiantes no reaccionan por igual a todos los centros de interés, que dependerá de su nivel de estudio o necesidades, y por otro lado, a mayor número de unidades léxicas se puede traducir en un mayor dominio de la competencia lingüística. Es en este momento cuando podemos dar respuesta a estas hipótesis, por un lado, es cierto que los estudiantes, tanto nativos como no nativos no reaccionan por igual a todos los centros de interés, y que como hemos dicho, dependerá de factores externos, como sus estudios o residencia. Sin embargo, el mayor número de unidades léxicas (palabras) no se puede relacionar con una mayor competencia léxica, ya que encontramos otros parámetros como el total vocablos (palabras diferentes) vinculado directamente con la riqueza léxica.

⁵ Véase el apartado 4.2.1.2. Relaciones familiares y sociales.



Gráfica 2. Diferencia en el número de vocablos por centro de interés

A continuación, se analiza el promedio de palabras por informante y por centro de interés. Dicha magnitud deriva del total de palabra y se considera el índice más adecuado para realizar comparaciones interdialectales entre distintas investigaciones (Serrano, 2014, p. 328). El grupo de informantes nativos supera al otro en todas las categorías, y, además, las diferencias son muy significativas con un promedio de 16.45 palabras el grupo de nativos frente a 8.75 palabras el grupo de no nativos; es decir, los nativos producen casi el doble de palabras que el grupo de no nativos. Al igual que ocurre con los totales de palabra, el centro más productivo para los informantes nativos es *Informática*, que produce como mínimo 9 palabras por informante, y con un promedio de 17.2. En cambio, el grupo de los no nativos, el centro con mejor promedio es *Relaciones*, produce como mínimo 5 palabras por informante, con un promedio de 10.4 palabras. Se observa que incluso la peor área en el grupo de los nativos (*Salud*, con un promedio de 15.5) tiene un resultado superior al que se produce en la mejor categoría de los informantes no nativos (*Relaciones*, con un promedio de 10.4).

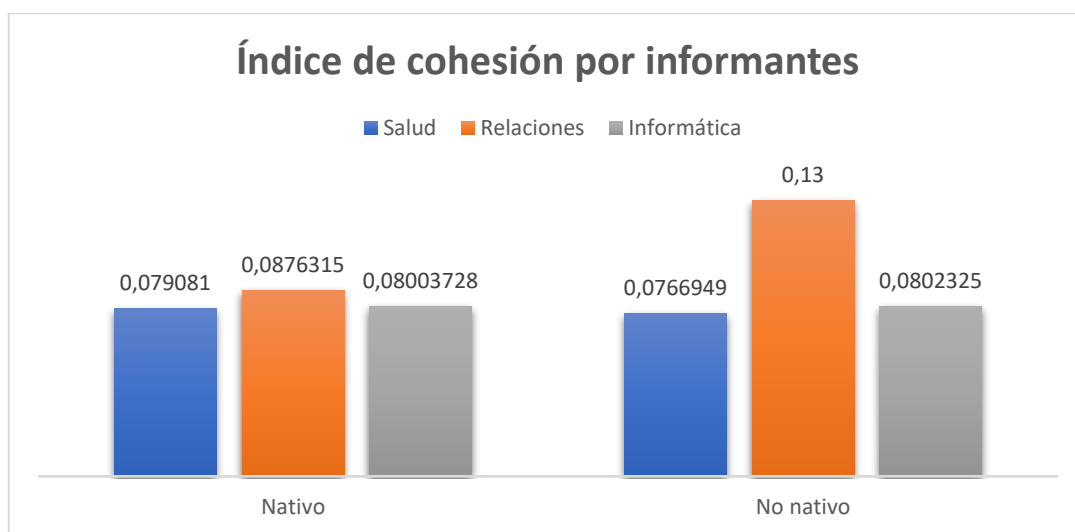


Gráfica 3. Promedio de palabras por informantes

El índice de cohesión y la densidad léxica son dos índices importantes para la disponibilidad léxica, pues ayudan a conocer el léxico que comparten los encuestados ante un estímulo. Hoy en día, todos los investigadores coinciden en que los valores obtenidos por el índice de cohesión resultan más adecuados (Serrano, 2014, p. 353), por lo que nos centraremos en el índice de cohesión (gráfica 4). Según Hernández Muñoz (2006, p. 155) “el índice de cohesión es una manifestación cuantitativa de los efectos que produce en las respuestas de los hablantes el hecho de que los centros de interés sean tan diferentes en su concepción epistemológica”. Además, clasifica los centros de interés según su estructura interna y concluye que las áreas que generalmente tienen un índice de cohesión bajo son aquellas que tienen estructuras relacionales o radiales, mientras que las que tienen índices altos, responden a categorías “naturales” (Serrano, 2014, p. 354).

Los resultados obtenidos son muy similares; aunque las cifras absolutas varían levemente, el rango que obtiene cada área es muy próximo (salvo en el centro *Relaciones*). A nuestro juicio esta es la categoría para analizar con más semejanzas, a excepción de la gran diferencia en el centro de *Relaciones* entre ambos grupos. Como se muestra en el gráfico las áreas más compactas, entendiendo estas como el resultado que más se aproxime a 1, en concreto, para el caso de los nativos es *Relaciones* (0.0876315), y para los no nativos es también *Relaciones*, (0.13). En cuanto a las categorías más difusas, entendiendo esto como las más lejanas a 1, en el caso de los nativos es *Salud* con 0.079081, y para los no nativos también es *Salud* con 0.0766, es decir, las áreas temáticas más difusas y compactas coinciden. La coincidencia en la cifra exacta o muy similar es

igual en el centro de *Informática* (0.080), con esto podemos concluir que, a pesar de las diferencias entre la productividad y riqueza, la cohesión interna en relación con este índice presenta gran similitud en ambos grupos, y las diferencias cuando ocurre, como en el caso de *Salud* parece poco significativas. Mención aparte son los datos obtenidos para *Relaciones*, (0.13 frente a 0.18395 del grupo nativo), pero en ambos casos es la categoría más compacta, especialmente la de los informantes no nativos. De cualquier modo, es llamativo que este sea el primer parámetro en el que coincidan las categorías superiores e inferiores en ambos grupos.



Gráfica 4. Índice de cohesión por informante

4.2. Análisis cualitativo

En este apartado se propone analizar las unidades léxicas con un índice superior a 0,1 para conocer qué palabras son las más utilizadas por nuestros informantes. También se analizará las diferentes categorías gramaticales y la proporción de elementos que se categorizan como sustantivos, adjetivos, verbos y otros junto con su proporción, para de esta forma conseguir una visión global del centro. Por último, se analizarán las palabras más frecuentes (producidas al menos por dos informantes) y se comparará con el PCIC y sus diferentes niveles.

4.2.1. Unidades léxicas con un índice superior a 0,1

Se analizará el tipo de léxico que compone los distintos centros de interés y se realizará una comparación de las palabras que coinciden (señaladas en negro), y aquellas palabras que se parecen. Se trabaja con las unidades léxicas con un índice de disponibilidad superior a 0,1, y que han sido utilizadas por al menos tres informantes, tanto de grupo nativo como no nativo. Nos hemos decantado por este valor por ser tradicionalmente muy usado en los trabajos de disponibilidad léxica, prescindiendo así de las unidades usadas de manera individual o por pocos informantes (Hidalgo, 2019, p. 68).

En las sucesivas tablas aparecen las unidades léxicas con un índice de disponibilidad léxica superior a 0, 1 pertenecientes al grupo de informantes nativos y no nativos y si hablamos de porcentajes y de número de unidades léxicas utilizadas todas por al menos tres informantes suponen porcentajes en torno al 10 % para grupo nativos, en concreto: *Salud* con 8.12%; *Relaciones familiares y sociales* con 12,10% e *Informática* con 9,34% para el grupo de nativos, en contraste el grupo de los no nativos tiene unos porcentajes de 7.627% para *Salud*, con un 20% el centro de interés de *Relaciones*, y por último, 8.3 para *Informática*.

Tal y como se puede comprobar hay una mayor presencia de unidades léxicas univocales, especialmente sustantivos. Entre los términos incluidos en las sucesivas tablas se puede ver sustantivos y sustantivos compuestos (*redes sociales* y *centro de salud*), verbos como *reír* y adjetivos como *divertido* o *sentimental*. En general, pese a tener todos los informantes no nativos un nivel entre B1 y B2 suelen recurrir, en su mayoría a un vocabulario más sencillo y práctico, al igual que les ocurre a los nativos.

No hay duda de que la posición de ciertas unidades léxicas puede estar influida por el perfil de los informantes y por la realidad inmediata. Así podemos explicar la aparición de *redes sociales* en la segunda posición en el grupo de nativos, que no aparece en grupo no nativos. Lo mismo ocurre con términos como *COVID* o *virus* que aparecen en posiciones muy elevadas dentro de los grupos.

Dentro de la variación diatópica se distingue entre geográfica o social, en nuestro caso nos interesa la variedad geográfica ya que vamos a encontrarnos con vocablos propios de algunas variedades del español. No obstante, dentro de la limitación que impone la metodología, ya que recoge las palabras fuera de contexto, únicamente permite detectar aquellos dialectismos que suponen una variante diatópica por su significante. En nuestra

investigación resulta muy poco frecuente, y de hecho suelen tratarse de caso de variantes del español de américa (Sánchez-Saus, 2011, p.381). A continuación, se desarrollará el análisis atendiendo a cada centro de interés

4.2.1.1.Salud

En primer lugar, se analiza este centro de interés atendiendo al grupo de nativos. Resulta de interés que este sea el centro con menos palabras seleccionadas con un índice superior o igual a 0.1. Está formado en primer lugar palabras que hacen referencia a profesionales como *médico, doctor, enfermero, técnico de rayos, oftalmólogo, celador, dentista o secretario*, también encontramos diferentes síntomas como *tos, sangrar, dolor de cabeza, jaqueca*, diferentes medicamentos o utensilios como *insulina, jeringuilla, vacuna, antiinflamatorios, anticuerpos* etc. Por último y aunque resulte dudoso que pueda tenerse en cuenta dentro de *Salud* aparecen numerosas partes del cuerpo como *corazón, pulmón, riñón*.

La palabra con el índice más alto es *médico* y ha sido repetida hasta 13 veces, en concreto este término aparece en el nivel A2 en diferentes inventarios (7. *Trabajo*; 11. *Servicios*; 13. *Salud e higiene*). El resto de las palabras se caracterizan por ser sustantivos muy sencillos y próximos al centro a excepción del término que ocupa la última posición de esta lista (*deporte*) que ha sido utilizado hasta 3 ocasiones diferentes, pero en nuestra opinión con una relación más lejana con este centro que otras unidades léxicas. Las primeras palabras pertenecientes al grupo de los nativos aparecen en los primeros niveles del MCER, como *médico, enfermera o enfermedad* en cambio a medida que descendemos encontramos palabras más complejas y específicas al ámbito, por lo que pertenecerá a niveles superiores como nivel B1- B2, en los casos de *vacuna; venda o termómetro*

En cuanto al grupo de los no nativos se observa una clara disminución de términos, en concreto aparecen 9, lo que supone un 7.69% del total de palabras. La palabra con mayor índice es *hospital*, que aparece hasta 10 veces, además este término se caracteriza por aparecer en los primeros niveles del MCER en diferentes inventarios (11. *Servicios*; 13. *Salud e higiene*; 20. *Geografía y naturaleza*), seguida de *enfermedad, sano, enfermo, etc.* Hay vocablos que pertenecen a otros niveles superiores, como es el caso de *sano*, que aparece en un B1; *fiebre o virus* en un B2 o incluso un C1 con *saludable*. Es significativo que los informantes no nativos introduzcan más variedad de categorías gramaticales; en nuestro ejemplo encontramos 3 adjetivos (*sano, saludable y enfermo*), y además son

términos muy cercanos al centro de interés. Es llamativo como los informantes no nativos han relacionado la salud con la alimentación, por esta razón encontramos términos como *vegetariano*, *caloría*, *azúcar*, o *alimentación*, también encontramos términos relacionados con las distintas partes del cuerpo como *corazón* o *riñón*.

Atendiendo a la comparación de ambos grupos, podemos decir que sus palabras no son muy similares: solo coinciden 4 palabras de 9 que tiene el grupo nativo, aunque el resto de las palabras también han sido producidas por los informantes nativos, pero con un índice de disponibilidad menor. Como es habitual en los trabajos de DL, se observa una mayor presencia de unidades léxicas univerbales, especialmente sustantivos, aun así, se incluyen 28 UL pluriverbales (23.72 % del total frente a 9.69% de los nativos) como *dolor de cabeza*, *salud pública*, *salud mental*, o *fortalecer la salud*.

