



**Disponibilidad y riqueza léxicas de un grupo de aprendientes de Español como Lengua
Extranjera de A2 y B2 en la ciudad de Medellín, Colombia.**

Sindy Melissa Metaute Arango

Asesor: Dr. Juan David Martínez Hincapié

Trabajo de grado para optar al título de Filóloga hispanista

**Universidad de Antioquia
Facultad de Comunicaciones
Filología Hispánica
Medellín 2020**

AGRADECIMIENTOS

Al Fondo para apoyar los trabajos de grado de pregrado de la Facultad de Comunicaciones y al Comité para el Desarrollo de la Investigación de la Universidad de Antioquia por la financiación de este proyecto de investigación.

Al profesor Juan David Martínez por su compromiso y acompañamiento en este proceso.

A la Universidad Eafit y a la escuela de español Elefun por brindarme la posibilidad de recoger los datos para mi investigación. A la escuela Learn Spanish Now por la oportunidad de ser parte de su equipo de trabajo.

A mi familia y amigos por su apoyo.

A Daniela por su amor.

A todas las personas que han sido parte de mi proceso académico.

TABLA DE CONTENIDO

Capítulo 1: Introducción y objetivos de la investigación.....	5
1.1. Introducción general.....	5
1.2. Planteamiento del problema	6
1.3. Justificación	6
1.4. Objetivos	8
1.4.1 Objetivo general:	8
1.4.2. Objetivos específicos	8
Capítulo 2: Antecedentes	9
Capítulo 3: Marco teórico	14
3.1. Conceptos sobre el léxico	14
3.1.1. Lexicología, léxico y vocabulario.....	14
3.1.2. Palabra y vocablo.....	14
3.2. Lexicón mental y competencia léxica	14
3.3. Frecuencia de las palabras en una lengua	15
3.4. Léxico común, léxico básico y léxico fundamental	15
3.4.1. Léxico frecuente y léxico disponible	16
3.5. Fórmulas para medir la disponibilidad léxica o el léxico disponible	16
3.6. Riqueza léxica	18
3.6.1 Fórmulas para medir la riqueza léxica	18
3.7. Marco Común Europeo	20
3.8. ELE en la virtualidad.....	22
Capítulo 4: Metodología	23
4.1. Paradigma y enfoque de investigación	23
4.2. Alcance de la investigación: exploratorio descriptivo.....	24
4.3. Diseño de la investigación: no experimental o <i>ex postfacto</i>	24
4.4. Variables de la investigación	25
4.5. Muestra	26
4.6. Instrumentos.....	27
4.7. Validación de instrumentos: centros de interés e instrucción texto escrito.	28

4.8. Software.....	29
4.9. Pruebas estadísticas para análisis de datos: distribución normal y test no paramétricos..	31
4.10. Etapas de la investigación.....	32
Capítulo 5: Resultados y discusión.....	33
5.1. Disponibilidad léxica.....	33
5.1.1. Selección del léxico analizado.....	33
5.1.2. Léxico no tenido en cuenta.....	33
5.1.3. Total de palabras y palabras diferentes	35
5.1.4. Promedio, varianza, desviación y mediana	36
5.1.5. Distribución normal	38
5.1.6. Test Kruskal-Wallis	40
5.1.7. Test Mann-Whitney	40
5.1.8. Características generales del léxico disponible encontrado	41
5.1.9. Unidades léxicas más disponibles y menos disponibles	43
5.2. Riqueza léxica	47
5.2.1. Selección de los textos analizados	47
5.2.2. Léxico no tenido en cuenta	47
5.2.4. Extensión de los textos.....	48
5.2.5. Promedio, varianza, desviación y mediana	48
5.2.6. Intervalo de aparición de palabras con contenido semántico nocional (IA)	49
5.2.7. Distribución normal	49
5.2.8. Porcentaje de vocablos (PV)	50
5.2.9. Cuestionamiento a la fórmula.....	52
5.3. Relación entre la disponibilidad léxica y la riqueza léxica.....	52
Capítulo 7: Conclusiones	55
Referencias bibliográficas	57

Capítulo 1: Introducción y objetivos de la investigación

1.1. Introducción general

El español se ha posicionado como una de las lenguas más habladas en el mundo y también como una de las más estudiadas desde el plano lingüístico por la difusión creciente que ha tenido en los últimos años (Sierra, 2016). En la actualidad, en Colombia, el Español como Lengua Extranjera (ELE) hace parte de la oferta educativa de diferentes Institutos de Educación Superior.

Ahora bien, entre los temas investigados alrededor del mundo con relación al español, se encuentran la disponibilidad léxica (DL) y la riqueza léxica (RL). Los estudios de DL están relacionados principalmente con el Proyecto Panhispánico coordinado por Humberto López Morales y los estudios de RL han estado direccionados a hablantes niños y jóvenes de Español como Lengua Materna (ELM) en diferentes países hispanohablantes, pero según la bibliografía encontrada, en Colombia no se han realizado estudios sobre la DL y la RL de aprendientes de ELE.

López (2014), dice que la importancia y los aportes teóricos de los estudios de disponibilidad y riqueza léxicas del ELE residen en sus alcances como “conocer el vocabulario realmente disponible de los estudiantes, determinar la influencia de factores socioculturales en el aprendizaje del léxico y graduar el vocabulario para su enseñanza” (p.412). Es decir, que los resultados de la investigación, además de servir como aportes teóricos y metodológicos para caracterizar el fenómeno lingüístico, sirven para conocer los aspectos lingüísticos mencionados en la variedad dialectal del español hablado en Medellín. A su vez, estos aportes contribuyen a la ampliación de la comunidad académica y discursiva sobre el tema en el país y al posicionamiento de la Universidad de Antioquia como eje en el conocimiento de los fenómenos lingüísticos del ELE.

1.2. Planteamiento del problema

En el Proyecto Panhispánico de disponibilidad léxica, a partir de grupos de muestra de estudiantes de nivel preuniversitario cuya lengua materna es el español, se busca “elaborar diccionarios de disponibilidad léxica para las diversas zonas del mundo hispánico con homogeneidad de criterios” (Grupo de investigación DispoLex, 2003-2019). Como aportes a este proyecto, existen múltiples estudios acerca de la disponibilidad léxica del ELM en diferentes países como Cuba (Pérez, 2011) y México (Cabrera, 2010), aunque la mayoría de ellos se han realizado en España: (López, 1995), (López, 2008). Por otra parte, en relación con los estudios de disponibilidad léxica de ELE, estos no son cuantiosos y se han realizado en países como Eslovenia (Sifrar, 2012), Chile (Blanco y Ferreira, 2018) Marruecos (Serfati, 2016), China (Jing, 2012), entre otros.

Para hablar del contexto colombiano, según la bibliografía consultada, no existen estudios que revisen la disponibilidad y riqueza léxicas de estudiantes extranjeros que se encuentren realizando un curso de español en Colombia ni específicamente en una universidad de la ciudad de Medellín. De lo anterior, surge la siguiente cuestión: ¿cuál es la disponibilidad y la riqueza léxica de los estudiantes de ELE que se encuentran realizando un curso de español en la ciudad de Medellín con niveles A2 y B2 y cuál es la diferencia entre ambos?

1.3. Justificación

La DL del Español como Lengua Extranjera será abordada en este proyecto desde el ámbito lingüístico. Al ser el español la segunda lengua más hablada en el mundo según Ethnologue (2018), los aspectos lingüísticos de este son importantes de ser estudiados, ya que son de interés y pueden tener impacto a nivel global.

En Colombia se tiene mención de que en la Universidad de Antioquia en los años 50's, se planea el programa de Español para extranjeros, con el que se pretende difundir la lengua castellana y la

cultura colombiana (Girón, 1998), es decir que los temas alrededor del ELE en el país han sido de interés de la comunidad académica desde hace varias décadas. En la actualidad, en el caso específico de Medellín, la oferta del ELE se encuentra en instituciones de educación superior como la Universidad de Antioquia, la Universidad Pontificia Bolivariana, la Universidad Eafit y la Universidad de Medellín, y en diversas escuelas privadas de español que se han creado en los últimos años como Elefun Spanish School, Toucan Spanish School y Learn Spanish Now.

Ahora bien, como se mencionó anteriormente, uno de los temas investigados alrededor del mundo con relación a ELE ha sido la DL y la RL del español en aprendientes extranjeros. Sin embargo, al no contar con investigaciones sobre este tema en Colombia, es importante realizar estudios que brinden aportes teóricos, con respecto a la descripción, caracterización y comprensión del fenómeno lingüístico, así como aportes metodológicos, en cuanto a la creación del precedente de esta línea de investigación en Colombia.

Tal como lo plantea López (2014), la importancia y los aportes teóricos de los estudios de disponibilidad léxica del Español como Lengua Extranjera residen en sus alcances que son:

Conocer el vocabulario realmente disponible de los estudiantes, determinar la influencia de factores socioculturales en el aprendizaje del léxico, precisar la influencia de la lengua materna en el léxico aprendido, examinar las distintas fases del proceso de aprendizaje/adquisición de ELE, comparar el léxico disponible aprendido con el léxico disponible de los nativos, seleccionar y graduar el vocabulario para su enseñanza (p.412).

Es importante tener en cuenta que los resultados de esta investigación sirven como insumo para Institutos de Educación y escuelas privadas que tienen ELE dentro de su oferta académica y para personas que se dediquen a la docencia o a la planeación curricular, para desarrollar aplicaciones metodológicas que ayuden a potenciar el conocimiento léxico de los estudiantes. Por último, es

pertinente resaltar que el proyecto aporta al fortalecimiento de la línea de investigación en ELE del pregrado de Filología Hispánica y a la construcción del programa de Español de la Universidad de Antioquia, en el que uno de sus campos de interés es conocer diferentes aspectos de ELE para realizar diversas aplicaciones y revitalizar a la Universidad de Antioquia como centro importante en investigación y enseñanza alrededor de ELE en el país.

1.4. Objetivos

1.4.1 Objetivo general: identificar la disponibilidad y la riqueza léxicas en un corpus de pruebas de aprendientes de Español como Lengua Extranjera con niveles A2 y B2 de dos instituciones de enseñanza en ELE de Medellín en el presente año (2020).

1.4.2. Objetivos específicos

- Identificar la disponibilidad léxica de los centros de interés: lugares de la ciudad, alimentos-bebidas y medios de transporte; y la riqueza léxica en aprendientes de niveles A2 y B2.
- Identificar la disponibilidad léxica de los centros de interés anteriormente mencionados y la riqueza léxica de un grupo control conformado por estudiantes universitarios de la Universidad de Antioquia.
- Comparar la disponibilidad léxica en los centros de interés seleccionados entre aprendientes A2 y B2, y asimismo con un grupo control.
- Comparar la riqueza léxica entre los aprendientes de ELE y el grupo control.
- Comparar la disponibilidad y la riqueza léxicas entre los grupos de datos analizados.

Capítulo 2: Antecedentes

Las preguntas por el léxico surgen muy tempranamente entre los griegos, ya que:

si tenemos en cuenta que durante tres siglos (del 600 al 300 a. C.) los poemas homéricos fueron objeto de estudio y que estos tenían ya para los griegos de esos siglos problemas de comprensión de ciertas parcelas del léxico, se comprende fácilmente que por razones escolares surgieran léxicos para entender muchas de sus expresiones (Serrano, 1977, p.61).

De esta manera, surgió una pregunta por las palabras y su significado, que posteriormente se trasladó a textos de temas específicos, como en los estudios de los que se tienen registro sobre los antiguos textos legales de Dracon y Solón. Así, es claro que los estudios del léxico han sido una preocupación que data de muchos siglos atrás. Ahora bien, dentro de los estudios del léxico se encuentran los estudios de DL, que según López (1996), iniciaron en Francia en la década de los 50's impulsados por los resultados de los trabajos léxico-estadísticos que se estaban realizando para la creación del *Francais élémentaire*, un manual de enseñanza del francés como lengua extranjera. De esta manera, fue en Francia donde se empezó la construcción y discusión de conceptos como *palabras frecuentes* y *palabras disponibles*. Así, se inició a vislumbrar que “la lengua poseía un amplio conjunto de palabras de contenido semántico muy concreto, que únicamente manejaba si lo permitía el tema del discurso y a partir de ahí se realizaron múltiples estudios en diferentes países” (López, 1996). Según López (1996), el lingüista Charles Müller en el año 1968 propuso estudiar la DL no solamente a partir de la frecuencia de las palabras, sino también a partir del orden de aparición de ellas en las listas, sin embargo, fue solamente hasta 1983 que comenzaron a diseñarse fórmulas estadísticas que tuvieran en cuenta ambos aspectos, una de ellas fue propuesta por Lorán y López Morales y, posteriormente, en 1987, Strassburger y López diseñaron una fórmula que sigue teniendo vigencia para los estudios de disponibilidad léxica (p.246) y asimismo realizaron los primeros estudios de la disponibilidad léxica del Español como Lengua Materna (ELM).

Sin embargo, fue solamente hasta 1998 que se realizó la primera investigación de DL del Español como Lengua Extranjera por Carcedo González, quien escribió la primera monografía de la que se tiene registro, a partir de una muestra de 350 estudiantes finlandeses (López, 2014, p.399). Después de este trabajo, según López (2014) hasta ese año en el que hizo su rastreo bibliográfico, podían encontrarse 18 investigaciones sobre disponibilidad léxica del ELE en el mundo (p.401) y según la bibliografía encontrada, en los últimos cinco años se han realizado diferentes estudios en países como Marruecos (Serfati, 2016), Cuba (Sierra y Pérez, 2016), Italia (Paolini, 2017), Corea del Sur (Mendoza, 2018), Taiwán (Rubio, 2018), entre otros, lo que quiere decir que las investigaciones sobre la disponibilidad léxica del ELE siguen en vigencia.

El trabajo que realizó Paolini (2017), titulado *Estudio sobre disponibilidad léxica en alumnos de ELE en la Universidad de Padua*, contó con una muestra de 112 aprendientes de ELE que participaban en el curso de español de nivel B2 y que se encontraban inscritos en segundo año de la universidad en los pregrados Mediación Lingüística y Cultural, y Lenguas, Literaturas y Culturas Modernas pertenecientes al Departamento de Estudios Lingüísticos y Literarios de la Universidad de Padua. Paolini (2017), escogió 8 de los 16 centros de interés propuestos por el Proyecto Panhispánico y a través del procesamiento de los datos en *Dispolex* encontró que la edad, el nivel sociocultural y la zona geográfica de origen de los participantes no eran factores influyentes en sus niveles de DL, pero sí la cantidad de lenguas que conocían los hablantes participantes y también el tiempo de estudio del español.

Por otra parte, el trabajo titulado *Léxico disponible en tres centros de interés de aprendices de Español como Lengua Extranjera*, realizado por Blanco y Ferreira (2017) en Chile, contó con una muestra de 21 aprendientes de ELE con nivel B1, que se encontraban realizando un curso de español en la Universidad de Concepción. En esta investigación se buscó identificar la DL de los

aprendientes de español en tres centros de interés: la ciudad, alimentos y bebidas, medios de transporte. Las autoras procesaron los datos en el programa *Dispogen*, y encontraron que el centro de interés alimentos y bebidas arrojó mayor cantidad de léxico mientras que medios de transporte arrojó la menor cantidad, pero a la misma vez fue el centro de interés que presentó mayores coincidencias entre los aprendientes. Por otra parte, al hacer la caracterización de la DL del grupo analizado Blanco y Ferreira (2017), encontraron que el léxico disponible del grupo se encuentra acorde con el nivel B1 de competencia lingüística según el Marco Común Europeo de Referencia. Ahora bien, en Colombia, según Garzón y Penagos (2016), hasta ese año solo se tenía referencia de una investigación terminada de DL de Español como Lengua Materna realizada por Mateus y Santiago (2006) en la ciudad de Bogotá y posteriormente Garzón y Penagos (2016) realizaron una investigación también de ELM en la ciudad de Villavicencio, siendo estos, según la bibliografía encontrada, los únicos estudios terminados realizados hasta el momento en Colombia. En relación al Español como Lengua Extranjera, según la bibliografía consultada no existen estudios que revisen la disponibilidad léxica del Español como Lengua Extranjera de estudiantes que se encuentren realizando un curso de español en Colombia. Por otra parte, sobre los primeros estudios de riqueza léxica de hablantes nativos de una lengua López (2010) dice que estos aparecen en 1954 de la mano de Giraud.

Según López Morales (2010), los estudios sobre riqueza léxica en el mundo hispánico no son cuantiosos y la mayoría de ellos se han realizado para medir la RL en la producción escrita de niños y adolescentes. Uno de los trabajos más reconocidos alrededor de este tema, es el realizado por Ávila (1986) titulado *Léxico infantil de México; palabra, tipos y vocablos* que hizo a partir de un corpus de 4500 textos de tema libre, escritos por niños de tercero a sexto grado. En esta investigación, Ávila (1986) encontró diferencias porcentuales de RL mayores para las niñas, así

como diferencias acentuadas según el estrato socioeconómico de los participantes, revelando así una relación estrecha entre la riqueza léxica y el nivel sociocultural.

En este mismo sentido Filomena Cintrón (1993) hizo un estudio en Puerto Rico, titulado *Nuevos índices de Riqueza Léxica en escolares de Barranquitas*, en el que tomó una muestra de estudiantes de los grados tercero, sexto, noveno y décimo, encontrando un crecimiento asistemático en el progreso léxico de los estudiantes según la variable grado de escolaridad, y asimismo encontró incidencia en el factor sexo a favor de las niñas y del factor socioeconómico a favor de los estudiantes de un nivel social más alto. Ahora bien, en el caso específico de los estudios de RL de aprendientes de Español como Lengua Extranjera se tiene registro de algunas investigaciones realizadas en otros países.

La investigación realizada por Berton (2014), se titula *La riqueza léxica en la producción escrita de estudiantes suecos de ELE* y tuvo como muestra 180 escritos de estudiantes suecos de entre 15 y 19 años. Para analizar la riqueza léxica de los aprendientes, se les presentó una serie de viñetas a partir de las cuales debían escribir una historia en un máximo de 15 minutos. En esta investigación, se encontró que en los valores había gran influencia de la edad de los estudiantes, donde entre a mayor edad tenían mayor RL. Además de eso, se plantea que la longitud de los textos influye en los valores de la riqueza léxica, teniendo resultados más altos los textos más cortos, puesto que al ser más cortos contienen menos repeticiones.