RELACIONES FAMILIARES Y SOCIALES			
NATIVOS		NO NATIVOS	
Vocablo	IDL	Vocablo	IDL
Médico	0, 5999102	Hospital	0, 33152789
Enfermero	0, 49031895	Enfermedad	0, 23561911
Hospital	0, 4298583	Sano	0.21996601
Vacuna	0.33427423	Enfermero	0, 20269756
Enfermedad	0, 31300217	Fiebre	0, 1522105
Medicina	0, 26306528	Saludable	0, 14367294
Centro de salud	0, 17613271	Virus	0, 1373459
Cáncer	0, 176113271	Médico	0, 12015184
Termómetro	0, 14256755	Enfermo	0, 11031181
Farmacia	0,1306881		
Jeringuilla	0,12084819		
Diabetes	0, 11931143		
Covid	0,118641		
Jarabe	0, 11506802		
Venda	0, 10905311		
Deporte	0,10857002		
	0, 1057172		
Total: 16		Total: 9	

Tabla 9. Palabras del centro de interés de Salud con un índice igual o superior a 0.1 en ambos grupos

4.2.1.2. Relaciones familiares y sociales

Como se ha dicho anteriormente este centro es el más cohesionado del estudio, debido a que ambos grupos realizan una serie de relaciones de parentescos, por lo que sus elementos más disponibles serán muy similares. Esta categoría está formada por el nombre de las diferentes relaciones de parentesco: *madre, padre, tíos, abuelos, cuñado, suegro, yerno, etc*, así como por otros sustantivos relacionado con ello y con las personas cercanas como, *novio, vecino, pariente, amigo*, los sentimientos que generan, *amor, respeto, comprensión, confianza o amistad* u otros más alejados de su temática como *secretos, estudiante, profesionalidad o taxista*.

En cuanto al análisis del centro de interés de *Relaciones familiares y sociales* a partir del grupo de los nativos, destaca la palabra *amigos*, que ha sido utilizada hasta 12 veces, y al igual que el caso anterior es una palabra que aparece en los primeros niveles del MCER, en concreto, en un nivel A1 en 4. *Relaciones personales*. Las palabras de este centro se caracterizan por ser principalmente sustantivos. A partir de este listado queda patente que se trata de un campo léxico que ha generado relaciones semánticas muy cerradas y es compartido por muchos informantes, aunque se puede ver como hay dos tipos de informantes nativos con diferentes interpretaciones del centro de interés, por un lado aquellos que lo han entendido como una sucesión de diferentes relaciones de parentesco (*tíos, padres, novios...*), y por otro lado, aquellos que lo han entendido como el sentimiento que te generan (*respeto, tolerancia, empatía*), es por esta razón por la que este grupo tiene mayor grado de dispersión al comparación con los no nativos que se han centrado más en el listado de las diferentes relaciones familiares o sociales. Al igual que el centro anterior las primeras palabras pertenecientes al grupo de los nativos aparecen en los primeros niveles del MCER, en cambio a medida que descendemos encontramos palabras más complejas como *amor; amistad o confianza* que pertenecen a un nivel B1-B2, lo mismo ocurre con palabras como *cuñado; suegro o matrimonio*, aunque los nativos no han utilizado términos con un nivel de competencia léxica mayor.

En cuanto al análisis de los informantes no nativos y a diferencia del anterior caso, aparecen más términos respecto a sus totales, en concreto se trata del 20% del total de vocablos frente al 12%. Destaca la palabra *padre*, y ha sido utilizada hasta 10 veces, seguida de las diferentes relaciones de parentesco y al igual que el caso anterior, es una palabra que aparece en un nivel A1 (4. *Relaciones personales*). Al igual que en el grupo

de los nativos encontramos una todos los tipos de relaciones familiares y los sentimientos asociados a ello como *amor*, *amistad*, etc. A diferencia del otro grupo, los no nativos han introducido diferentes profesiones que poco tiene que ver con esta categoría, como *profesor*, *bombero o taxista* y aun así sigue siendo el centro más compacto.

Atendiendo a la comparación de ambos grupos, podemos decir que sus palabras son muy similares, en concreto, coinciden 13 de 16 palabras que tiene el grupo no nativo, debido a que las palabras con mayor disponibilidad son las relacionadas con el listado de tipos de relaciones familiares que hay. Este conjunto de términos y de similitudes pone de manifiesto el elevado grado de pertinencia de las unidades léxicas respecto al centro de interés. Las palabras de este listado se caracterizan por ser unidades léxicas univerbales y en concreto, sustantivos, aunque en este caso aparecen otras categorías gramaticales en posiciones avanzadas, manteniendo una relación muy estrecha con la categoría, en concreto, no aparece un verbo hasta la posición 31 (*ser uña y carne*) frente a la posición 20 del grupo de los nativos (*reír*). El adjetivo también aparece en ambos casos y en posiciones avanzadas en la lista, por un lado, el grupo nativo con *divertido* (posición 74) frente a *sentimental* (posición 46). Las unidades léxicas pluriverbales predominan en el grupo no nativo, con un porcentaje de 13.75% (11 expresiones pluriverbales) frente a 7.89% (15 expresiones pluriverbales) del grupo nativo, y encontramos ejemplos como *familia nuclear*, *ser uña y carne o reunirse con la familia*.

RELACIONES FAMILIARES Y SOCIALES			
NATIVOS		NO NATIVOS	
Vocablo	IDL	Vocablos	IDL
Amigos	0,41330367	Padre	0,42883524
Hermanos	0,32712802	Abuelos	0,42385229
Amistad	0,32371157	Madre	0,38229579
Abuelos	0,31399727	Hermanos	0,32952142
Padres	0,27084678	Hijos	0,30347642
Primos	0,26748309	Tíos	0,28956378
Hijos	0,25378013	Padres	0,21539387
Tíos	0,24526665	Familia	0,18264022
Amor	0,22590092	Yerno	0,15395156
Confianza	0,16488248	Primos	0,15025571
Sobrinos	0,16159435	Sobrinos	0,13089283
Familia	0,16151519	Amigos	0,12939881
Cariño	0,15863459	Nuera	0,1265316
Unión	0,15305945	Novios	0,1210077
Suegros	0,14839663	Amistad	0,10798772
Madre	0,14727235	Nietos	0,10215533
Padre	0,14454469		
Compañeros	0,12582259		
Colegas	0,11620265		
Reír	0,11133559		
Cuñados	0,10623236		
Novios	0,10229844		
Matrimonio	0,10161138		
Total: 23		Total: 16	

Tabla 10. Palabras del centro de interés de Relaciones con un índice igual o superior a 0.1 en ambos grupos

4.2.1.3. Informática y nuevas tecnologías

Por último, en el centro de interés de Informática está formado por numerosos dispositivos electrónicos y las partes que lo forman como *móvil, computadora, ordenador, televisión, impresora, disco duro, memoria USB, microondas, pizarra digital*, en cuanto a las partes que lo forman *encontramos pantalla, ratón, teclado, o cables*. También aparecen las diferentes acciones que podemos hacer con ellos *navegar, chatear, descargar, teletrabajo, conectarse, cargar*. Por último, aparecen numerosas aplicaciones y marcas comerciales como *Twitter, IOS, Apple, Google, Samsung, etc*

Los vocablos más disponibles son 19, y tienen una estrecha relación con el centro, van desde aparatos electrónicos hasta acciones o componentes electrónicos. Destaca la palabra *ordenador* en el grupo de nativos y ha sido usada 14 veces, dicha palabra aparece repetida varias veces a lo largo del inventario de nociones específicas, en concreto aparece en 3. *Identidad personal*; 6. *Educación*; 7. *Trabajo*; 16. *Ciencia y tecnología*. Hay que destacar que los informantes nativos no han introducido anglicismos o neologismos hasta este centro que encontramos palabras como *Wifi* o *software*, es cierto que estas palabras se encuentran más que extendidas en nuestra sociedad a pesar de que tengan una traducción al español como es el caso de *red inalámbrica*. Todas las palabras que han introducido son sustantivos, y se corresponden con aquellos que usan en su día a día, quizá por esta razón tiene una frecuencia mayor que el resto de los centros, con una media en su frecuencia de 6. En este caso los nativos no han optado por utilizar palabras sencillas (entendiendo esto como las pertenecientes a los primeros niveles del MCER), a excepción de *móvil* o *Internet*.

He de destacar este centro de interés y en especial el grupo de los nativos, ya que se muestra claramente influido por proyección académica, ya que han utilizado numerosos términos como *innovación didáctica*, *plataforma virtual*, *ABP*

En cuanto al análisis del grupo de los no nativos, las palabras con mayor índice de disponibilidad léxica se reduce significativamente, sólo aparecen 7 palabras. La palabra con mayor disponibilidad es *Internet*, y ha sido utilizada hasta en ocho ocasiones diferentes. Esta palabra aparece en el primer nivel de MCER en diferentes inventarios (6. *Educación*; 7. *Trabajo*; 9. *Información y medios de comunicación*; 16. *Ciencia y tecnología*), el resto de las palabras también pertenecen a los niveles inferiores a excepción de *ratón* que pertenece a un B1.

Atendiendo a la comparación, hay que destacar que la mayoría de las palabras con mayor disponibilidad de los informantes no nativos coinciden con los nativos, en concreto 5 de sus 7 palabras están presentes en la otra lista. Es significativo la convivencia de las variedades diatópicas de *ordenador/ computadora*, en concreto aparece hasta 7 veces en los informantes no nativos frente a 2 veces en los nativos, por lo que predomina la variante peninsular. En ambos casos son frecuentes los anglicismos, pues todos los avances e investigaciones se realizan en esta lengua, y encontramos ejemplos como *hacker*, *software*, *Tablet*. Además, se registran en ambos casos los nombres comerciales de

aplicaciones como *Facebook*, o en el caso de los no nativos *WeChat*. Hay que destacar la gran presencia de siglas especialmente en el grupo de los nativos que producen hasta 10, frente a 2 del otro grupo, ejemplo de ellos son *ADSL*, *TIC*, o *5G*. Por último, relacionado con las unidades pluriverbales y siguiendo los ejemplos anteriores los informantes no nativos producen más unidades pluriverbales con un porcentaje de 26, 19% frente a 17, 28%, y algunos ejemplos son, *inteligencia artificial*, *modificación genética*, *enseñanza online* o *pasar por Internet*.

RELACIONES FAMILIARES Y SOCIALES			
NATIVOS		NO NATIVOS	
Vocablo	IDL	Vocablo	IDL
Ordenador	0,66376311	Internet	0,30657083
Redes sociales	0,32823709	Ordenador	0,30018049
Wifi	0,31271586	Computadora	0,28326792
Ratón	0,27429357	Móvil	0,22965693
Teclado	0,25059706	Teléfono	0,15956722
Móvil	0,24296674	Ratón	0,11680652
Aplicación	0,24023373	5G	0,11582424
Web	0,21663828		
Tablet	0,21122356		
Internet	0,20919156		
Software	0,18433514		
Blog	0,17297831		
Portátil	0,15635972		
Facebook	0,14735369		
Red	0,14134337		
Google	0,12813625		
Pantalla	0,11418649		
Conexión	0,10185039		
Computadora	0,1		
Total:	19	Total:	7

Tabla 11. Palabras del centro de interés de Informática con un índice igual o superior a 0.1 en ambos grupos

4.2.2. Análisis de las categorías gramaticales

En la tabla 12 se observa la distribución o peso de las categorías gramaticales según los distintos centros de interés. La categoría predominante en ambos grupos es el sustantivo, algo propio de esta disciplina, lo que en algunos casos ha podido considerarse como una limitación (Hidalgo, 2019, p. 110). Existen trabajos donde se modifica la nomenclatura de algunos centros o se incorporan nuevos con el propósito de aumentar la

presencia de otras categorías gramaticales, como es el caso de López Chavea (1994), que utiliza directamente el rótulo *Adjetivos*, o el caso de Ayora Estaban (2006), que introdujo *Defectos y cualidades físicas y morales* o *Acciones*, con el objetivo de captar adjetivos y verbos (Tomé, 2015, p. 235). Nuestra investigación no lo tenía como fin, por eso se encuentran cifras tan poco significativas.

Primero, se analizará el grupo de los nativos. Los verbos destacan en *Relaciones* con un total de 29 (15.2%) destacan frente a un 6 (3.04%) y 4 (1.86%) del resto. En cambio, los adjetivos tienen mayor aparición en *Informática* (4.67%), aunque cifras muy poco representativas si lo comparamos con los sustantivos.

En cuanto a los informantes no nativos destaca una mayor presencia de otras categorías gramaticales como adjetivos, verbos e incluso adverbios como *saludablemente*. Al igual que el grupo de nativos destacan los sustantivos, pero con diferencias menores. Los verbos destacan en *Salud* con 14.520%, frente a un 10% o 11.90%. En cambio, los adjetivos destacan en *Relaciones* con un 12.5%, frente a un 10.25 % o 5,95%.

Es significativo que sea en la categoría *Informática* donde más adjetivos producen los informantes nativos, y donde menos produzcan los informantes no nativos, con un 4.67 % frente a un 5.95%. A pesar de ello, los informantes no nativos tienen unos porcentajes mayores de producción de adjetivos. Es significativo el caso de los verbos, pues los informantes nativos tienen un porcentaje mayor, 15.2% en *Relaciones*. Esta es la única categoría gramatical donde los nativos tienen porcentajes mayores.