Por su parte, en la investigación *El aprendizaje de locuciones verbales por parte de alumnos ELE rumanos* por (Prundeanu, 2013), si bien el trabajo se centra en el aprendizaje de las locuciones verbales, se estudia además la riqueza léxica de los textos recogidos. Para esto, el estudio fue realizado con cuatro grupos de aprendientes rumanos de ELE con un nivel de competencia lingüística de B2 y como resultado principal de esta investigación, se encuentra una

correspondencia entre la riqueza léxica y el uso de las locuciones verbales, así, quienes conocían más locuciones verbales a la misma vez tenían mayor riqueza léxica.

Si bien en Colombia, según la bibliografía encontrada, no se han realizado investigaciones sobre disponibilidad y riqueza léxicas del Español como Lengua Extranjera, se han realizado múltiples investigaciones en diferentes temas alrededor del ELE y asimismo se han desarrollado encuentros académicos con el fin de dar a conocer experiencias sobre la enseñanza de ELE. De esta manera, el estudio de la enseñanza de ELE en Colombia, se ha materializado teniendo en cuenta distintos enfoques: el educativo, el comunicativo, el lingüístico, entre otros.

Uno de los ámbitos más desarrollados es el educativo, ya que es la disciplina que se relaciona directamente con lo que se enseña. Se puede dar cuenta de esto, en investigaciones como: *Estado del arte de ELE en Colombia, una mirada holística: metodologías y enfoques, material didáctico, variedades lingüísticas y dificultades en el área* (García, et al 2001), *Estado de la enseñanza del ELE en Santander* (Quintero, Aviles y Suárez, 2013) y *Una aproximación a las experiencias didácticas en la enseñanza del español como lengua extranjera en la ciudad de Medellín* (Higueta, Martínez y Patiño, 2012). Estas investigaciones, tienen en común la importancia de desarrollar y estudiar la didáctica, es decir, la intervención en los procesos de enseñanza - aprendizaje y los métodos y técnicas que se implementan dentro de estos procesos.

Por otra parte, otro de los campos en el que se han desarrollado investigaciones sobre este tema, es el lingüístico. Allí, se pueden encontrar análisis como: *Variantes del español colombiano y su efecto en la enseñanza del ELE* (Areiza y Flórez, 2015), *Los estudiantes extranjeros y su proceso de aprendizaje del Español en Colombia: primeros indicios del translanguaje* (Nielsen, 2018) y *El registro lingüístico en el discurso literario colombiano para el fomento de la competencia sociolingüística en la enseñanza del español como lengua extranjera* (Piñeros, 2014).

Capítulo 3: Marco teórico

3.1. Conceptos sobre el léxico

3.1.1. Lexicología, léxico y vocabulario

La lexicología es la "disciplina que se ocupa del estudio, análisis y descripción de las palabras y de las unidades significativas que las componen, y de las relaciones que existen entre ellas" (Luna, et al., 2005, p.136).

Es pertinente hacer una distinción entre los términos léxico y vocabulario. Si bien ambos términos se relacionan entre sí, podría decirse que el léxico se refiere al sistema lingüístico, mientras que el vocabulario se relaciona directamente con el uso de la lengua. Justicia (1995) plantea que "el léxico hace referencia al sistema de palabras que componen una lengua" (p.16), mientras que "cuando esas unidades virtuales de la lengua se actualizan en el discurso del hablante reciben el nombre de vocablos y el conjunto de los mismos lo llamamos vocabulario" (p.16).

3.1.2. Palabra y vocablo

López (2010) plantea que la diferencia entre *palabra* y *vocablo* consiste en que "la primera, es el material gráfico comprendido entre dos espacios en blanco de un texto, y el segundo, son palabras diferentes que aparecen en un texto, sin contar las repeticiones" (p.17). Es decir, que la diferencia entre ambos conceptos puede ser entendida por el parámetro de repetición o no, tal y como lo expresa (Ávila, 2006, p.82), siendo palabra cada una de las formas repetidas o no con las que se trabaja y los vocablos son cada una de las palabras diferentes.

3.2. Lexicón mental y competencia léxica

Según Baralo (1997) el *lexicón mental* se puede representar dividido en tres niveles: en primer lugar, se encuentra una lista finita de morfemas, es decir, los datos de entrada de las reglas de formación de palabras; en segunda posición, las reglas de formación de palabras; luego, el

producto de estas reglas, que gracias a los procesos de formación de palabras, teóricamente puede estar formado por un número infinito de palabras, aunque en realidad el cerebro puede retener solo un conjunto de palabras “existentes”. Así, el lexicón mental permite la construcción, la comprensión y la capacidad de usar las unidades léxicas. Además de lo anterior, Baralo (1997) plantea que los hablantes no nativos poseen el mismo tipo de competencia léxica y gramatical que poseen los hablantes nativos.

3.3. Frecuencia de las palabras en una lengua

Cuando se habla de *frecuencia léxica*, se hace referencia al número de ocurrencias que tiene una palabra en una lengua. Ahora bien, según Alvar (2004) las cinco mil palabras más frecuentes de una lengua cubren el 90% de esta y permiten al aprendiente de una lengua extranjera tener buen manejo del idioma en una situación cotidiana, por lo tanto, es de gran utilidad conocer cuáles son las palabras más frecuentes de una lengua. La importancia de conocer la frecuencia de las palabras en una lengua, también reside en que es mucho más útil aprender una palabra altamente usada en una lengua que se está aprendiendo (Almela et al, 2005, p.15) que aprender un vocabulario específico sin tener todavía las bases suficientes de la lengua.

3.4. Léxico común, léxico básico y léxico fundamental

Según García Hoz (1953), el *léxico común* de una lengua está conformado por el vocabulario que un hablante debe utilizar de manera necesaria, por lo tanto, son las que se encuentran con mayor frecuencia en todas las situaciones comunicativas y distribuidas en todos los tipos de texto. De esta manera, el léxico común sirve para “mantener la estructura del mensaje, en su mayoría figuran palabras gramaticales y las que no lo son pertenecen, en orden de frecuencia, a verbos, adjetivos y sustantivos de significado general” (García, 1953, p.385), es decir, el léxico común es aquel considerado como más estable.

El *léxico básico* o también llamado *léxico usual* es definido por (Gómez, 2004) como

el conjunto de vocablos habitualmente utilizados en las situaciones de comunicación de la vida cotidiana. Este, abarca los vocablos más usuales de una comunidad, los cuales se caracterizan por un alto grado de estabilidad que les permite aparecer con mucha frecuencia y en todo tipo de discursos, independientemente de la temática de que se trate (p.38).

Es posible hablar del *léxico fundamental*, que es aquel formado por el léxico básico común y el léxico disponible (López, 1999). Esto, ya que el léxico fundamental recoge las palabras más estables de una lengua, mientras que el léxico disponible recoge las palabras que se usan en una situación comunicativa específica, así, ambos léxicos combinados forman el léxico fundamental que es el perfil léxico de una comunidad lingüística.

3.4.1. Léxico frecuente y léxico disponible

El *léxico frecuente*, puede ser entendido como el número de casos que se repite un vocablo en una medición o cuantificación (Gómez, 2004, p.37), de esta manera es aquél que recoge “las palabras de mayor estabilidad lingüística” (López, 1999, p.8) y el *léxico disponible* se define como “el que se usa en una situación comunicativa específica, actualizándose solo ante la necesidad de comunicar algo referente a esa situación concreta” (López, 1999, p.9). De esta manera, el léxico disponible se compone por las palabras que acuden primero a la memoria del hablante con relación a un tema concreto. Asimismo, Gómez (2004), plantea que el léxico disponible es el que, según las necesidades concretas en una producción lingüística, puede ser activado por el hablante.

3.5. Fórmulas para medir la disponibilidad léxica o el léxico disponible

A través de la historia, se han realizado diferentes propuestas para buscar calcular por medio de la léxico-estadística la disponibilidad léxica de los hablantes. Según Callealta y Gallego (2016), en los primeros estudios sobre el tema realizados a partir de 1956, se consideraba que la frecuencia de aparición de las palabras era el único factor ponderador de la disponibilidad (p.41). Así, en

estudios posteriores surgió el interés de buscar una fórmula matemática en la que se pudiera tener en cuenta tanto la frecuencia como la posición que ocupaban las palabras en las listas.

La primera de ellas fue la planteada por Lorán y López Morales (1983). Posteriormente, en el año 1987, Butrón presentó cuatro fórmulas matemáticas diferentes que permitieran establecer diferentes órdenes entre los centros de interés analizados. Sin embargo, según López (1999), las fórmulas mencionadas anteriormente “lograron controlar bastante el desajuste, pero no eliminarlo” (p.18) y esa precisión se logra solamente con el planteamiento de la fórmula de Strassburger y López en el año 1991 (p.18).

La fórmula propuesta por Strassburger y López (2000), se aplicó en la primera muestra escrita de este trabajo de investigación, es decir en la lista de palabras por cada centro de interés. Se seleccionó esta fórmula porque tiene en cuenta aspectos que se consideran importantes para la investigación, tales como la frecuencia con que fue dicha cada palabra en cada posición, la suma de esas frecuencias (frecuencia absoluta de la palabra), el número de informantes que participaron en la encuesta, el número de informantes que llegó a cada posición, el número de posiciones y el número de vocablos diferentes. Además, según la bibliografía consultada, esta fórmula es la utilizada en las investigaciones más recientes de DL y según López (1996) esta fórmula demostró superioridad porque logra producir una adecuación descriptiva sumamente plausible. Por último, es preciso señalar que este cálculo estadístico se realizará de manera automática con un *software* creado especialmente para esta investigación que trabaja a partir de la fórmula de Strassburger y López (2000). La fórmula es la siguiente:

$$D(P_j) = \sum_{i=1}^n e^{-c((i-1)/(n-1))} ((f_{ji}/I_1)$$

donde

- n = máxima posición alcanzada
- i = número de posición
- j = índice de la palabra
- f_{ji} = frecuencia absoluta de la palabra j en la posición i
- I₁ = número de informantes que participaron en la encuesta
- c = coeficiente de dispersión (se recomienda 2.3)
- D(P_j) = disponibilidad de la palabra j

Figura 1: Fórmula de disponibilidad léxica propuesta Strassburger y López (2010)

3.6. Riqueza léxica

Por otra parte, la riqueza léxica de un texto se obtiene según López (2010) considerando la cantidad de vocablos o unidades léxicas diferentes y el total de palabras de contenido nocional, es decir las unidades léxicas con contenido semántico como sustantivos, verbos y adjetivos. De esta manera, el léxico es uno de los factores determinantes para la organización del discurso y para la llamada calidad de la escritura. Tal y como lo plantea López (2010) diciendo que la calidad de la escritura “está integrada por una serie de factores, entre los que destacan: la riqueza léxica, la madurez sintáctica, los esquemas de cohesión y la coherencia discursiva. La amplitud del vocabulario está apoyado en la DL del hablante” (p.15)

3.6.1 Fórmulas para medir la riqueza léxica

A través de la historia, se han realizado diferentes propuestas para buscar calcular por medio de la léxico-estadística la riqueza léxica de los textos. Según López Morales (2010), las fórmulas iniciales fueron planteadas por Giraud en el año 1954 quien tomó en consideración dos tipos de palabras: nocionales o palabras con contenido semántico y gramaticales o palabras funcionales del discurso y para cada tipo de palabras propuso una fórmula diferente asumiendo que las palabras nocionales representaban la mitad del texto. Posteriormente, Tesitelová (1992) propuso una fórmula en la que se recogían nuevos aspectos como la repetición de palabras de un texto, la

fuerza de las zonas de palabras en baja frecuencia, la dispersión del vocabulario y también su concentración.

Otra de las propuestas más conocidas es la presentada por Ávila (1986), quien propone tres procedimientos comparativos, que consisten en el número de vocablos recogidos en total, la densidad promedio para cien palabras y el número de palabras acumuladas por deciles.

Por último, es pertinente presentar las dos fórmulas realizadas por López Morales (2010), que fueron las aplicadas en la segunda muestra escrita, es decir, en el texto expositivo escrito por los estudiantes. López Morales propone dos fórmulas interrelacionadas que representan dos medidas obtenidas a base de cálculos empíricos. Por una parte, toma en cuenta el porcentaje de vocablos (PV) del total de palabras de un texto (N), y por otra parte toma el intervalo de aparición en el texto de palabras de contenido semántico nocional (IAT). Con palabras nocionales (PN) se refiere a aquellas unidades léxicas con contenido semántico. La riqueza léxica se obtiene aquí al considerar la cantidad de vocablos o unidades léxicas diferentes y el total de palabras de contenido nocional (PN). López Morales (2010) plantea que el primer cálculo que se realiza es el que determina el porcentaje de vocablos (PV), así, el procedimiento requiere que se divida el total de vocablos (V) entre el total de las unidades léxicas comprendidas en el texto (N) y luego se multipliquen por 100, de la siguiente manera:

$$PV = \frac{V \times 100}{N}$$

aparición

Morales (2010) propone la siguiente fórmula:

$$IAT = \frac{N}{PN}$$

Figura 2: Primera parte de la fórmula de riqueza léxica

Posteriormente, para cuantificar el intervalo de palabras de contenido nocional (IAT), López

Figura 3: Segunda parte de la fórmula de riqueza léxica

Para esta investigación, se utilizó la propuesta para medir la Riqueza Léxica de López Morales (2010), ya que como él lo plantea, el resultado de esta operación matemática refleja cifras relacionadas con la proporción de palabras nocionales en el texto. Esto es, a mayor número de palabras nocionales, menor es el intervalo, lo que se interpreta como mejor índice de riqueza léxica. Por lo anterior, se considera que es una fórmula pertinente para lo que se buscaba observar en la segunda muestra escrita. Además, según la bibliografía consultada, no hay propuestas más recientes con relación a este tema y la formulada por López Morales (2010) es una de las más utilizadas en los estudios lingüísticos de riqueza léxica en ELE.

3.7. Marco Común Europeo

En el año 2001, el Consejo de Europa publicó el *Marco Común Europeo de Referencia para las lenguas: aprender, enseñar, evaluar* (MER), el cual consiste en un documento que busca establecer los elementos comunes que se pretenden alcanzar en las diferentes etapas del aprendizaje de todas las lenguas que se hablan en los países europeos. Además de eso, este documento describe las capacidades lingüísticas, así como los saberes que se deben desarrollar en cada uno de los niveles de lengua extranjera para afrontar las situaciones y los diferentes contextos comunicativos (Clouet, 2010,). En el Marco Común Europeo de Referencia (2002), se propone una diversificación de las metodologías comunicativas, teniendo en cuenta la diversificación de los recursos y los currículos, así como la diversidad de los aprendices, los cuales son considerados como los sujetos de su propio aprendizaje.

Es posible encontrar muchos docentes e investigadores en el área de ELE que consideran el MER como una herramienta positiva y útil para el campo, tal y como lo plantea Clouet (2010) “consideramos el MER como un enfoque metodológico integrador y equilibrado para la

enseñanza de lenguas, que redefine la competencia comunicativa como una competencia compleja que toma en cuenta los siguientes parámetros” (p.73).

En el Marco Común Europeo (2002) se define la competencia léxica como “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo y se compone de elementos léxicos y elementos gramaticales” (p.58). Sin embargo, según Baralo (2005) el tratamiento conceptual que se le da al léxico en el MER es muy superficial, se dejan por fuera muchos elementos importantes y además de eso no se trata de manera explícita, sino que se incluye en las demás competencias.

Con respecto a los niveles de dominio de la competencia léxica en el Marco Común Europeo (2002) se definen según las situaciones comunicativas, es decir, al nivel A1 le corresponde el repertorio léxico de situaciones concretas; al A2 el léxico de los entornos cotidianos y habituales, al B1 una ampliación de la vida diaria, con temas relacionados con el trabajo, las aficiones e intereses, los viajes y los hechos de actualidad; al nivel B2 le corresponde los asuntos relacionados con su especialidad.

Además de eso, en el MER (2002) se establecen algunas características sobre el dominio del vocabulario por parte de los aprendientes según los niveles, planteando que en el nivel A1 no hay descriptor disponible, en el A2 hay un dominio limitado de vocabulario referente a necesidades concretas y cotidianas, en el B1 los estudiantes manifiestan un buen dominio del vocabulario elemental, pero todavía cometen errores importantes cuando expresa pensamientos más complejos, en el B2 la precisión léxica es generalmente alta, aunque se presentan algunas confusiones o incorrecciones al seleccionar las palabras.

Si bien Colombia no pertenece a los países del Consejo europeo, la mayoría de las instituciones que actualmente se dedican en el país a la enseñanza de español como lengua extranjera, siguen el modelo del Marco de Referencia europeo, mientras que otras instituciones proponen nuevos enfoques y estrategias de enseñanza (Espejo, et al, 2011).

3.8. ELE en la virtualidad

En la actualidad, la aparición de las Nuevas Tecnologías ha permitido que diferentes aspectos de la vida cotidiana se trasladen al plano virtual. Un ejemplo de esto es la transformación que han tenido las aulas de clase, ya que las exigencias de la digitalización las ha mutado, para convertirlas en aulas virtuales.

Según Villanueva y Perdrix (2011) en su investigación sobre *El aula virtual en la enseñanza online de una lengua extranjera: tipos y materiales*, la enseñanza de las lenguas extranjeras, se ha ido transformando, ya que se han implementado lo que se conoce, en la actualidad, como clases online, siendo estas una herramienta fundamental para el aprendizaje de una segunda lengua. Esta modalidad es relativamente nueva y aunque todavía le falta camino para seguir afianzándose, ha traído como consecuencia la existencia de muchos programas en el mercado que muestran ofertas y novedosas metodologías para la realización de procesos de enseñanza de una segunda lengua.

Además, esto ha desarrollado consecuencias positivas, algunas de ellas según Kumar (2014) son la oportunidad de tener aprendizajes inmensurables y el fomento del aprendizaje autónomo de los estudiantes, asimismo, permite la formación de aprendientes desde cualquier parte del mundo.