	C1	Sustantivos	%	Adjetivos	%	Verbos	%	Otros	%
NT	Salud	185	93.9%	6	3.04%	6	3.04%	-	-
	Relaciones	157	82.6%	4	2.2%	29	15.2%	-	-
	Informática	200	93.45 %	10	4.67%	4	1.86%	-	-
NO NT	Salud	88	75.21 %	12	10.25	17	14.52%	1	1.17%
	Relaciones	62	77.5%	10	12.5%	8	10%	-	-
	Informática	68	80.9%	5	5.95%	10	11.90%	1	1.19%

Tabla 12: Análisis de las categorías gramaticales

4.2.3. Comparativa y evaluación de las unidades léxicas con el PCIC

En la tabla 13 se encuentra la distribución atendiendo a los diferentes niveles del MCER de las unidades léxicas obtenidos de los informantes no nativos, en concreto se trata de 75 UL. Para ello he utilizado el inventario de nociones específicas y la medida de corte a seguir es la utilización de esas UL por al menos dos informantes debido a que obtenemos más muestras si lo realizamos con aquellas UL con un índice mayor o igual a 0,1. No todas las palabras aparecen en el inventario exacto que les pertenecer, sino que al ser categorías más abierta establecen relaciones de otros inventarios, como ocurre con *cuerpo*, o *deporte*.

Las unidades léxicas producidas por los informantes que tienen mayor representación en el PCIC son aquellas que pertenecen a *Relaciones familiares y sociales*, a medida que el nivel es más elevado, las unidades léxicas disminuyen. Es llamativo como casi todas las palabras producidas por los informantes no nativos aparecen en el PCIC, y además se corresponden con niveles inferiores a los que están, puesto que la mayoría de los informantes se encuentran en B1 y B2.

Para comprobar la eficacia de los índices de disponibilidad nos basaremos en el centro de *Salud*, ya que encontramos déficits importantes en relación con el léxico y su nivelación significativos. Encontramos palabras como *sano*, *fiebre*, o *vacuna* con un gran ID, pero no aparecen hasta niveles como B1 o incluso B2 en el caso de *fiebre*. Por esta razón, es importante destacar las pruebas de disponibilidad léxica y con ello los índices de disponibilidad como una herramienta complementaria para la evaluación y nivelación del léxico, pues el PCIC podría servirse de sus listas para aportar nuevas palabras. Otros ejemplos pertenecientes a otro centro de interés, pero no menos significativos, es la incorporación de palabras *redes sociales*, *pagar con el móvil*, es decir, una actualización de las unidades léxicas que encontramos en él pues cada vez se usan más, y no aparecen hasta niveles superiores, como por ejemplo en el C1 aparece, *buscar en Internet* (9.6 *Internet*), *aplicación o actualizar una página web* (16.4 *Informática y nuevas tecnologías*) palabras más que extendidas en el día a día. Por último, también es necesaria una actualización o mejor dicho, incorporación de extranjerismos, especialmente los relacionados con *Informática y nuevas tecnologías*.

	Salud	Relaciones familiares y sociales	Informática y nuevas tecnologías
A1	Hospital, enfermero, médico, enfermo, deporte, verdura, farmacia	Padre, madre, abuelos, hermanos, hijos, tíos, padres, familia, sobrino, primos, novio, amigo, jefe, pareja, profesor. Mujer	Internet, ordenador, móvil, teléfono, televisión
A2	Tener una+ enfermedad, tener + dolor, cuerpo, tomar+ una medicina, tener + tos, tener+ dolor de cabeza, ambulancia.		Portátil, navegar+ por Internet
B1	Estar+ sano, vacuna, doctor, tener+ cáncer	Amistad, tener+ una relación de amor, matrimonio	Ratón, descargarse + programa, videojuego
B2	Fiebre, coger/contagiar+ virus, estar + a dieta, ataque al corazón	Yerno, nuera, cuñado, suegro	Inteligencia artificial, enlace
C1	Saludable, dar/ poner + mascarilla	Parientes	
C2			
No aparecen	COVID, operación	Mejores amigos	Computadora, 5G, redes sociales, pagar con el móvil, Ipad, Tablet, WeChat

Tabla 13. Distribución de UL utilizada por al menos dos informantes del grupo de los no nativos según los niveles de referencia del MCER.

Tras esto hemos podido concluir que la categoría más utilizada por todos los informantes, tanto nativos como no nativos, es el sustantivo y que de introducir otras categorías serán los estudiantes de ELE quienes incorporan más palabras diferentes al sustantivo. Al igual que en caso anterior predominan las formas uniberbales (sobre todo el sustantivo), pero cuando se introduce los informantes no nativos son los que más producen y con grandes diferencias. Son frecuente los extranjerismos en ambos grupos

especialmente en *Informática y Nuevas tecnologías*. Por último, los casos de variación diatópica son muy escasos, lo que puede demostrar un especial apego a la lengua estándar.

En definitiva, este último apartado ha estado dedicado al análisis detallado de los elementos que configuran internamente las diferentes categorías analizadas en este trabajo, es decir, al análisis de las unidades léxicas con un índice superior o igual a 0,1; el análisis de las categorías gramaticales y análisis de las UL según su aparición y nivelación en el PCIC.

5. Conclusiones

En este último apartado se ofrece una síntesis de los resultados y conclusiones más relevantes obtenidas tras las distintas revisiones teóricas y tras el análisis de los resultados de nuestra investigación. Se partirá de las hipótesis formuladas al comienzo de este trabajo, con el fin de comprobar si han quedado corroboradas en nuestro estudio. La investigación está formada por 40 estudiantes universitarios, 20 nativos y 20 estudiantes sinohablantes de ELE. Una vez analizados los resultados se puede comprobar:

- El índice de disponibilidad léxica es una herramienta adecuada para evaluar el léxico. Se ha podido constatar que alumnos nativos y no nativos producen palabras que en su mayoría aparecen en el PCIC, como es el caso de *Relaciones familiares y sociales*, aparecen todas las palabras salvo una. Por el contrario, hay centros en los que muchas de sus palabras no aparecen, como es el caso de *Informática*, en concreto, son cinco las palabras que no aparecen en el PCIC, por lo que sería conveniente una revisión y actualización de sus términos, especialmente relacionados con las nuevas tecnologías. Además, se ha podido observar que en niveles inferiores son más las palabras que aparecen, en cambio según aumenta el nivel, disminuye la cantidad de palabras producidas.
- No todo el vocabulario que aprenden los estudiantes se corresponde con el que realmente emplean los nativos. Esta fue la primera hipótesis en ser comprobada, a través de los diferentes listados de palabras se puede ver sus diferencias.
- Los estudiantes no reaccionan por igual en todos los centros de interés, pues depende de su nivel de estudio o necesidades. Esta hipótesis también es cierta; en concreto, se puede ver que los estudiantes nativos producen más palabras en *Informática*, seguido de *Relaciones* y por último *Salud*, con un total de 344, 333 y 310 palabras respectivamente. Por el contrario, los alumnos no nativos producen más palabras en *Relaciones* (208 palabras), seguido de *Salud* (181 palabras) y por último *Informática* (136 palabras).
- Un mayor número de unidades léxicas se puede traducir en un mayor dominio de la competencia lingüística. Esta hipótesis no es cierta, ya que se puede diferenciar entre productividad léxica (relacionada con el total de palabras que han aparecido) o riqueza léxica (relacionado con el número de palabras no repetidas, también llamado vocablos).

Una vez comprobada las diferentes hipótesis, se detallan los aspectos más relevantes de cada punto. En el marco teórico se realiza una aproximación al concepto de competencia léxica, desde dos visiones, a través del MCER y el PCIC, para después continuar con el tratamiento del léxico según los distintos enfoques metodológicos. En el siguiente apartado se exponen los criterios para la selección léxica en la enseñanza de lenguas, los cuales se pueden resumir en: selección subjetiva; selección objetiva selección subjetiva-correcta; se concluyó que un único criterio no parece el proceder más adecuado para la selección del léxico, sino la conjugación de varios es el proceder más adecuado. Después se profundiza en la disponibilidad léxica y al acercamiento teórico del concepto, acercándonos al origen de los estudios de disponibilidad y la disponibilidad léxica como instrumento de evaluación para la enseñanza de español como lengua extranjera. Por último, se explican las principales aportaciones y líneas de investigación.

Después de este punto estrictamente teórico, los dos restantes se centran en la investigación llevado a cabo para este trabajo, en la obtención de los índices de disponibilidad léxica de los tres centros de interés con 40 informantes y su posterior análisis. Por lo tanto, el punto 3 se ha dedicado a la metodología de la investigación basado en los trabajos del *Proyecto Panhispánico* y a la descripción de los informantes. En lo que respecta a la edición de los datos recogidos, se ha seguido los propuestos por el “Proyecto”, además de los criterios empleados en otras investigaciones con informantes no nativos, como Sánchez-Saus.

Respecto a la investigación y en relación con los resultados cuantitativos, en concreto atendiendo a los totales de palabras y vocablos, se puede comprobar que, tanto en un parámetro como en otro, los informantes nativos superan ampliamente a los estudiantes de ELE, por lo que la productividad léxica (relacionado con el total de respuestas por categorías) como la riqueza léxica (relacionada por el número de vocablos o palabras disponibles actualizadas) es mayor en el grupo de los nativos. En este sentido, las hipótesis que manejábamos en la introducción se ven comprobados por los datos obtenidos.

Atendiendo a los cuantitativo de los resultados, concretamente, sobre el total de palabras, el grupo de los nativos cuenta con 462 respuestas más (987 palabras para los nativos frente a 525 palabras de los estudiantes de ELE), lo que supone una diferencia de 154 palabras en el promedio de palabras por cada categoría. En cuanto al número de

vocablos en total, hay una diferencia de 319 vocablos entre ambos grupos, en concreto los nativos ofrecen 600 vocablos en total frente a los 281 de los estudiantes de ELE. Las diferencias no parecen reducirse al evaluar el comportamiento individual de los informantes, debido a que los nativos tienen una media 16.45 palabras, frente a 8.75 de los estudiantes de ELE, lo que supone una diferencia de 7,7 palabras. Es importante recordar en este punto que, como bien argumenta Hernández Muñoz (2006, p. 320), no siempre un caudal léxico mayor conlleva un conocimiento superior de vocabulario, sino que las diferencias que evalúa la prueba de disponibilidad léxica pueden radicar simplemente en una mayor habilidad para evocar palabras. Según el índice de cohesión podemos clasificar las áreas temáticas en: centros de cohesión alto, o también nombrados a lo largo del presente trabajo como, centros muy compactos o cerrados, en este caso hablamos de *Relaciones* para ambos grupos. Por el contrario, centros muy dispersos o difusos, que hacen referencia a *Salud*, también para ambos grupos, con esto podemos concluir que, a pesar de las diferencias entre la productividad y riqueza, la cohesión interna en presenta gran similitud en ambos grupos, y las diferencias cuando ocurre parecen poco significativas.

Atendiendo al análisis cualitativo, en que se analizan las palabras evocadas con un índice superior o igual a 0,1, las categorías gramaticales predominantes y por último, la clasificación de las palabras producidas por al menos 2 informantes según los niveles del MCER. Tras este estudio hemos podido encontrar parámetros en que los resultados se invierten y son los alumnos de ELE quién obtienen resultados mayores, es decir, la categoría más utilizada por todos los informantes, tanto nativos como no nativos, es el sustantivo y que de introducir otras categorías serán los estudiantes de ELE quienes incorporan más palabras diferentes al sustantivo (89.98% de sustantivos frente a 77.87% de sustantivo, dejando el resto a otras categorías gramaticales). Al igual que en caso anterior predominan las formas uniberbales (sobre todo el sustantivo), pero cuando se introduce los informantes no nativos son los que más producen y con grandes diferencias. Es llamativo como casi todas las palabras producidas por los informantes no nativos aparecen en el PCIC, y además se corresponden con niveles inferiores a los que están, puesto que la mayoría de los informantes se encuentran en B1 y B2.

Por último, en los anexos se encuentra todo el material recogido durante este estudio, ordenado por centro de interés, grupo y según el índice de disponibilidad. Esto podrá utilizarse como una herramienta para la optimalización del léxico de una lengua, es decir,

utilizarlo como una herramienta previa a la elaboración de metodologías o estrategias didácticas.

Diferentes limitaciones han estado presentes a lo largo de la investigación. En primer lugar, realizar la encuesta de forma online no permite supervisar si todos los informantes están realizándolo de manera correcta, es decir, respetando los dos minutos, sin utilizar ningún apoyo, etc. La herramienta por la que se administraba la encuesta a los informantes no nativos también supuso un debate, debido a que numerosos informantes sinohablantes no podían tener acceso a ningún servicio proporcionado por Google, y a pesar de buscar diferentes herramientas, se optó por utilizar sus propias aplicaciones, como WeChat. Gracias a esta herramienta nos asegurábamos el acceso de todos los estudiantes; sin embargo, no podíamos controlar el tiempo exacto que tardaban en escribir todos los vocablos.

Como futuras líneas de investigación, propongo aplicar la disponibilidad léxica de forma real en el aula y poder influir directamente en el proceso de enseñanza- aprendizaje. Para lograr esto será necesario incluir la disponibilidad léxica en los programas y materiales ELE, pues esto dará como resultado unos materiales y cursos más adecuados a los aprendices, que son, en definitiva, el elemento esencial. En general, el índice de disponibilidad léxica es una herramienta que puede ser explotada para distintas finalidades.

6. Referencia bibliográficas.

- Baralo, M. (2005). La competencia léxica en el Marco Común Europeo de Referencia. *Carabela*, 58, 27-48. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/carabela/pdf/58/58_027.pdf
- Bartol, J. (2010). Disponibilidad léxica y selección del vocabulario. 85-107. Recuperado de: <https://ifc.dpz.es/recursos/publicaciones/29/95/09bartol.pdf>
- Cervantes, I (2001). Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación, Madrid: MECD- Anaya.
- Cervantes, I. (2006). Plan curricular del Instituto Cervantes, Madrid: Instituto Cervantes y Biblioteca Nueva.
- Gómez, B. (2004). La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones. Tesis doctoral, Valencia: Universidad de Valencia. Departamento de Filología Hispánica. Recuperado de: <https://roderic.uv.es/bitstream/handle/10550/15248/gomez.pdf?sequence=1>
- Gómez, J. (2000). La competencia léxica en la enseñanza de español como lengua extranjera. *Mosaico*, 5, 23-29. Disponible en: https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=13139_19
- Grethel, L. & Pérez, C. (2018). La selección léxico en el español como lengua extranjera. *Maestro y Sociedad*, 15 (1). Recuperado de: <https://maestroysociedad.uo.edu.cu/index.php/MyS/article/view/3287>
- Hernández, N (2006). Hacia una teoría integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellanos-manchegos. Salamanca: ediciones de la Universidad de Salamanca.
- Hidalgo, M (2017). La disponibilidad léxica como herramienta de evaluación interventiva del vocabulario en ELE: aplicación en un contexto universitario sinohablante. *SinoELE*, 16, 43-63. Recuperado de: https://www.sinoele.org/images/Revista/16/Articulos/Hidalgo_SinoELE_16_2017_43-63.pdf

- Hidalgo, M. (2019). Factores y fuentes que inciden en el léxico disponible de estudiantes sinohablantes de ELE, *Asele*, 23. Recuperado de: https://www.academia.edu/41156817/Factores_y_fuentes_que_inciden_en_el_lexico_disponible_de_estudiantes_sinohablantes_de_ELE
- Higueras, M. (2012). Claves prácticas para la enseñanza del léxico. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 6. (11). Recuperado de: <https://revistas.nebrija.com/revista-linguistica/article/view/183>
- Izquierdo, G. (2003). La selección del léxico en la enseñanza del español como lengua extranjera. Tesis doctoral, Valencia: Universidad de Valencia. Departamento de Filología Española. Recuperado de: <https://www.tdx.cat/handle/10803/9815?show=full>
- López- Mezquita, M. (2007). La evaluación de la competencia léxica: test de vocabulario, su fiabilidad y validez. Tesis doctoral, Granada: Universidad de Granada. Departamento de Filología Inglesa y Alemana. Recuperado de: <https://digibug.ugr.es/handle/10481/4571>
- López, E. (2008). Estudio de disponibilidad léxica en 43 estudiantes de ELE. Memoria de máster, Madrid: Universidad Antonio Nebrija. Departamento de Lenguas Aplicadas. Recuperado de: <http://www.educacionyfp.gob.es/dam/jcr:15c7f59f-8d8d-4a93-8313-e8b5d84ff3a7/2008-bv-09-16lopez-rivero-pdf.pdf>
- Martín, M. (2010). Aproximación histórica a la enseñanza de lenguas extranjeras. *Aula*, 16, 137-154. Recuperado de: <https://revistas.usal.es/index.php/0214-3402/article/view/7437/8480>
- Moreno, J.A. y Madrona, A. (2020). Los índices de disponibilidad léxica como criterio complementario para la selección del lecionario de un diccionario de ELE. *Universitat Rovira i Virgili*.
- Núñez, L. (2019) Metodologías para la enseñanza del léxico en el aprendizaje de lenguas extranjeras: un recorrido holístico. *Foro de profesores de E/LE*, 15. Recuperado de: <https://ojs.uv.es/index.php/foroele/article/view/15626/14511>
- Palapanidi, K. (2012). La aplicación de la disponibilidad léxica a la didáctica del léxico de LE. *Revista Nebrija de Lingüística Aplicada*, 11. Recuperado de: [La aplicación de la disponibilidad léxica a la didáctica del léxico de LE \(nebrija.com\)](https://www.nebrija.com/la-aplicacion-de-la-disponibilidad-lexica-a-la-didactica-del-lexico-de-le)

- Paredes, F. (2015). Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula. *Linred*. Recuperado de: http://www.linred.es/numero13_monografico_Art2.html
- Pérez, M. (2015). Tratamiento de la combinatoria léxica en documentos de referencia y curriculares: el caso del MCER y del PCIC. *Revista de Investigación Lingüística*, 18, 213-232. Recuperado de: <https://revistas.um.es/ril/article/view/248071/188831>
- Pérez, M. (2015). Un enfoque léxico a prueba: efectos de la instrucción en el aprendizaje de las colocaciones léxicas. Tesis doctoral, Salamanca: Universidad de Salamanca. Departamento de lengua española. Recuperado de: <https://gredos.usal.es/handle/10366/128255>
- Salazar, V. (2004). Acercamiento crítico a la selección objetiva de contenidos léxicos en la enseñanza de E/LE. *ELUA*, 18, 243-273. Recuperado de: https://rua.ua.es/dspace/bitstream/10045/6140/1/ELUA_18_13.pdf
- Samper, M. (2002). Disponibilidad léxica en alumnos de español como lengua extranjera. *ASELE*, Colección monografías, 4.
- Sánchez, M. (2010). Aproximación histórica a la enseñanza de lenguas extranjeras. *Aula*, 16, 137-154. Recuperado de: <https://revistas.usal.es/index.php/0214-3402/article/download/7437/8480/>
- Sánchez- Saus, M (2011). Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas. Tesis doctoral, Cádiz: Universidad de Cádiz. Departamento de Filología. Recuperado de: <https://rodin.uca.es/xmlui/handle/10498/15862>
- Sanz, I. (2000). El léxico en la enseñanza de ELE: una aplicación práctica para el nivel intermedio. *Centro Virtual Cervantes*, 635-645. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/11/11_0635.pdf
- Saralegui, C & Tabernero, C. (2008). Aportación al proyecto panhispánico de léxico disponible: Navarra. *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)*, Departamento de Lingüística hispánica y Lenguas modernas. Pamplona, Servicio de Publicaciones de la Universidad de Navarra. Recuperado de: <http://www.unav.es/linguis/simposiosel/act>

Serrano, M (2014). Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto. Tesis doctoral, Lleida: Universitat de Lleida. Departamento de Filología Clásica, francesa e Hispánica. Recuperado de: <https://repositori.udl.cat/handle/10459.1/64151>

Tomé Cornejo, C. (2015). Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE. Tesis doctoral, Salamanca: Universidad de Salamanca.

Vidiella, M. (2011). El enfoque léxico en la didáctica de ELE: la puesta en práctica. Trabajo fin de grado, Barcelona: Universidad de Barcelona. Recuperado de: http://diposit.ub.edu/dspace/bitstream/2445/131178/1/tfm_merce_vidiella_andreu.pdf

7. Anexos

Anexo 1. Documento principal para la explicación de las encuestas.

¡Hola a todos! Muchas gracias por participar en esta encuesta.

Soy Blanca Corral y estoy estudiando un Máster en Enseñanza de Lenguas Extranjeras. Español como lengua extranjera, y esta encuesta se encuentra dentro de mi investigación para el TFM. Estudiaré la disponibilidad léxica de estudiantes nativos y no nativos.

Esta prueba se dividirá en dos partes, por un lado, la recopilación de datos sociolingüísticos y por otro lado, 3 encuestas de disponibilidad léxica relacionadas con tres temas concretos (Relaciones familiares y sociales; Salud; Internet y digitalización). Para la realización de estas encuestas habrá un tiempo determinado de 2 min. Esta primera encuesta está relacionado con los datos sociolingüísticos para poder facilitar el proceso estadístico. Los datos son anónimos y se incorporarán a una base de datos. Tras completar los datos sociolingüísticos, se pasará a realizar las encuestas de disponibilidad léxica. Es necesario hacer todas las encuestas en el mismo momento, no llevará más de 2 minutos cada encuesta. **Se anotarán todos los términos que vengan a la mente en relación con el área de interés (se podrá escribir cualquier término, pues no importa si es un verbo, adjetivo, sustantivo o expresiones pluriverbales).**

ES MUY IMPORTANTE QUE INTENTÉIS REALIZARLO EN DOS MINUTOS

Los datos que se facilitan se tratarán de forma anónima y se incorporarán a una base de datos para su tratamiento estadístico

Enlaces a las encuestas:

Datos sociolingüísticos: <https://es.surveymonkey.com/r/JWG9YFT>

Salud: <https://es.surveymonkey.com/r/JTKWJVV>

Relaciones personales y sociales: <https://es.surveymonkey.com/r/JMKF5RN>

Informática y Nuevas Tecnologías: <https://es.surveymonkey.com/r/JMPNHGM>

Anexo 2. Índices de disponibilidad léxica

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
Médico	0,5999102	13	64,99999762	0,041935484
Enfermero	0,49031895	11	55,00000119	0,077419355
Hospital	0,4298583	10	50	0,109677419
Vacuna	0,33427423	8	40,0000006	0,135483871
enfermedad	0,31300217	7	34,9999994	0,158064516
Medicina	0,26306528	6	30,00000119	0,177419355
centro de salud	0,17613271	5	25	0,193548387
Cáncer	0,14256755	4	20,0000003	0,206451613
termómetro	0,1306881	4	20,0000003	0,219354839
Farmacia	0,12084819	3	15,0000006	0,229032258
Jeringuilla	0,11931143	3	15,0000006	0,238709677
Diabetes	0,118641	3	15,0000006	0,248387097
COVID	0,11506802	3	15,0000006	0,258064516
Jarabe	0,10905311	3	15,0000006	0,267741935
Venda	0,10857002	3	15,0000006	0,277419355
Deporte	0,1057172	3	15,0000006	0,287096774
Operación	0,09677843	4	20,0000003	0,3
Herida	0,09605844	3	15,0000006	0,309677419
Quirófano	0,09238607	3	15,0000006	0,319354839
Saludable	0,09153937	3	15,0000006	0,329032258
medicamento	0,08972668	2	10,00000015	0,335483871
radiografía	0,08914873	3	15,0000006	0,34516129
Camilla	0,08680581	3	15,0000006	0,35483871
apendicitis	0,08392195	3	15,0000006	0,364516129
Cuerpo	0,08343232	2	10,00000015	0,370967742
medicamentos	0,08261712	2	10,00000015	0,377419355
Tirita	0,08207513	2	10,00000015	0,383870968
Diarrea	0,08207513	2	10,00000015	0,390322581
coronavirus	0,0818046	2	10,00000015	0,396774194
Dolor	0,08156418	2	10,00000015	0,403225806
Celador	0,07945336	2	10,00000015	0,409677419
fármaco	0,07945336	2	10,00000015	0,416129032
colesterol	0,07723351	2	10,00000015	0,422580645
ambulancia	0,07187835	2	10,00000015	0,429032258
sala de espera	0,07082204	2	10,00000015	0,435483871
nutrición	0,06864101	2	10,00000015	0,441935484
urgencias	0,06443231	2	10,00000015	0,448387097
ecografía	0,06354626	2	10,00000015	0,45483871
jaqueca	0,06318319	2	10,00000015	0,461290323
pastilla	0,06156756	2	10,00000015	0,467741935
especialista	0,05908847	2	10,00000015	0,474193548
cuidarse	0,05664299	2	10,00000015	0,480645161