Por otra parte, alrededor del mundo existen diferentes estudios sobre la enseñanza de ELE en la virtualidad. Entre ellos, se pueden encontrar: *El potencial de la realidad aumentada en la enseñanza de español como lengua extranjera* (Sánchez, 2016) y *Enseñanza virtual de ELE: del estudio empírico a la aplicación didáctica* (Agray y Baralo, 2020). Estos análisis son los encargados de dar indicios sobre el auge que tiene el ámbito virtual en la enseñanza de ELE, mostrando con claridad que se ha convertido en un espacio indispensable para ir más allá de las barreras espacio temporales que se presentan.

Capítulo 4: Metodología

4.1. Paradigma y enfoque de investigación

La investigación se realizó desde el paradigma positivista, a través del cual se puede estudiar la realidad a partir de una parte de ella. En este caso, dentro del paradigma positivista comporta que el investigador se diferencia y se aparta del objeto de investigación en un intento por alcanzar la objetividad, así, el investigador puede ser observador neutral, por lo tanto, no afecta a los resultados de la investigación (Kerlinger, 2002).

Por otro lado, en el paradigma positivista, podemos encontrar tres *enfoques*, a saber, el cuantitativo, el cualitativo y el mixto. Dado que esta investigación busca ser objetiva sin llegar a generalizar los resultados, se enmarca en un enfoque metodológico cualitativo. Además de esto, el fin prioritario de la investigación cualitativa consiste en cualificar y describir el fenómeno a partir de los rasgos y variables del objeto que se está estudiando, tal y como se planteó para esta investigación (Bonilla, E; Rodríguez, P. 2005).

La razón de este enfoque es que no se siguieron métodos estandarizados para la recolección de las muestras, sino que ellas fueron seleccionadas siguiendo los procedimientos característicos de un muestreo intencionado, llevado a cabo según el criterio del investigador sin controlar las variables experimentalmente (Hernández et al., 2006). De igual manera, no se busca llegar a inferencias generalizables, ya que dicha significancia solo se da para la población de la cual se toma la muestra y no para una población mayor, por ejemplo, un número de muestras representativas de una universidad, todas las universidades de Medellín o del país, que se dedican a la enseñanza de Español como Lengua Extranjera (Hernández et al., 2006).

Por otro lado, es preciso señalar que, si bien el enfoque es cualitativo, se va a hacer uso de estadística descriptiva y pruebas de hipótesis no paramétricas para la obtención y manejo de datos,

pero para interpretarlos se hará uso principalmente del lenguaje natural. Esto, puesto que el interés principal reside en el sentido y la interpretación de situaciones que va más allá del *simple* procesamiento de datos numéricos (Rodríguez, Gil & García, 1999; Campos, 2009).

4.2. Alcance de la investigación: exploratorio descriptivo

El alcance de la investigación presentada es exploratorio-descriptivo (Vieytes, 2004; Hernández et al., 2006), en este sentido, es exploratorio porque según la bibliografía encontrada no hay trabajos realizados sobre la Disponibilidad y Riqueza Léxica de estudiantes de Español como Lengua Extranjera en el país, ni específicamente en la ciudad Medellín. De esta manera, uno de los propósitos de la investigación es examinar un problema de investigación que no se ha abordado antes en el ámbito local, que es el objetivo asignado para los estudios de corte exploratorio (Hernández et al., 2006). Por otra parte, el alcance también es de corte descriptivo porque en la investigación se busca realizar una descripción de la Disponibilidad y Riqueza Léxicas de estudiantes aprendientes de ELE de niveles A2 y B2, de esta manera se va a presentar cómo es y cómo se manifiesta el fenómeno lingüístico (Hernández et al., 2006).

4.3. Diseño de la investigación: no experimental o *ex postfacto*

En la presente investigación, se busca observar una situación a partir de la recolección de datos sin manipular las variables, ya que solo se observa su comportamiento. En este caso particular, no se manipula el nivel que tienen los estudiantes a través de entrenamientos, solo se observa el nivel de español que ya traen. Dentro de los diseños de investigación, se optó por el no experimental toda vez que no se manipularon ni se controlaron variables con el fin de encontrar respuestas a dicha manipulación. En la investigación propuesta, las variables tenidas en cuenta son variables discretas o discontinuas pues tienen valores enteros.

4.4. Variables de la investigación

A continuación, se presenta el sistema de variables que sustenta esta investigación:

- **Nivel de español:** en esta investigación se tomaron el nivel de Español como Lengua Extranjera y un grupo control con un nivel de Español como lengua Materna (nivel nativo). En el primer caso, se toman muestras de aprendientes de los niveles A2 y B2, en segundo lugar, se toman hablantes nativos de español. La intención con esta variable es hacer una comparación entre niveles de aprendientes (A2 y B2), y entre hablantes aprendientes de ELE y nativos.

- **Centros de interés:** para las investigaciones sobre disponibilidad léxica se tienen establecidos dieciséis centros de interés, pero para esta investigación, se tendrán en cuenta solamente tres: la ciudad, alimentos-bebidas y medios de transporte de ellos. Esto, en la medida en que se quiere profundidad de estos, así como la vinculación de la muestra de los centros de interés con un texto escrito de tipo expositivo en el que se incluyan los mismos centros de interés. Por lo anterior, se seleccionaron tres centros de interés, los cuales además según el Marco Común Europeo (2002) hacen parte importante del proceso inicial de aprendizaje de la lengua.

- **Modo de organización discursiva:** el texto que se asignará para la escritura tanto de aprendientes de ELE como de hablantes nativos es de carácter expositivo, según el Marco Común Europeo (2002), esta tipología textual empieza a ser explorada por los aprendientes de ELE desde el nivel inicial A1.

- **Variables no consideradas en la investigación:** si bien según el Marco Común Europeo (2002) existe una estandarización de la competencia léxica de los aprendientes según los niveles, existen otras variables que no fueron controladas en la presente investigación y que pudieron influir en la distribución de los datos.

Una de las variables no consideradas fue la lengua materna de los aprendientes de ELE, que según Agustín (2017), es una de las variables más influyentes en el aprendizaje del léxico de una lengua. De esta manera, diferentes lenguas maternas pueden generar diferentes dificultades en el aprendizaje del léxico, así como características diferentes en su selección y su producción; por ejemplo, los hablantes cuya lengua materna es alemán o inglés pueden tener mayor conocimiento de palabras con origen anglosajón, mientras que los hablantes cuya lengua materna es el francés pueden tener mayor conocimiento de palabras del español cuyo origen es grecolatino (Jarvis, 2009), ya que la percepción de similitud léxica con relación a la propia lengua favorece el aprendizaje del léxico en una lengua extranjera (Ringbom, 2006).

Por otra parte, otra de las variables no consideradas en esta investigación y que también influye en el aprendizaje de las lenguas y por lo tanto del léxico, es la edad. Si bien se considera que los adultos tienen un nivel mayor de consciencia metalingüística y demuestran mayor diversidad léxica que los niños o los jóvenes, en los adultos de mayor edad el aprendizaje puede ser más lento (Celaya y Torras, 2001). Finalmente, entre las variables no consideradas también se encuentra el tiempo que los aprendientes de ELE llevan estudiando la lengua ni tampoco el contacto del aprendiente con la lengua fuera de las aulas.

Es importante tener en cuenta que estas variables no se controlaron por el tamaño de las muestras, ya que al estandarizarlas se quedaría con una muestra muy pequeña y pasaría a ser un estudio de caso.

4.5. Muestra

La población de la cual se tomó la muestra estaba constituida por 18 estudiantes de Español como Lengua Extranjera que en el momento de la prueba se encontraban realizando un curso de A2 y 14 estudiantes que se encontraban realizando un curso de B2. Se tomaron estos niveles sobre la

base del supuesto de que hay más diferencias entre ellos que entre niveles contiguos, como por ejemplo entre A1 y A2 o A2 y B1, en cualquier caso, para esta variable también prima el criterio del investigador. Esta muestra fue tomada de una universidad privada y un instituto que tienen reconocimiento en el área de enseñanza del Español como Lengua Extranjera en la ciudad de Medellín. Además, se tomaron 23 hablantes nativos de español que estudian Licenciatura en Literatura y Lengua Castellana en la Universidad de Antioquia, que es una universidad pública con reconocimiento en el ámbito nacional e internacional por su nivel académico e investigativo.

4.6. Instrumentos

Para llevar a cabo el estudio, cada estudiante, tanto aprendientes de ELE como hablantes nativos realizaron dos muestras escritas. Cabe señalar que estos instrumentos fueron validados por cinco docentes expertos en las áreas de lingüística, lexicografía y Didáctica de la lengua. La primera parte del instrumento correspondió a un espacio para completar datos personales de los estudiantes: nombre, edad, lengua materna, lenguas extranjeras que habla y profesión. A continuación, se describen cada uno de los instrumentos:

- Muestras escritas

- Para la primera muestra escrita, que era el instrumento para recoger los datos para disponibilidad léxica, cada estudiante recibió la indicación de escribir una lista de palabras relacionadas con cada centro de interés seleccionado (la ciudad, alimentos-bebidas y medios de transporte) en un tiempo máximo de dos minutos por centro de interés. Esto lo realizaron tanto los aprendientes de ELE como el grupo control. Algunas muestras fueron recogidas de manera presencial y otras de manera virtual, sin embargo, en las dos modalidades se hizo una aplicación rigurosa del instrumento, contabilizando el tiempo y dando las indicaciones para que todos realizaran el ejercicio de manera sincrónica.