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

pulsómetro	0,05526815	2	10,00000015	0,487096774
análisis	0,05321405	2	10,00000015	0,493548387
Pierna	0,05	1	5,000000075	0,496774194
Enfermera	0,0472061	1	5,000000075	0,5
Belleza	0,0472061	1	5,000000075	0,503225806
Higiene	0,0472061	1	5,000000075	0,506451613
Cura	0,0472061	1	5,000000075	0,509677419
Laboratorio	0,04648905	2	10,00000015	0,516129032
Oftalmólogo	0,04537563	2	10,00000015	0,522580645
Pastillas	0,04456831	1	5,000000075	0,525806452
resfriado	0,04456831	1	5,000000075	0,529032258
Medico	0,04456831	1	5,000000075	0,532258065
pediatra	0,04456831	1	5,000000075	0,535483871
estómago	0,04456831	1	5,000000075	0,538709677
bienestar	0,04312385	2	10,00000015	0,54516129
técnico de rayos	0,04207791	1	5,000000075	0,548387097
dolor de cabeza	0,04207791	1	5,000000075	0,551612903
biberón	0,04207791	1	5,000000075	0,55483871
lactancia	0,04207791	1	5,000000075	0,558064516
corazón	0,04207791	1	5,000000075	0,561290323
sanitario	0,04207791	1	5,000000075	0,564516129
VIH	0,03972668	1	5,000000075	0,567741935
Cita	0,03972668	1	5,000000075	0,570967742
maternidad	0,03972668	1	5,000000075	0,574193548
Agua	0,03972668	1	5,000000075	0,577419355
abductores	0,03972668	1	5,000000075	0,580645161
anestesia	0,03914202	2	10,00000015	0,587096774
Familia	0,03778941	2	10,00000015	0,593548387
Salud	0,03750683	1	5,000000075	0,596774194
inyección	0,03750683	1	5,000000075	0,6
Dolencia	0,03750683	1	5,000000075	0,603225806
Aspirinas	0,03750683	1	5,000000075	0,606451613
Sano	0,03750683	1	5,000000075	0,609677419
nutricionista	0,03750683	1	5,000000075	0,612903226
cardiopatía	0,03541102	1	5,000000075	0,616129032
Alta	0,03541102	1	5,000000075	0,619354839
Eutanasia	0,03541102	1	5,000000075	0,622580645
Urgencia	0,03541102	1	5,000000075	0,625806452
confinamiento	0,03541102	1	5,000000075	0,629032258
Vacunas	0,03541102	1	5,000000075	0,632258065
alimentación	0,03541102	1	5,000000075	0,635483871
Garganta	0,03541102	1	5,000000075	0,638709677
celador.	0,03343232	1	5,000000075	0,641935484
Aneurisma	0,03343232	1	5,000000075	0,64516129
antibiótico	0,03343232	1	5,000000075	0,648387097
MIR	0,03343232	1	5,000000075	0,651612903

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

ingreso	0,03343232	1	5,000000075	0,65483871
centro médico	0,03343232	1	5,000000075	0,658064516
operación quirúrgica	0,03343232	1	5,000000075	0,661290323
nacimiento	0,03343232	1	5,000000075	0,664516129
cabeza	0,03343232	1	5,000000075	0,667741935
ejercicio	0,03343232	1	5,000000075	0,670967742
gastroenteritis	0,03156418	1	5,000000075	0,674193548
neumotorax	0,03156418	1	5,000000075	0,677419355
Héroes	0,03156418	1	5,000000075	0,680645161
Virus	0,03156418	1	5,000000075	0,683870968
Analítica	0,03156418	1	5,000000075	0,687096774
Físico	0,03156418	1	5,000000075	0,690322581
Pubertad	0,03156418	1	5,000000075	0,693548387
Muelas	0,03156418	1	5,000000075	0,696774194
Tratamiento	0,02980044	1	5,000000075	0,7
Dentista	0,02980044	1	5,000000075	0,703225806
Aguja	0,02980044	1	5,000000075	0,706451613
Sacrificio	0,02980044	1	5,000000075	0,709677419
Bacteria	0,02980044	1	5,000000075	0,712903226
Insulina	0,02980044	1	5,000000075	0,716129032
actividad física	0,02980044	1	5,000000075	0,719354839
Ducharse	0,02980044	1	5,000000075	0,722580645
Paracetamol	0,02980044	1	5,000000075	0,725806452
Muela	0,02813524	1	5,000000075	0,729032258
Blanco	0,02813524	1	5,000000075	0,732258065
estar enfermo	0,02813524	1	5,000000075	0,735483871
Cirugía	0,02813524	1	5,000000075	0,738709677
Patología	0,02813524	1	5,000000075	0,741935484
Tensión	0,02813524	1	5,000000075	0,74516129
Hidratación	0,02813524	1	5,000000075	0,748387097
Testículo	0,02813524	1	5,000000075	0,751612903
Receta	0,0265631	1	5,000000075	0,75483871
Pasillo	0,0265631	1	5,000000075	0,758064516
Migraña	0,0265631	1	5,000000075	0,761290323
Intervención operación	0,0265631	1	5,000000075	0,764516129
Ciencia	0,0265631	1	5,000000075	0,767741935
Nutrientes	0,0265631	1	5,000000075	0,770967742
Pan	0,0265631	1	5,000000075	0,774193548
cáncer de testículo	0,0265631	1	5,000000075	0,777419355
Uretra	0,0250788	1	5,000000075	0,780645161
Secretaria	0,0250788	1	5,000000075	0,783870968
Reposo	0,0250788	1	5,000000075	0,787096774
microbiología	0,0250788	1	5,000000075	0,790322581
Glucosa	0,0250788	1	5,000000075	0,793548387
Vitaminas	0,0250788	1	5,000000075	0,796774194
cáncer de páncreas	0,0250788	1	5,000000075	0,8

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Clínica	0,0250788	1	5,000000075	0,803225806
Ovarios	0,02367745	1	5,000000075	0,806451613
rayos x	0,02367745	1	5,000000075	0,809677419
Paciente	0,02367745	1	5,000000075	0,812903226
Fiebre	0,02367745	1	5,000000075	0,816129032
Minerales	0,02367745	1	5,000000075	0,819354839
Liposucción	0,02367745	1	5,000000075	0,822580645
bisturí	0,0223544	1	5,000000075	0,825806452
Útero	0,0223544	1	5,000000075	0,829032258
salud familiar	0,0223544	1	5,000000075	0,832258065
médico de familia	0,0223544	1	5,000000075	0,835483871
prácticas en hospital	0,0223544	1	5,000000075	0,838709677
prescripción médica	0,0223544	1	5,000000075	0,841935484
entusiasmo	0,0223544	1	5,000000075	0,84516129
mamografía	0,02110527	1	5,000000075	0,848387097
salud comunitaria	0,02110527	1	5,000000075	0,851612903
sangrar	0,02110527	1	5,000000075	0,85483871
atención primaria	0,02110527	1	5,000000075	0,858064516
consulta	0,02110527	1	5,000000075	0,861290323
glóbulos rojos	0,02110527	1	5,000000075	0,864516129
Ánimo	0,02110527	1	5,000000075	0,867741935
derechos	0,01992595	1	5,000000075	0,870967742
guardias	0,01992595	1	5,000000075	0,874193548
cirujano	0,01992595	1	5,000000075	0,877419355
glóbulos blancos	0,01992595	1	5,000000075	0,880645161
equilibrio	0,01992595	1	5,000000075	0,883870968
tobillo	0,01992595	1	5,000000075	0,887096774
Ojos	0,01881253	1	5,000000075	0,890322581
gratuidad	0,01881253	1	5,000000075	0,893548387
escayola	0,01881253	1	5,000000075	0,896774194
balance	0,01881253	1	5,000000075	0,9
hombros	0,01881253	1	5,000000075	0,903225806
psicóloga	0,01776132	1	5,000000075	0,906451613
digestivo	0,01776132	1	5,000000075	0,909677419
Meditación	0,01776132	1	5,000000075	0,912903226
Kinesiólogo	0,01676885	1	5,000000075	0,916129032
Endocrinología	0,01676885	1	5,000000075	0,919354839
Proteína	0,01676885	1	5,000000075	0,922580645
Prevención	0,01583184	1	5,000000075	0,925806452
Anatomía	0,01583184	1	5,000000075	0,929032258
Calcio	0,01583184	1	5,000000075	0,932258065
Vida	0,01494719	1	5,000000075	0,935483871
Pabellón	0,01411196	1	5,000000075	0,938709677
Quirúrgico	0,01332341	1	5,000000075	0,941935484
Plenitud	0,01257893	1	5,000000075	0,94516129
Reanimación	0,01187604	1	5,000000075	0,948387097

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Felicidad	0,01187604	1	5,000000075	0,951612903
Calma	0,01121243	1	5,000000075	0,95483871
Paz	0,0105859	1	5,000000075	0,958064516
Estrés	0,00999438	1	5,000000075	0,961290323
Vitalidad	0,00943591	1	5,000000075	0,964516129
Energía	0,00890865	1	5,000000075	0,967741935
Enérgico	0,00841085	1	5,000000075	0,970967742
Vital	0,00794087	1	5,000000075	0,974193548
Activo	0,00749715	1	5,000000075	0,977419355
Animado	0,00707822	1	5,000000075	0,980645161
Alma	0,00668271	1	5,000000075	0,983870968
Comer	0,00630929	1	5,000000075	0,987096774
Comida	0,00595674	1	5,000000075	0,990322581
Beber	0,00562389	1	5,000000075	0,993548387
Bebida	0,00530963	1	5,000000075	0,996774194

Tabla 14. Índice de disponibilidad léxica centro de interés Salud, grupo nativo

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
Amigos	0,41330367	12	60,00000238	0,036036036
Hermanos	0,32712802	8	40,0000006	0,06006006
Amistad	0,32371157	9	44,99999881	0,087087087
Abuelos	0,31399727	9	44,99999881	0,114114114
Padres	0,27084678	7	34,9999994	0,135135135
Primos	0,26748309	7	34,9999994	0,156156156
Hijos	0,25378013	7	34,9999994	0,177177177
Tíos	0,24526665	7	34,9999994	0,198198198
Amor	0,22590092	6	30,00000119	0,216216216
Confianza	0,16488248	5	25	0,231231231
Sobrinos	0,16159435	5	25	0,246246246
Familia	0,16151519	4	20,0000003	0,258258258
Cariño	0,15863459	4	20,0000003	0,27027027
Unión	0,15305945	6	30,00000119	0,288288288
Suegros	0,14839663	5	25	0,303303303
Madre	0,14727235	3	15,0000006	0,312312312
Padre	0,14454469	3	15,0000006	0,321321321
Compañeros	0,12582259	4	20,0000003	0,333333333
Colegas	0,11620265	5	25	0,348348348
Reír	0,11133559	3	15,0000006	0,357357357
Cuñados	0,10623236	3	15,0000006	0,366366366
Novios	0,10229844	3	15,0000006	0,375375375
Matrimonio	0,10161138	3	15,0000006	0,384384384
Vínculo	0,09850896	3	15,0000006	0,393393393
Pareja	0,09727235	2	10,00000015	0,399399399
Nietos	0,09274743	3	15,0000006	0,408408408

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Reunión	0,09103931	3	15,00000006	0,417417417
Lealtad	0,09064968	3	15,00000006	0,426426426
Respeto	0,08995016	2	10,00000015	0,432432432
Nuera	0,08785732	3	15,00000006	0,441441441
cercanía	0,0850431	2	10,00000015	0,447447447
Hablar	0,08376215	2	10,00000015	0,453453453
Conocidos	0,08116278	3	15,00000006	0,462462462
Sentimientos	0,08040373	2	10,00000015	0,468468468
Parientes	0,08002608	2	10,00000015	0,474474474
empatía	0,0785326	2	10,00000015	0,48048048
Disfrutar	0,07601746	2	10,00000015	0,486486486
Quedar	0,07487245	2	10,00000015	0,492492492
Grupo	0,07376869	3	15,00000006	0,501501502
Escuchar	0,07187047	2	10,00000015	0,507507508
Compromiso	0,07142049	2	10,00000015	0,513513514
Cena	0,07019792	2	10,00000015	0,51951952
Afectividad	0,06969107	2	10,00000015	0,525525526
Bisabuelos	0,06763055	2	10,00000015	0,531531532
Apoyo	0,06692623	2	10,00000015	0,537537538
Yerno	0,06406325	2	10,00000015	0,543543544
Hogar	0,06121468	2	10,00000015	0,54954955
Alegría	0,05982333	2	10,00000015	0,555555556
Relación	0,05850789	2	10,00000015	0,561561562
Infancia	0,05395212	2	10,00000015	0,567567568
Contacto	0,05347427	2	10,00000015	0,573573574
Complicidad	0,05	1	5,000000075	0,576576577
Integración	0,05	1	5,000000075	0,57957958
Empatizar	0,05	1	5,000000075	0,582582583
mejores amigos	0,04839435	2	10,00000015	0,588588589
Comida	0,04727235	1	5,000000075	0,591591592
Comunicación	0,04727235	1	5,000000075	0,594594595
Comprensión	0,04727235	1	5,000000075	0,597597598
Tolerancia	0,04727235	1	5,000000075	0,600600601
Conocer	0,04469349	1	5,000000075	0,603603604
Compartir	0,04469349	1	5,000000075	0,606606607
Interrelación	0,04469349	1	5,000000075	0,60960961
Hijo	0,04469349	1	5,000000075	0,612612613
Sociedad	0,04469349	1	5,000000075	0,615615616
Lazos	0,04469349	1	5,000000075	0,618618619
redes sociales	0,04469349	1	5,000000075	0,621621622
Querer	0,04225532	1	5,000000075	0,624624625
Risas	0,04225532	1	5,000000075	0,627627628
Salida	0,04225532	1	5,000000075	0,630630631
Novio	0,04225532	1	5,000000075	0,633633634
Madrastra	0,04225532	1	5,000000075	0,636636637
Herencia	0,04225532	1	5,000000075	0,63963964