- La segunda muestra escrita, es decir, la muestra para riqueza léxica, consistió en un texto expositivo en relación con los centros de interés seleccionados para esta investigación (la ciudad, alimentos-bebida y medios de transporte). Este texto lo realizaron tanto los aprendientes de ELE como los hablantes nativos de español. Para esto, los estudiantes recibieron la indicación de escribir un texto de no más de 150 palabras y contaron con máximo quince minutos, estos criterios fueron escogidos siguiendo los parámetros de certificación de nivel que se emplean en las pruebas llevadas a cabo por el Diploma de Español como Lengua Extranjera de los niveles A2 y B2, DELE, en los que cada tarea de escritura cuenta con un máximo de 15 minutos y 150 palabras. El enunciado del texto a escribir fue el siguiente:

“Un amigo que no conoce tu ciudad quiere visitarte. Escríbele un correo donde le hables sobre lo que más te gusta de tu ciudad (tus lugares favoritos, la comida, el clima, etc), recomiéndale qué hacer cuando te visite. Para esto tienes 15 minutos. (150 palabras aproximadamente)”.

4.7. Validación de instrumentos: centros de interés e instrucción texto escrito.

La validación de los instrumentos la realizaron cinco docentes investigadores en el área de lingüística y Didáctica de la lengua. Para esta validación se utilizó una escala Likert de cinco valores, en la que los ítems se evaluaron de la siguiente manera: 1 como totalmente en desacuerdo (TA), 2 en desacuerdo (ED), 3 ni en desacuerdo ni de acuerdo (NI), 4 de acuerdo (DA) y 5 como totalmente de acuerdo (TD). En esta escala, se encontró un valor de aprobación del 88% para el instrumento de disponibilidad léxica y un valor de 72% para el instrumento de riqueza léxica, estos valores fueron tomados como valores aprobatorios, ya que se considera que hay acuerdo si hay un valor mayor a 60%.

Además, los instrumentos fueron validados a través del coeficiente *Kappa de Fleiss* que permite identificar el nivel de acuerdo entre evaluadores, en el que los valores varían de -1 a +1, mientras

más alto es el valor mayor es la concordancia y un valor de al menos 0.75 indica una concordancia adecuada. En este caso, se encontró un nivel de concordancia de 60% y -0.25 en el instrumento de disponibilidad léxica puesto que uno de los cinco evaluadores estuvo en desacuerdo. En el instrumento de riqueza léxica se encontró un nivel de acuerdo del 20% y -0.25 entre los evaluadores, ya que dos evaluadores estuvieron en desacuerdo mientras que los otros tres estuvieron de acuerdo. Frente a esto, se hicieron los ajustes y correcciones sugeridas por los evaluadores para precisar el instrumento. Después de validado el instrumento se procedió a su aplicación.

4.8. Software

Para el análisis de los datos se crearon dos *software*, uno para disponibilidad léxica y otro para riqueza léxica. Estos aplicativos fueron construidos en lenguaje Python por un experto en ello, pero bajo las indicaciones matemáticas y procedimentales de la investigadora de este proyecto.

El *software* creado para analizar la disponibilidad léxica con la fórmula de Strassburger y López (2000) permite ingresar la lista total de palabras en *Excel* por cada uno de los centros de interés y este hace el cálculo automático ordenando las palabras en orden descendente desde la palabra más disponible. Además de eso, en el software es posible ingresar manualmente las palabras que no se van a tener en cuenta en el análisis como interferencias lingüísticas, palabras repetidas, errores léxicos, etc. A continuación, se presenta un pantallazo de la interfaz del software y un ejemplo de cómo son organizados automáticamente los resultados:

The screenshot shows a software interface with three input fields: 'Nombre de archivo xlsx', 'Nombre de campo', and 'Nombre de archivo Exportación'. Below these are three yellow buttons: 'Generar reporte', 'Borrar', and 'Salir'. To the right is a table with two columns: '0' and '1'. The table lists 15 items with their corresponding values in column '1'.

	0	1
0	Palabra	Disponibilidad
1	Agua	0,493
2	Cerveza	0,4634
3	Arepa	0,2667
4	Leche	0,2414
5	Pollo	0,2137
6	Café	0,2117
7	Pan	0,2028
8	Aguardiente	0,1237
9	Arroz	0,1145
10	Huevos	0,1135
11	Jugo	0,1128
12	Naranja	0,1056
13	Té	0,0946
14	Fríjoles	0,0938
15	Queso	0,092

Figura 4: Pantallazo del software y muestra de los resultados

El ejemplo presentado corresponde al centro de interés de comidas y bebidas de A2, como se observa, en primer lugar, se encuentra la palabra más disponible de la muestra, es decir *agua* con su valor correspondiente, y así sucesivamente.

Por otra parte, para realizar el análisis de riqueza léxica de los textos expositivos a partir de la fórmula de López Morales (2010), se creó un software también en lenguaje *Python*. En este caso, es necesario ingresar los textos escritos por los estudiantes en formato *.txt* para calcular de manera automática todos los valores de la fórmula. Además de eso, en el *software* es posible introducir manualmente las unidades fraseológicas para que sean contadas como una sola y también se pueden excluir algunas palabras que no se van a tener en cuenta para el conteo como interferencias lingüísticas y errores léxicos. La siguiente, es una muestra de la interfaz y un ejemplo de cómo quedan organizados los datos automáticamente:

The screenshot shows a software interface with two input fields. The first field is labeled 'Ingrese las diferentes unidades fraseológicas y terminológicas separadas por espacio'. The second field is labeled 'Ingrese los términos que desea apartar de los resultados (anglicismos o palabras en otro idioma, palabras inventadas, etc) separados por espacios'. Below the fields is a yellow button labeled 'ANALIZAR ARCHIVO'. At the bottom left is the logo of the 'UNIVERSIDAD DE ANTIOQUIA Facultad de Comunicaciones'. To the right of the interface is a table with the following data:

	1.Nombre	2.N	3.V	4.PV	5.IA
0	AntonioNapoletani	65	32	49,23077	2,03125
1	AurelienAchard	128	54	42,1875	2,37037
2	ChristopherCropper	142	65	45,77465	2,184615
3	EdwardWalls	71	38	53,52113	1,868421
4	Francis	97	46	47,42268	2,108696
5	GaryWood	98	46	46,93878	2,130435
6	JeromeBollier	100	50	50	2
7	Luba	185	78	42,16216	2,371795
8	Manling	66	37	56,06061	1,783784
9	RachelThom	110	60	54,54545	1,833333

Figura 5: Pantallazo del software y muestra de los resultados

En el ejemplo anterior, se muestran los resultados de diez aprendientes de A2 en el que en cada uno de los textos se observa el porcentaje de vocablos (PV) del total de palabras de un texto (N), el total de vocablos o palabras con contenido semántico del texto (V) y el intervalo de aparición en el texto de palabras de contenido semántico nocional (IAT).

4.9. Pruebas estadísticas para análisis de datos: distribución normal y test no paramétricos

Después de aplicar el instrumento, se tabularon los datos en *Excel*, para calcular el promedio, la varianza, la desviación y la media del total de palabras escritas por los estudiantes de cada uno de los niveles, así como en cada centro de interés de cada uno de los grupos.

Posteriormente, se calculó la distribución normal de los datos. La distribución normal, también conocida como *La campana de Gauss*, hace referencia a la simetría que hay alrededor de la media de un grupo de datos, de esta manera, en un grupo de datos con distribución normal, la media tiene un valor muy próximo o coincide con la mediana, es decir que los datos deben estar distribuidos simétricamente en la muestra. (Martínez, M; Marí, M. 2016). Cuando se dice que las distribuciones de los datos no tienen una distribución normal, eso hace referencia a que las desviaciones sucesivas con relación a la media no son normales o paramétricas y hay mucha variación dentro de un mismo grupo de datos (Mayén, 2009).

En los datos de esta investigación se encontró que estos están distribuidos asimétricamente y con gran variabilidad, siendo los del grupo control los que más se acercan a tener una distribución normal. Por lo tanto, al no tener datos con una distribución normal se decidió hacer la comparación de ellos a través de pruebas no paramétricas.

En un primer momento, se realizó la prueba de Kruskal-Wallis entre todos los cruces posibles. Esta prueba es la alternativa no paramétrica al test ANOVA para datos simétricos (Amat, 2016), en la que se tiene como factor principal de análisis a las medianas de las muestras y permite analizar datos que no cuentan con una distribución normal. Si bien la prueba Kruskal-Wallis expresa si hay o no diferencias estadísticamente significativas entre tres o más grupos o muestras, no indica entre cuáles de ellos se encuentran esas diferencias (Amat, 2016). Por este motivo, fue

necesario realizar el test Mann-Whitney, que es la alternativa no paramétrica al T-test (Amat, 2017) y el cual permite contrastar dos muestras con datos independientes.

4.10. Etapas de la investigación

ETAPA 1: Determinación de la población

ETAPA 2: Determinación de los grupos para la muestra

ETAPA 3: Validación de instrumentos

ETAPA 4: Ajuste de instrumentos

ETAPA 5: Aplicación de los instrumentos

ETAPA 6: Obtención de los datos a partir de las fórmulas matemáticas.