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Videollamada	0,04225532	1	5,000000075	0,642642643
Divertido	0,03995016	1	5,000000075	0,645645646
Pariente	0,03995016	1	5,000000075	0,648648649
Bisabuelo	0,03995016	1	5,000000075	0,651651652
Marido	0,03995016	1	5,000000075	0,654654655
Parentesco	0,03995016	1	5,000000075	0,657657658
Padrastro	0,03995016	1	5,000000075	0,660660661
Fraternidad	0,03855632	2	10,00000015	0,666666667
Entretenimiento	0,03777076	1	5,000000075	0,66966967
Monótono	0,03777076	1	5,000000075	0,672672673
Ocio	0,03777076	1	5,000000075	0,675675676
Socialización	0,03777076	1	5,000000075	0,678678679
Mujer	0,03777076	1	5,000000075	0,681681682
Entender	0,03777076	1	5,000000075	0,684684685
Jefe	0,03655783	2	10,00000015	0,690690691
Protección	0,03571025	1	5,000000075	0,693693694
Padrinos	0,03571025	1	5,000000075	0,696696697
Vecino	0,03571025	1	5,000000075	0,6996997
Comprender	0,03571025	1	5,000000075	0,702702703
Felicidad	0,03376214	1	5,000000075	0,705705706
Frustración	0,03376214	1	5,000000075	0,708708709
Bautizo	0,03376214	1	5,000000075	0,711711712
Casados	0,03376214	1	5,000000075	0,714714715
Convivencia	0,03376214	1	5,000000075	0,717717718
Entorno	0,03376214	1	5,000000075	0,720720721
primos segundos	0,03376214	1	5,000000075	0,723723724
Ansiedad	0,03376214	1	5,000000075	0,726726727
Dialogar	0,03192031	1	5,000000075	0,732732733
Preocupación	0,03192031	1	5,000000075	0,735735736
Prometidos	0,03192031	1	5,000000075	0,738738739
Honor	0,03192031	1	5,000000075	0,741741742
Concuñados	0,03192031	1	5,000000075	0,744744745
Ideas	0,03192031	1	5,000000075	0,747747748
Agradecer	0,03017896	1	5,000000075	0,750750751
Besar	0,03017896	1	5,000000075	0,753753754
Diálogo	0,03017896	1	5,000000075	0,756756757
Celebraciones	0,03017896	1	5,000000075	0,75975976
Sinceridad	0,03017896	1	5,000000075	0,762762763
Tatarabuelos	0,03017896	1	5,000000075	0,765765766
Nieto	0,03017896	1	5,000000075	0,768768769
Predisposición	0,03017896	1	5,000000075	0,771771772
Red	0,03017896	1	5,000000075	0,774774775
Opiniones	0,03017896	1	5,000000075	0,777777778
Abrazar	0,0285326	1	5,000000075	0,780780781
Perdón	0,0285326	1	5,000000075	0,783783784
Encuentros	0,0285326	1	5,000000075	0,786786787

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Adolescencia	0,0285326	1	5,000000075	0,78978979
Divertirse	0,0285326	1	5,000000075	0,792792793
Esfuerzo	0,0285326	1	5,000000075	0,795795796
Proximidad	0,0285326	1	5,000000075	0,798798799
Cercano	0,0285326	1	5,000000075	0,801801802
dar la mano	0,02697606	1	5,000000075	0,804804805
Cenar	0,02697606	1	5,000000075	0,807807808
Constancia	0,02697606	1	5,000000075	0,810810811
Divorciado	0,02697606	1	5,000000075	0,813813814
Recuerdos	0,02697606	1	5,000000075	0,816816817
satisfacción.	0,02550443	1	5,000000075	0,81981982
Desahogarse	0,02550443	1	5,000000075	0,822822823
Mascota	0,02550443	1	5,000000075	0,825825826
Vejez	0,02550443	1	5,000000075	0,828828829
Conveniencia	0,02550443	1	5,000000075	0,831831832
Tapear	0,02550443	1	5,000000075	0,834834835
compañeros de piso	0,02550443	1	5,000000075	0,837837838
Cooperación	0,02550443	1	5,000000075	0,840840841
Fotografías	0,02550443	1	5,000000075	0,843843844
tía abuela	0,02550443	1	5,000000075	0,846846847
Sentir	0,02411309	1	5,000000075	0,84984985
Acuerdos	0,02411309	1	5,000000075	0,852852853
quedar a comer	0,02411309	1	5,000000075	0,855855856
Jugar	0,02411309	1	5,000000075	0,858858859
Vivienda	0,02411309	1	5,000000075	0,861861862
Enemigos	0,02411309	1	5,000000075	0,864864865
Empatía	0,02411309	1	5,000000075	0,867867868
tío abuelo	0,02411309	1	5,000000075	0,870870871
Trabajo	0,02411309	1	5,000000075	0,873873874
Quedadas	0,02279764	1	5,000000075	0,876876877
Incondicional	0,02279764	1	5,000000075	0,87987988
vivir juntos	0,02279764	1	5,000000075	0,882882883
pasar tiempo juntos	0,02279764	1	5,000000075	0,885885886
Carisma	0,02279764	1	5,000000075	0,888888889
Filial	0,02279764	1	5,000000075	0,891891892
Tristeza	0,02279764	1	5,000000075	0,894894895
Hermanastros	0,02279764	1	5,000000075	0,897897898
Votar	0,02279764	1	5,000000075	0,900900901
Superficial	0,02155396	1	5,000000075	0,903903904
Charlar	0,02155396	1	5,000000075	0,906906907
Vacaciones	0,02155396	1	5,000000075	0,90990991
Asertivo	0,02155396	1	5,000000075	0,912912913
grado de consanguinidad	0,02155396	1	5,000000075	0,915915916
Disciplina	0,02155396	1	5,000000075	0,918918919
pareja de hecho	0,02037813	1	5,000000075	0,921921922
Dar	0,02037813	1	5,000000075	0,924924925

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Educación	0,02037813	1	5,000000075	0,927927928
amigos cercanos	0,01926644	1	5,000000075	0,930930931
lengua de signos	0,01926644	1	5,000000075	0,933933934
compañero de clase	0,01628219	1	5,000000075	0,936936937
compañero de trabajo	0,01455416	1	5,000000075	0,93993994
Amantes	0,01455416	1	5,000000075	0,942942943
Socios	0,01376018	1	5,000000075	0,945945946
Subordinado	0,01300952	1	5,000000075	0,948948949
Asistente	0,01229981	1	5,000000075	0,951951952
Colaborador	0,01229981	1	5,000000075	0,954954955
primo de segundo grado	0,01162882	1	5,000000075	0,957957958
Colaboración	0,01162882	1	5,000000075	0,960960961
familia política	0,01099443	1	5,000000075	0,963963964
Secta	0,01099443	1	5,000000075	0,966966967
lazos sanguíneos	0,01039465	1	5,000000075	0,96996997
Clan	0,01039465	1	5,000000075	0,972972973
Agrupación	0,00929146	1	5,000000075	0,975975976
Conglomerado	0,00878459	1	5,000000075	0,978978979
Paternalismo	0,00830536	1	5,000000075	0,981981982
Patriarcado	0,00785228	1	5,000000075	0,984984985
Matriarcado	0,00742391	1	5,000000075	0,987987988
Materialismo	0,00701891	1	5,000000075	0,990990991
Asociarse	0,00627399	1	5,000000075	0,993993994
Vincularse	0,00593173	1	5,000000075	0,996996997
Unirse	0,00560813	1	5,000000075	1

Tabla 15. índices de disponibilidad léxica centro de interés Relaciones, grupo nativo

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
Ordenador	0,66376311	14	69,99999881	0,040697674
redes sociales	0,32823709	10	50	0,069767442
Wifi	0,31271586	9	44,99999881	0,095930233
Ratón	0,27429357	7	34,9999994	0,11627907
Teclado	0,25059706	7	34,9999994	0,136627907
Móvil	0,24296674	6	30,00000119	0,154069767
Aplicación	0,24023373	7	34,9999994	0,174418605
Web	0,21663828	7	34,9999994	0,194767442
Tablet	0,21122356	5	25	0,209302326
Internet	0,20919156	5	25	0,223837209
Software	0,18433514	5	25	0,238372093
Blog	0,17297831	6	30,00000119	0,255813953
Portátil	0,15635972	4	20,0000003	0,26744186
Facebook	0,14735369	5	25	0,281976744

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Red	0,14134337	4	20,0000003	0,293604651
Google	0,12813625	5	25	0,308139535
Pantalla	0,11418649	4	20,0000003	0,319767442
Conexión	0,10185039	3	15,0000006	0,328488372
Computadora	0,1	2	10,00000015	0,334302326
Teléfono	0,09397305	2	10,00000015	0,340116279
Apple	0,08724835	3	15,0000006	0,348837209
Programación	0,08616044	3	15,0000006	0,35755814
Whatsapp	0,08114201	3	15,0000006	0,36627907
disco duro	0,08048619	2	10,00000015	0,372093023
Instagram	0,07852646	3	15,0000006	0,380813953
Torre	0,07651353	2	10,00000015	0,386627907
Hardware	0,07456328	2	10,00000015	0,39244186
TIC	0,07371427	2	10,00000015	0,398255814
5G	0,0727554	3	15,0000006	0,406976744
Windows	0,0719021	2	10,00000015	0,412790698
Ipad	0,07100831	2	10,00000015	0,418604651
Android	0,06834713	2	10,00000015	0,424418605
Futuro	0,06786894	2	10,00000015	0,430232558
Webcam	0,06584634	2	10,00000015	0,436046512
memoria USB	0,06128783	2	10,00000015	0,441860465
Cableado	0,05966938	2	10,00000015	0,447674419
Televisor	0,05814847	2	10,00000015	0,453488372
Comunicación	0,05542954	2	10,00000015	0,459302326
Word	0,05050596	3	15,0000006	0,468023256
Hotmail	0,05	1	5,000000075	0,470930233
pantalla digital interactiva	0,05	1	5,000000075	0,473837209
mecanografía	0,05	1	5,000000075	0,476744186
Bulos	0,05	1	5,000000075	0,479651163
Router	0,04890228	2	10,00000015	0,485465116
Videollamada	0,04778767	2	10,00000015	0,49127907
correo electrónico	0,04698652	1	5,000000075	0,494186047
Aplicaciones	0,04698652	1	5,000000075	0,497093023
Avance	0,04698652	1	5,000000075	0,5
maker space	0,04698652	1	5,000000075	0,502906977
Ofimática	0,04698652	1	5,000000075	0,505813953
Ciberacoso	0,04698652	1	5,000000075	0,50872093
Cargador	0,0459993	2	10,00000015	0,514534884
Hacker	0,04595496	2	10,00000015	0,520348837
Cable	0,04491494	2	10,00000015	0,526162791
Youtube	0,04423271	2	10,00000015	0,531976744
banda ancha	0,04415467	1	5,000000075	0,534883721
Lápiz de memoria	0,04415467	1	5,000000075	0,537790698
Gamificación	0,04415467	1	5,000000075	0,540697674
IP	0,04415467	1	5,000000075	0,543604651
fibra óptica	0,04415467	1	5,000000075	0,546511628

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Estafas	0,04415467	1	5,000000075	0,549418605
Power Point	0,04220794	2	10,00000015	0,555232558
Conectarse	0,04149349	1	5,000000075	0,558139535
Tuenti	0,04149349	1	5,000000075	0,561046512
Macbook	0,04149349	1	5,000000075	0,563953488
escape room	0,04149349	1	5,000000075	0,566860465
telefonía móvil	0,04149349	1	5,000000075	0,569767442
Antivirus	0,04149349	1	5,000000075	0,572674419
Alexa	0,04149349	1	5,000000075	0,575581395
Virus	0,0389927	1	5,000000075	0,578488372
Interactivo	0,0389927	1	5,000000075	0,581395349
Iphone	0,0389927	1	5,000000075	0,584302326
Programa	0,0389927	1	5,000000075	0,587209302
ABP	0,0389927	1	5,000000075	0,590116279
Conectividad	0,0389927	1	5,000000075	0,593023256
Samsung	0,0389927	1	5,000000075	0,595930233
tele(visión)	0,0389927	1	5,000000075	0,598837209
Chatear	0,03664263	1	5,000000075	0,601744186
Inmediatez	0,03664263	1	5,000000075	0,604651163
Progreso	0,03664263	1	5,000000075	0,60755814
sistema informático	0,03664263	1	5,000000075	0,610465116
Proyector	0,03664263	1	5,000000075	0,613372093
IOS	0,03664263	1	5,000000075	0,61627907
fibra de internet	0,03664263	1	5,000000075	0,619186047
Radio	0,03664263	1	5,000000075	0,622093023
Telegram	0,03443419	1	5,000000075	0,625
Chat	0,03443419	1	5,000000075	0,627906977
Arroba	0,03443419	1	5,000000075	0,630813953
Disponibilidad	0,03443419	1	5,000000075	0,63372093
Digitalización	0,03443419	1	5,000000075	0,636627907
Electrónica	0,03443419	1	5,000000075	0,639534884
Drive	0,03443419	1	5,000000075	0,64244186
Sticker	0,03443419	1	5,000000075	0,645348837
Data Base	0,03443419	1	5,000000075	0,648255814
pantalla digital	0,03443419	1	5,000000075	0,651162791
Emoticonos	0,03235886	1	5,000000075	0,654069767
Accesibilidad	0,03235886	1	5,000000075	0,656976744
Comunidad	0,03235886	1	5,000000075	0,659883721
Informática	0,03235886	1	5,000000075	0,662790698
tarjeta de memoria	0,03235886	1	5,000000075	0,665697674
conexión a internet	0,03235886	1	5,000000075	0,668604651
conexión con personas que no están próximas	0,03235886	1	5,000000075	0,671511628
Robot	0,03235886	1	5,000000075	0,674418605
Procesador	0,03235886	1	5,000000075	0,677325581
Flipped Classrom	0,03235886	1	5,000000075	0,680232558
Excel	0,03127582	2	10,00000015	0,686046512

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Chips	0,03040861	1	5,000000075	0,688953488
Click	0,03040861	1	5,000000075	0,691860465
Robótica	0,03040861	1	5,000000075	0,694767442
Nube	0,03040861	1	5,000000075	0,697674419
Dispositivo	0,03040861	1	5,000000075	0,700581395
mando a distancia	0,03040861	1	5,000000075	0,703488372
Proxy	0,03040861	1	5,000000075	0,706395349
memoria Ram	0,03040861	1	5,000000075	0,709302326
Periódico	0,0285759	1	5,000000075	0,712209302
Velocidad	0,0285759	1	5,000000075	0,715116279
Cobertura	0,0285759	1	5,000000075	0,718023256
Archivo	0,0285759	1	5,000000075	0,720930233
Navegador	0,0285759	1	5,000000075	0,723837209
Cookies	0,0285759	1	5,000000075	0,726744186
Pulgadas	0,0285759	1	5,000000075	0,729651163
Relación	0,02685364	1	5,000000075	0,73255814
Fibra	0,02685364	1	5,000000075	0,735465116
Enchufe	0,02685364	1	5,000000075	0,738372093
Email	0,02685364	1	5,000000075	0,74127907
Plataforma	0,02685364	1	5,000000075	0,744186047
programas informáticos	0,02685364	1	5,000000075	0,747093023
4G	0,02685364	1	5,000000075	0,75
Megapíxeles	0,02685364	1	5,000000075	0,752906977
Robar	0,02685364	1	5,000000075	0,755813953
Microondas	0,02523519	1	5,000000075	0,75872093
ADSL	0,02523519	1	5,000000075	0,761627907
Aproximación	0,02523519	1	5,000000075	0,764534884
banco de imágenes	0,02523519	1	5,000000075	0,76744186
Teletrabajo	0,02523519	1	5,000000075	0,770348837
herramienta digital	0,02523519	1	5,000000075	0,773255814
Memoria	0,02523519	1	5,000000075	0,776162791
PlayStation	0,02523519	1	5,000000075	0,779069767
Yahoo respuestas	0,02371427	1	5,000000075	0,781976744
Virtual	0,02371427	1	5,000000075	0,784883721
Foro	0,02371427	1	5,000000075	0,787790698
Tarjeta sim	0,02371427	1	5,000000075	0,790697674
3G	0,02371427	1	5,000000075	0,793604651
Audio	0,02371427	1	5,000000075	0,796511628
Clases	0,02228503	1	5,000000075	0,799418605
dispositivos electrónicos	0,02228503	1	5,000000075	0,802325581
Imágenes	0,02228503	1	5,000000075	0,805232558
pizarra digital	0,02228503	1	5,000000075	0,808139535
diario colaborativo	0,02228503	1	5,000000075	0,811046512
aprendizaje online	0,02228503	1	5,000000075	0,813953488
Telecomunicaciones	0,02228503	1	5,000000075	0,816860465
Película	0,02228503	1	5,000000075	0,819767442

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Teclear	0,02228503	1	5,000000075	0,822674419
Monitor	0,02094192	1	5,000000075	0,825581395
Actividades	0,02094192	1	5,000000075	0,828488372
Portabilidad	0,02094192	1	5,000000075	0,831395349
video	0,02094192	1	5,000000075	0,834302326
Táctil	0,02094192	1	5,000000075	0,837209302
reloj inteligente	0,02094192	1	5,000000075	0,840116279
Código	0,02094192	1	5,000000075	0,843023256
Diapositiva	0,02094192	1	5,000000075	0,845930233
Exámenes	0,01967976	1	5,000000075	0,848837209
Innovación	0,01967976	1	5,000000075	0,851744186
Noticia	0,01967976	1	5,000000075	0,854651163
Excell	0,01967976	1	5,000000075	0,85755814
Dígito	0,01967976	1	5,000000075	0,860465116
ingenieros tecnológicos	0,01967976	1	5,000000075	0,863372093
Marketing	0,01849367	1	5,000000075	0,86627907
Canción	0,01849367	1	5,000000075	0,869186047
Inalámbrico	0,01849367	1	5,000000075	0,872093023
Contraseña	0,01849367	1	5,000000075	0,875
Microsoft	0,01849367	1	5,000000075	0,877906977
Cables	0,01737906	1	5,000000075	0,880813953
Twitter	0,01737906	1	5,000000075	0,88372093
innovación didáctica	0,01737906	1	5,000000075	0,886627907
plataforma virtual	0,01737906	1	5,000000075	0,889534884
JAVA	0,01737906	1	5,000000075	0,89244186
Microprocesador	0,01737906	1	5,000000075	0,895348837
Antena	0,01633164	1	5,000000075	0,898255814
aprendizaje colaborativo	0,01633164	1	5,000000075	0,901162791
Música	0,01633164	1	5,000000075	0,904069767
Cámara	0,01633164	1	5,000000075	0,906976744
Enchufar	0,01633164	1	5,000000075	0,909883721
Amplificador	0,01534734	1	5,000000075	0,912790698
Redes	0,01534734	1	5,000000075	0,915697674
Escribir	0,01534734	1	5,000000075	0,918604651
Desconectar	0,01534734	1	5,000000075	0,921511628
Messenger.	0,01442236	1	5,000000075	0,927325581
Navegación	0,01442236	1	5,000000075	0,930232558
Leer	0,01442236	1	5,000000075	0,933139535
URL	0,01442236	1	5,000000075	0,936046512
acoso virtual	0,01355313	1	5,000000075	0,938953488
Dirección	0,01355313	1	5,000000075	0,941860465
Cyberbullying	0,01273629	1	5,000000075	0,947674419
Link	0,01273629	1	5,000000075	0,950581395
sala de computación	0,01196868	1	5,000000075	0,956395349
Hipervínculo	0,01196868	1	5,000000075	0,959302326
técnico en computación	0,01124733	1	5,000000075	0,962209302

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Diseño	0,01124733	1	5,000000075	0,965116279
ingeniero en informática	0,01056946	1	5,000000075	0,968023256
Usabilidad	0,01056946	1	5,000000075	0,970930233
Rapidez	0,00993245	1	5,000000075	0,973837209
Peso	0,00993245	1	5,000000075	0,976744186
Almacenaje	0,00933382	1	5,000000075	0,979651163
Byte	0,00877128	1	5,000000075	0,98255814
Giga	0,00824264	1	5,000000075	0,985465116
Mega	0,00774586	1	5,000000075	0,988372093
Office	0,00684032	1	5,000000075	0,99127907
HTTP	0,00642805	1	5,000000075	0,994186047
Publisher	0,00567657	1	5,000000075	0,997093023
Googlear	0,00501294	1	5,000000075	1

Tabla 16. índice de disponibilidad léxica centro de interés Informática, grupo nativos

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
Hospital	0,33152789	10	50	0,055248619
Enfermedad	0,23562911	8	40,0000006	0,099447514
Sano	0,21996601	5	25	0,127071823
Enfermero	0,20269756	6	30,00000119	0,160220994
Fiebre	0,1522105	5	25	0,187845304
Saludable	0,14367294	3	15,0000006	0,20441989
Virus	0,1373459	3	15,0000006	0,220994475
Médico	0,12015184	4	20,0000003	0,243093923
Enfermo	0,11031181	3	15,0000006	0,259668508
COVID	0,0958669	4	20,0000003	0,281767956
Cuerpo	0,09367295	2	10,00000015	0,29281768
Operación	0,09152573	3	15,0000006	0,309392265
Vacuna	0,0828165	3	15,0000006	0,325966851
Doctor	0,0805068	3	15,0000006	0,342541436
Cáncer	0,07233682	3	15,0000006	0,359116022
Mascarilla	0,07070085	3	15,0000006	0,375690608
Resfriado	0,06939401	2	10,00000015	0,386740331
Medicina	0,06693987	2	10,00000015	0,397790055
Dieta	0,06292395	2	10,00000015	0,408839779
Deporte	0,06035022	2	10,00000015	0,419889503
Verdura	0,05873984	2	10,00000015	0,430939227
Dolor	0,05773836	3	15,0000006	0,447513812
Corazón	0,05659689	2	10,00000015	0,458563536
Tos	0,0550864	2	10,00000015	0,46961326
dolor de cabeza	0,0502593	2	10,00000015	0,480662983
Pandemia	0,05	1	5,000000075	0,486187845
Adelgazamiento	0,05	1	5,000000075	0,491712707
salud pública	0,05	1	5,000000075	0,497237569
Parecetamol	0,05	1	5,000000075	0,502762431
Caloría	0,05	1	5,000000075	0,508287293
Coronavirus	0,05	1	5,000000075	0,513812155

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Vegetariano	0,05	1	5,000000075	0,519337017
Farmacia	0,04389944	2	10,00000015	0,53038674
Curar	0,04367295	1	5,000000075	0,535911602
ejercicio diario	0,04367295	1	5,000000075	0,541436464
Hipertensión	0,04367295	1	5,000000075	0,546961326
salud materno-infantil	0,04367295	1	5,000000075	0,552486188
centro de salud	0,04367295	1	5,000000075	0,55801105
Dentadura	0,04367295	1	5,000000075	0,563535912
Saludablemente	0,04367295	1	5,000000075	0,569060773
Radiografía	0,03814653	1	5,000000075	0,574585635
Fuerte	0,03814653	1	5,000000075	0,580110497
Calor	0,03814653	1	5,000000075	0,585635359
Correr	0,03814653	1	5,000000075	0,591160221
salud mental	0,03814653	1	5,000000075	0,596685083
Ojos	0,03814653	1	5,000000075	0,602209945
Descansar	0,03331943	1	5,000000075	0,607734807
salud de hierro	0,03331943	1	5,000000075	0,613259669
Ejercicio	0,03331943	1	5,000000075	0,61878453
Frío	0,03331943	1	5,000000075	0,624309392
comida saludable	0,03331943	1	5,000000075	0,629834254
Vitamina	0,03331943	1	5,000000075	0,635359116
salud sexual	0,03331943	1	5,000000075	0,640883978
Alimentación	0,03331943	1	5,000000075	0,64640884
Indispuesto	0,03331943	1	5,000000075	0,651933702
Ambulancia	0,03231795	2	10,00000015	0,662983425
guardar cama	0,02910315	1	5,000000075	0,668508287
Bienestar	0,02910315	1	5,000000075	0,674033149
Beneficioso	0,02910315	1	5,000000075	0,679558011
Resfriarse	0,02910315	1	5,000000075	0,685082873
Favorable	0,02910315	1	5,000000075	0,690607735
salud digital	0,02910315	1	5,000000075	0,696132597
dolor de estómago	0,02910315	1	5,000000075	0,701657459
Hemorragia	0,02910315	1	5,000000075	0,70718232
Órgano	0,02910315	1	5,000000075	0,712707182
salud física	0,02910315	1	5,000000075	0,718232044
Dormir	0,02910315	1	5,000000075	0,723756906
dolor físico	0,02910315	1	5,000000075	0,729281768
Insano	0,02910315	1	5,000000075	0,73480663
Nacimiento	0,02910315	1	5,000000075	0,740331492
Bueno	0,02542041	1	5,000000075	0,745856354
Gripe	0,02542041	1	5,000000075	0,751381215
tener salud	0,02542041	1	5,000000075	0,756906077
Muerte	0,02542041	1	5,000000075	0,762430939
salud reproductiva	0,02542041	1	5,000000075	0,767955801
Alcohol	0,02542041	1	5,000000075	0,773480663
Operar	0,02542041	1	5,000000075	0,779005525