- Disponibilidad Léxica de aprendientes A2 y B2, y de hablantes nativos a través del software.
- Riqueza Léxica de aprendientes de A2 y de B2, y de hablantes nativos a través del *software*.

ETAPA 7: Comparación de variables

- Disponibilidad Léxica entre aprendientes de A2 y B2.
- Disponibilidad Léxica entre estudiantes de A2 y B2 y hablantes nativos.
- Disponibilidad Léxica (lista centros de interés) y Riqueza Léxica (texto expositivo), de los aprendientes de ELE y los hablantes nativos.

ETAPA 8: Análisis y discusión de resultados.

Capítulo 5: Resultados y discusión

5.1. Disponibilidad léxica

5.1.1. Selección del léxico analizado

En las listas de palabras referentes a los tres centros de interés analizados en esta investigación, se encontró que en el nivel A2, cuya muestra contó con 18 estudiantes se escribieron un total de 749 palabras. Sin embargo, en ese total de palabras se encontraron 86 que no se tuvieron en cuenta en la investigación, puesto que correspondían a errores léxicos, interferencias léxicas o a palabras que si bien se relacionaban con los centros de interés no hacían parte de ellos. Por lo tanto, se excluyó el 10% de los datos e hicieron parte del análisis 673 palabras. En el caso de B2, cuya muestra contó con 14 estudiantes se escribieron un total de 769 palabras, de las cuales no se tuvieron en cuenta 94 palabras por los mismos motivos anteriormente descritos, es decir que se excluyó el 12% de los datos. Por lo tanto, hicieron parte del análisis 668 palabras.

Por su parte, en el grupo control, en la cual se contó con 23 estudiantes, se escribieron un total de 1276 palabras de las cuales no se tuvieron en cuenta 93 palabras, es decir el 7% de la muestra, porque las palabras estaban repetidas en la misma lista de un estudiante o si bien se relacionaban con los centros de interés no hacían parte de ellos. Por lo tanto, se trabajó con un total de 1183 palabras. La siguiente tabla representa la distribución del total de palabras seleccionadas:

Tabla 1

Total de palabras seleccionadas para el análisis

	Comidas y bebidas	Lugares de la ciudad	Medios de transporte
A2	280	222	171
B2	290	229	148
Grupo control	509	342	332

5.1.2. Léxico no tenido en cuenta

- Errores léxicos

Un error léxico según Dulay (1982) es cualquier desvío con relación a la lengua meta, es decir, escribir una unidad léxica de manera incorrecta según las reglas lexicales de la lengua meta, que en este caso es el español. En el nivel A2 se encontró un error en lugares de la ciudad con la palabra *pensino* y en el nivel B2 se encontraron dos errores en alimentos y bebidas que fueron *smarten* y *monitoricé* y un error en medios de transporte con la palabra *troque*. Mientras que en el nivel nativo no se encontró ningún error léxico.

Es decir, que, de un total de palabras escritas por aprendientes de ELE de 1518, se encontraron solamente cuatro errores consistentes en *errores de forma equivocada*, o sea palabras que no existen en la lengua meta (Ringbom, 2001).

- Interferencias de otras lenguas

En el nivel A2 fue posible encontrar cinco palabras de préstamo lingüístico, cuatro de ellas del inglés que fueron *steak*, *library*, *zoo* y *train*; y una palabra del italiano, *strada*. Además, se encontró una acuñación (Ringbom, 2001), con la unidad léxica *jugo de avocado*. Mientras que en el nivel B2 solamente se encontró un préstamo del inglés que fue *lemon*. Es importante tener en cuenta que, para la mayoría de los aprendientes participantes en esta investigación, el inglés era su lengua materna o su L2, por lo que la mayoría de interferencias lingüísticas son de este idioma.

Finalmente, en el grupo de estudiantes nativos no se encontró ninguna interferencia lingüística, pero sí el uso de algunas palabras provenientes del inglés que actualmente hacen parte del uso cotidiano del español como *smoothie*, *brownie*, *pancake*, entre otras.

- Palabras no pertenecientes al centro de interés

Ahora bien, la mayoría de las palabras excluidas en esta investigación fueron palabras que se relacionaban con el centro de interés pero que no pertenecían a él. En los tres grupos analizados en esta investigación la mayoría de los casos se encontraron en los medios de transporte, en donde

los estudiantes escribieron palabras relacionadas con el desplazamiento o la movilización pero que no eran un medio de transporte, como: *caminar, autopista, tráfico, correr, piernas, pies, a pie, zapatos, conducir, gasolina, pasaje, caminata, congestión, dirección, recorrido, vagón, torniquete*, entre otras.

Además de eso, en los tres grupos analizados se encontraron muchos casos en el centro de interés de alimentos y bebidas como: *comer, cocinar, cuchara, tenedor, restaurante, vaso, plato, mesa, bar*, entre otras.

En este caso, se puede hablar entonces de un mecanismo de asociación léxica en el que se crean redes semánticas (Lehmann, 1992) y se relacionan ciertas palabras a partir de su significado o lo que representan para el hablante, por ejemplo, si bien *cuchara* o *tenedor* no son alimentos, son elementos que se usan para ingerir los alimentos, por lo que pueden ser activados en el lexicón mental.

5.1.3. Total de palabras y palabras diferentes

En el caso de A2, si bien en total los aprendientes escribieron 673 palabras, de ellas 447 fueron repetidas y 226 fueron diferentes. En el caso de B2, se escribieron 668, de las cuales 415 fueron repetidas y 253 palabras fueron diferentes. Finalmente, los pertenecientes al grupo control escribieron 1183 palabras, de las cuales 717 fueron repetidas y 466 fueron palabras diferentes.

En la siguiente tabla se presentan los datos mencionados anteriormente:

Tabla 2

Total de palabras, palabras repetidas y palabras diferentes.

	Total de palabras	Palabras repetidas	Palabras diferentes
A2	673	447	226
B2	668	415	253
Grupo control	1183	717	466

En la siguiente tabla se presenta la distribución por cada centro de interés del total de palabras diferentes:

Tabla 3

Distribución de las palabras diferentes

	Comidas y bebidas	Lugares de la ciudad	Medios de transporte
A2	104	81	41
B2	119	95	39
Grupo control	192	197	77

5.1.4. Promedio, varianza, desviación y mediana

De cada una de las listas de datos se calculó la media, es decir, el valor más representativo que identifica las medidas de tendencia central que describen el valor típico en un grupo de datos (Mayén, 2009). La media más alta de palabras lo tuvo el grupo control con un valor de 51,43 mientras que el promedio de palabras más bajo lo tuvo el nivel A2 con 37,9. Los valores de esta media son evidentes además porque en el grupo control el máximo de palabras escritas por un estudiante fue de 76 mientras que en A2 fue de 54, y el mínimo de palabras escritas por un estudiante nativo fue de 25 mientras que de un estudiante de A2 fueron 10 palabras, por lo que es claro que la tendencia media de ambos grupos sea diferente. Por otra parte, la media de palabras escritas por el grupo de B2 si bien dista del de A2 se acerca a la media de palabras escritas por el grupo control con un valor de 47,71.

Además del promedio, se calculó la varianza de cada uno de los grupos de datos, es decir, la variabilidad de los datos con relación a su media (Mayén, 2009). En esta investigación, la varianza más alta la tuvo el grupo de hablantes nativos con un valor de 178,44 y la más baja la tuvo el grupo de B2 con 94,99, es, lo que quiere decir que en el grupo control varió más la cantidad de palabras escritas por los estudiantes. Esto se evidencia además en la desviación estándar de cada uno de los grupos, es decir, la medida de dispersión de los datos en relación a su media (Mayén, 2009), ya que los del grupo control están más dispersos con un valor en desviación de 13,36,

mientras que A2 tiene una desviación de 12,65 y B2 de 9,74. Además de eso, se realizó el cálculo del promedio, la varianza, la desviación y la mediana de cada uno de los centros de interés por cada uno de los niveles.

Tabla 4

Promedio, varianza, desviación y mediana por cada nivel.

	A2	B2	Grupo control
Promedio	37,39	47,71	51,43
Varianza	160,02	94,99	178,44
Desviación	12,65	9,74	13,36
Mediana	40	48	48

Tabla 5

Promedio, varianza, desviación y mediana por cada centro de interés.

	Comidas y bebidas			Lugares de la ciudad			Medios de transporte		
	A2	B2	Control	A2	B2	Control	A2	B2	Control
Promedio	16,47	20,71	22,13	12,33	16,43	14,87	9,5	10,57	14,43
Varianza	42,97	31,45	53,20	26,59	20,72	28,30	7,68	7,18	14,71
Desviación	6,55	5,61	7,29	5,16	4,55	5,32	2,77	2,68	3,83
Mediana	17	21,5	22	13	17,5	14	9	10	14

Según Blanco y Ferreira (2018), los tres centros de interés escogidos para esta investigación hacen parte del vocabulario básico de cualquier aprendiente de ELE. Por su parte, los alimentos y las bebidas son un tema de primera necesidad para los extranjeros que aprenden una lengua, ya que deben identificarlos para decidir sobre su consumo, por otra parte, los lugares de la ciudad hacen parte de la cotidianidad de los extranjeros, quienes se desplazan a través de ella y finalmente, los medios de transporte son un tema de interés relevante para los aprendientes, puesto que necesitan hacer uso de ellos para su desenvolvimiento académico y personal (Ferreira y Blanco, 2018).