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Recuperación	0,02542041	1	5,000000075	0,784530387
Dentista	0,02542041	1	5,000000075	0,790055249
Azúcar	0,02542041	1	5,000000075	0,79558011
salud metal	0,02542041	1	5,000000075	0,801104972
dolor psicológico	0,02542041	1	5,000000075	0,806629834
Insalubre	0,02542041	1	5,000000075	0,812154696
comida sana	0,02220369	1	5,000000075	0,817679558
mal cuerpo	0,02220369	1	5,000000075	0,82320442
Termómetro	0,02220369	1	5,000000075	0,828729282
estar enfermo	0,02220369	1	5,000000075	0,834254144
salud genésica	0,02220369	1	5,000000075	0,839779006
sentirse bien	0,02220369	1	5,000000075	0,845303867
Seguridad	0,02220369	1	5,000000075	0,850828729
medicina tradicional				
China	0,02220369	1	5,000000075	0,856353591
Infeccioso	0,01939401	1	5,000000075	0,861878453
tener dolor	0,01939401	1	5,000000075	0,867403315
salud ocupacional	0,01939401	1	5,000000075	0,872928177
sentirse mal	0,01939401	1	5,000000075	0,878453039
sector de la salud	0,01939401	1	5,000000075	0,883977901
Vomitar	0,01693987	1	5,000000075	0,889502762
Digestión	0,01693987	1	5,000000075	0,895027624
Sangre	0,01479628	1	5,000000075	0,900552486
Jaquica	0,01479628	1	5,000000075	0,906077348
Enfermarse	0,01479628	1	5,000000075	0,91160221
Vacunación	0,01479628	1	5,000000075	0,917127072
Caminar	0,01292394	1	5,000000075	0,922651934
Infectar	0,01292394	1	5,000000075	0,928176796
Sudar	0,01128854	1	5,000000075	0,933701657
Nutrición	0,01128854	1	5,000000075	0,939226519
Pulmón	0,01128854	1	5,000000075	0,944751381
consultorio médico	0,01128854	1	5,000000075	0,950276243
Herida	0,01128854	1	5,000000075	0,955801105
Sanitarios	0,01128854	1	5,000000075	0,961325967
Inyección	0,00986007	1	5,000000075	0,966850829
Riñón	0,00986007	1	5,000000075	0,972375691
Auscultar	0,00861237	1	5,000000075	0,977900552
Tiritas	0,00861237	1	5,000000075	0,983425414
Antinflamatorios	0,00861237	1	5,000000075	0,988950276
Anticuerpos	0,00752255	1	5,000000075	0,994475138
fortalecer la salud	0,00657064	1	5,000000075	1

Tabla 17. índice de disponibilidad léxica centro de interés Salud, grupo no nativos

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
padre	0,42883524	10	50	0,048076923
abuelos	0,42385229	13	64,99999762	0,110576923
madre	0,38229579	10	50	0,158653846
hermanos	0,32952142	11	55,00000119	0,211538462
hijos	0,30347642	10	50	0,259615385
Tíos	0,28956378	11	55,00000119	0,3125
padres	0,21539387	5	25	0,336538462
familia	0,18264022	5	25	0,360576923
yerno	0,15395156	7	34,9999994	0,394230769
primos	0,15025571	7	34,9999994	0,427884615
sobrinos	0,13089283	6	30,00000119	0,456730769
amigos	0,12939881	5	25	0,480769231
nuera	0,1265316	6	30,00000119	0,509615385
novios	0,1210077	6	30,00000119	0,538461538
amistad	0,10798772	3	15,0000006	0,552884615
nietos	0,10215533	5	25	0,576923077
Jefe	0,09441184	6	30,00000119	0,605769231
pareja	0,08934981	3	15,0000006	0,620192308
cuñados	0,07606731	4	20,0000003	0,639423077
mejores amigos	0,06989374	3	15,0000006	0,653846154
Amor	0,06583184	2	10,00000015	0,663461538
mujer	0,05967133	2	10,00000015	0,673076923
suegros	0,05798772	2	10,00000015	0,682692308
empleado	0,05343191	5	25	0,706730769
familia nuclear	0,05	1	5,000000075	0,711538462
profesor	0,04269845	3	15,0000006	0,725961538
amigo	0,04242498	1	5,000000075	0,730769231
convivencia	0,04242498	1	5,000000075	0,735576923
bisabuelos	0,04242498	1	5,000000075	0,740384615
afecto	0,04242498	1	5,000000075	0,745192308
ser uña y carne	0,03599758	1	5,000000075	0,75
socializar	0,03599758	1	5,000000075	0,754807692
estudiante	0,03209194	2	10,00000015	0,764423077
matrimonio	0,03182451	2	10,00000015	0,774038462
confiar	0,03054393	1	5,000000075	0,778846154
Socio	0,03054393	1	5,000000075	0,783653846
compañero	0,03054393	1	5,000000075	0,788461538
parientes	0,03019626	2	10,00000015	0,798076923
nietos	0,02591651	1	5,000000075	0,802884615
ingeniero	0,02591651	1	5,000000075	0,807692308
respeto	0,02591651	1	5,000000075	0,8125
relaciones estrechas	0,02591651	1	5,000000075	0,817307692

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y
sinohablantes de ELE

maestro	0,02591651	1	5,000000075	0,822115385
esposo	0,02199015	1	5,000000075	0,826923077
bombero	0,02199015	1	5,000000075	0,831730769
sentimental	0,02199015	1	5,000000075	0,836538462
ser sincero	0,02199015	1	5,000000075	0,841346154
desconocido	0,01865863	1	5,000000075	0,846153846
emociones	0,01865863	1	5,000000075	0,850961538
leal	0,01865863	1	5,000000075	0,855769231
vecinos	0,01583184	1	5,000000075	0,860576923
confiar en alguien	0,01583184	1	5,000000075	0,865384615
comprensión	0,01583184	1	5,000000075	0,870192308
reunirse con la familia	0,01583184	1	5,000000075	0,875
inferior trabajo	0,01343331	1	5,000000075	0,879807692
Traicionar	0,01343331	1	5,000000075	0,884615385
Secretos	0,01343331	1	5,000000075	0,889423077
médico	0,01139816	1	5,000000075	0,894230769
sociedad humana	0,01139816	1	5,000000075	0,899038462
marido	0,01139816	1	5,000000075	0,903846154
economía	0,01139816	1	5,000000075	0,908653846
compartir	0,01139816	1	5,000000075	0,913461538
compañero de piso	0,01139816	1	5,000000075	0,918269231
Alumno	0,01139816	1	5,000000075	0,923076923
animal social	0,00967133	1	5,000000075	0,927884615
paciencia	0,00967133	1	5,000000075	0,932692308
Hija	0,00967133	1	5,000000075	0,9375
empleador	0,00967133	1	5,000000075	0,942307692
Taxista	0,00967133	1	5,000000075	0,947115385
divorciado	0,00820612	1	5,000000075	0,951923077
profesionalidad	0,00820612	1	5,000000075	0,956730769
científico	0,00820612	1	5,000000075	0,961538462
Suegra	0,00696289	1	5,000000075	0,966346154
Casado	0,00696289	1	5,000000075	0,971153846
contactos	0,00696289	1	5,000000075	0,975961538
Colega	0,00696289	1	5,000000075	0,980769231
Vecino	0,00696289	1	5,000000075	0,985576923
profesora	0,00590801	1	5,000000075	0,990384615
enemigo	0,00501294	1	5,000000075	0,995192308
Boda	0,00501294	1	5,000000075	1

Tabla 18. índice de disponibilidad léxica centro de interés Relaciones grupo no nativos

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Vocablo	IDL	Frecuencia	% Aparición	Frec. Acumulada
internet	0,30657083	8	40,0000006	0,058823529
ordenador	0,30018049	7	34,9999994	0,110294118
computadora	0,28326792	7	34,9999994	0,161764706
móvil	0,22965693	6	30,00000119	0,205882353
teléfono	0,15956722	5	25	0,242647059
ratón	0,11680652	4	20,0000003	0,272058824
5G	0,11582424	3	15,0000006	0,294117647
redes sociales	0,08283384	4	20,0000003	0,323529412
inteligencia artificial	0,08152866	3	15,0000006	0,345588235
pantalla	0,06426027	2	10,00000015	0,360294118
portátil	0,06421307	3	15,0000006	0,382352941
pagar con el móvil	0,05853919	3	15,0000006	0,404411765
Navegar	0,05775319	3	15,0000006	0,426470588
vehículos eléctricos	0,05	1	5,000000075	0,433823529
software	0,05	1	5,000000075	0,441176471
digitalización	0,05	1	5,000000075	0,448529412
Cargar	0,05	1	5,000000075	0,455882353
genéticamente modificado	0,05	1	5,000000075	0,463235294
Ipad	0,04818172	2	10,00000015	0,477941176
tele(visión)	0,04448179	2	10,00000015	0,492647059
Enlace	0,04427891	2	10,00000015	0,507352941
Carpeta	0,04056608	1	5,000000075	0,514705882
vehículos de nueva energía	0,04056608	1	5,000000075	0,522058824
técnica	0,04056608	1	5,000000075	0,529411765
comprobar	0,04056608	1	5,000000075	0,536764706
aplicación	0,04056608	1	5,000000075	0,544117647
seguridad biológica	0,04056608	1	5,000000075	0,551470588
Desarrollo	0,04056608	1	5,000000075	0,558823529
Tablet	0,0351532	2	10,00000015	0,573529412
Economía	0,03291213	1	5,000000075	0,580882353
Teletrabajo	0,03291213	1	5,000000075	0,588235294
información	0,03291213	1	5,000000075	0,595588235
Chips	0,03291213	1	5,000000075	0,602941176
VE	0,03291213	1	5,000000075	0,610294118
Mouse	0,03291213	1	5,000000075	0,617647059
Impresora	0,03291213	1	5,000000075	0,625
correo electrónico	0,03291213	1	5,000000075	0,632352941
Tecnología	0,03291213	1	5,000000075	0,639705882
energía solar	0,03291213	1	5,000000075	0,647058824
Descargar	0,03105088	2	10,00000015	0,661764706
videojuegos	0,02914627	2	10,00000015	0,676470588
tren de alta velocidad	0,02670231	1	5,000000075	0,683823529
mainframes	0,02670231	1	5,000000075	0,691176471

Disponibilidad léxica en tres centros de interés relacionales en estudiantes nativos de español y sinohablantes de ELE

Incrementar	0,02670231	1	5,000000075	0,698529412
Autotransporte	0,02670231	1	5,000000075	0,705882353
Audio	0,02670231	1	5,000000075	0,713235294
Videojuego	0,02670231	1	5,000000075	0,720588235
Mensaje	0,02670231	1	5,000000075	0,727941176
inteligencia virtual	0,02670231	1	5,000000075	0,735294118
modificación genética	0,02670231	1	5,000000075	0,742647059
periódicos online	0,02670231	1	5,000000075	0,75
Cohete	0,02166416	1	5,000000075	0,757352941
realidad aumentada	0,02166416	1	5,000000075	0,764705882
noticia virtual	0,02166416	1	5,000000075	0,772058824
tener problemas con internet	0,02166416	1	5,000000075	0,779411765
disco duro	0,02166416	1	5,000000075	0,786764706
Bolsa	0,02166416	1	5,000000075	0,794117647
Cargador	0,02166416	1	5,000000075	0,801470588
Google	0,02166416	1	5,000000075	0,808823529
Avión	0,02166416	1	5,000000075	0,816176471
Interaccional	0,02166416	1	5,000000075	0,823529412
enseñanza online	0,0175766	1	5,000000075	0,830882353
Zoom	0,0175766	1	5,000000075	0,838235294
pila de carga	0,0175766	1	5,000000075	0,845588235
Email	0,0175766	1	5,000000075	0,852941176
Cable	0,0175766	1	5,000000075	0,860294118
Correo	0,0175766	1	5,000000075	0,867647059
Globalización	0,0175766	1	5,000000075	0,875
mando a distancia	0,01426027	1	5,000000075	0,882352941
escribir un email	0,01426027	1	5,000000075	0,889705882
Enviar	0,01426027	1	5,000000075	0,897058824
Electricidad	0,01426027	1	5,000000075	0,904411765
Virtual	0,01426027	1	5,000000075	0,911764706
Batería	0,01426027	1	5,000000075	0,919117647
Wechat	0,01262859	2	10,00000015	0,933823529
Teclado	0,01156967	1	5,000000075	0,941176471
pasar por internet	0,01156967	1	5,000000075	0,948529412
Robótica	0,01156967	1	5,000000075	0,955882353
Contraseña	0,01156967	1	5,000000075	0,963235294
Analógico	0,00938672	1	5,000000075	0,970588235
Digital	0,00761565	1	5,000000075	0,977941176
Estéreo	0,00761565	1	5,000000075	0,985294118
Micrófono	0,00617874	1	5,000000075	0,992647059
coches sin conductor	0,00501294	1	5,000000075	1

Tabla 19. Índice de disponibilidad léxica centro de interés Informática, no nativos