Sin embargo, a pesar de que los tres centros de interés hacen parte del vocabulario básico de los aprendientes de ELE o de un hablante nativo, fue posible encontrar diferencias entre ellos. Para

los tres niveles analizados, las comidas y las bebidas fue el centro de interés con un promedio de palabras mayor, en A2 con 16,47, en B2 con un valor de 20,71 y en el grupo control con un promedio de 22,13 palabras. Mientras que el centro de interés de los medios de transporte fue el que tuvo promedios más bajos en las tres muestras, con un valor de 9,5 en A2, 10,57 en B2 y 14,43 en el grupo control.

Ferreira y Blanco (2018), encontraron este mismo comportamiento de los promedios en su trabajo de investigación con un grupo de aprendientes de ELE de nivel B1, y las autoras expresan que esto se debe a que de los tres centros de interés el de la alimentación es un tema de primera necesidad más que los otros porque además a través de ese tema los estudiantes pueden hablar de sus propios gustos. Además de eso, los ítems léxicos en español con relación a los medios de transporte son más limitados, con relación a la cantidad de alimentos y bebidas que se pueden llegar a conocer (Ferreira y Blanco, 2018). Asimismo, en el caso de los aprendientes de ELE estos dos centros de interés fueron los que tuvieron mayor número de palabras diferentes.

5.1.5. Distribución normal

Al realizar el cálculo de la distribución normal, también conocida como *La campana de Gauss* se encontró que los datos no tenían una distribución normal. Además, al graficar la distribución de los datos se evidenció que la gráfica no era simétrica a ambos lados de la media, siendo el grupo control la que más se acercaba a tener una distribución normal sin alcanzarla completamente. A continuación, se presenta la cantidad de palabras escritas por cada uno de los aprendientes en los diferentes niveles:

Tabla 6

Total de palabras disponibles por cada estudiante.

Estudiante	A2	B2	Grupo control	Estudiante	A2	B2	Grupo control
1	52	61	76	12	36	40	48

2	52	60	70	13	35	38	48
3	49	58	69	14	35	27	48
4	49	56	66	15	24		46
5	47	52	65	16	19		46
6	45	52	63	17	15		45
7	44	52	62	18	10		42
8	43	44	61	19			39
9	41	44	56	20			37
10	39	43	54	21			36
11	38	41	49	22			32
				23			25

La distribución no normal de los datos pudo ser ocasionada por la cantidad de variables que no se controlaron. En el caso de la variable no controlada lengua materna, en A2 se contó con hablantes nativos de inglés, francés, mandarín, árabe, alemán y ruso, es decir, aprendientes que tenían como lengua materna seis idiomas diferentes; y en B2 se contó con hablantes cuyas lenguas maternas eran cuatro idiomas diferentes, a saber, inglés, francés, alemán y coreano.

En el caso específico de A2, la aprendiente que escribió el menor número de palabras tiene como lengua materna el mandarín, lo que puede ser la causa de esto, ya que según Méndez (2005), a los chinos se les dificulta mucho el aprendizaje del español y necesitan más tiempo que extranjeros a otros países para llegar a tener fluidez en la lengua, ya que al ser su lengua muy distante de la lengua española, cuando los chinos empiezan a aprender español se encuentran en desventaja con relación a otros. Asimismo, esta puede ser una razón por la cual los datos del grupo control tienen una distribución más normal, ya que la lengua materna de todos era el español.

En la variable de la edad, la cual tampoco fue controlada, en el caso del grupo de A2 las edades de los aprendientes varían entre los 25 y los 61 años, mientras que en B2 varían entre los 20 y los

60 años, lo que pudo haber generado desviaciones. Mientras que en el grupo control las edades de los estudiantes se encuentran entre los 20 y los 30 años.

De la misma manera, tampoco se encontró una distribución normal en los datos por separado de cada uno de los centros de interés de cada uno de los niveles, siendo los centros de interés del grupo control los que más se acercaron a ella.

5.1.6. Test Kruskal-Wallis

En un primer momento, se realizó la prueba de Kruskal-Wallis para comparar y establecer diferencias entre la cantidad de palabras entre A2, B2 y el grupo control con los datos en general y se encontró que no había diferencias significativas, sin embargo, el valor era muy bajo por lo que probablemente sí podía existir una diferencia, pero solamente entre un par de datos, lo que fue verificado posteriormente con el Test Mann-Whitney. Además de lo anterior, con este test se encontró que entre los centros de interés de A2 existían diferencias significativas y de igual manera entre los centros de interés en B2 y en el grupo control. Finalmente, en la comparación que se estableció entre cada uno los centros de interés y los diferentes niveles también se encontró un valor de diferencia significativa. En este sentido, es evidente que en cada nivel en los tres centros de interés se encuentra una cantidad diferente de palabras disponibles.

5.1.7. Test Mann-Whitney

Al realizar la comparación entre pares de niveles, es decir A2 y B2, A2 y grupo control, B2 y grupo control, se encontró que estadísticamente entre los niveles A2 y B2 hay diferencias significativas en la cantidad de léxico disponible, así como entre A2 y el grupo control, pero también se encontró que entre el nivel B2 y el grupo control no hay diferencias significativas. Esto va en concordancia con el Marco Común de Referencia Europeo (2002), que indica que en el nivel B2 debe haber un mayor conocimiento léxico que en el nivel A2.

De la misma manera, se realizó la comparación con este test entre pares de centros de interés, por ejemplo entre Comidas y bebidas y lugares, y se encontró que en cada uno de los centros de cada nivel hay diferencias significativas, menos en el grupo control que no se encontró diferencia entre medios de transporte y lugares de la ciudad, esto se debe posiblemente a que en los medios de transporte, en el grupo control aparece más variedad de palabras con relación a los aprendientes de ELE y se incluyen medios de transporte locales como *chalupa, chiva, mototaxi, escalera*.

Finalmente, en la comparación que se estableció por cada uno de los centros de interés en los tres niveles, en comidas y bebidas se encontró una diferencia significativa entre A2, B2 y el grupo control, pero no entre B2 y el grupo control. En lugares de la ciudad se encontró que hay una pequeña diferencia entre A2 y B2, pero no hay diferencia significativa entre A2 y B2 con relación al grupo control. Y con respecto a los medios de transporte, se encontró que no hay diferencias significativas entre A2 y B2, pero que sí hay diferencias significativas entre A2 y B2 en cuanto al grupo control, esto puede ser debido a que como se expresó anteriormente, los ítems léxicos en español en relación con los medios de transporte son más limitados (Ferreira y Blanco, 2018) y por lo tanto los hablantes nativos pueden tener un mayor conocimiento de ellos que los aprendientes de ELE.

5.1.8. Características generales del léxico disponible encontrado

En los tres centros de interés se encontraron palabras que se relacionan con el contexto de la ciudad de Medellín y que se encuentran entre las palabras más disponibles. En el caso de las comidas y bebidas, se encuentran palabras que hacen parte de la gastronomía típica de la ciudad como *arepa, aguardiente, chicharrón, bandeja paisa, trago, sancocho, buñuelo, sopa de mondongo, aguapanela, guaro, morcilla*, entre otras. Posiblemente, esto se relacione por una parte con las experiencias de los aprendientes de ELE en la ciudad y por otra parte con los métodos

de enseñanza del léxico en el aula de clase, ya que los docentes y su metodología tienen influencia en la adquisición del léxico por parte de los aprendientes (Sierra y Pérez, 2016). En relación con los lugares de la ciudad, en los tres niveles la mayoría de unidades léxicas que escribieron los estudiantes fueron nombres propios de lugares de la ciudad de Medellín, como *Museo de Antioquia, Comuna 13, Parque Norte, Parque Lleras*, entre otras.

Finalmente, con relación a los medios de transporte, en los aprendientes de ELE es común encontrar el mismo medio de transporte con diferentes variantes como *bicicleta-bici, auto-carro-coche, Metrocable-teleférico*, mientras que esto no ocurre en el grupo de hablantes nativos. De esta manera, en este centro de interés es posible encontrar variaciones lingüísticas, o sea alternancias de dos o más expresiones para referirse al mismo elemento que se construyen en el uso de las palabras en un contexto socio-cultural específico (Mitkova, 2017), en el ejemplo anterior de *auto-carro-coche*, si bien en Medellín se utiliza comúnmente la palabra *carro*, se considera que el español es una lengua especialmente rica en variaciones dialectales y a la misma vez lexicales, lo que puede suponer un problema para la enseñanza de ELE (Mitkova, 2017), pero es posible que los estudiantes conozcan las otras variaciones para referirse al mismo medio de transporte a través de su contacto con la lengua en otros países o en otros contextos como el cine o la música (Mitkova, 2017).

Por otra parte, en los tres niveles se encontraron nombres de aplicaciones relacionadas con el transporte como *Uber, Didi, Indriver y Picap*. En este caso es evidente la existencia de un proceso de lexicalización que consiste en que nombres comunes, como los presentados anteriormente, que son además entidades lingüísticas nuevas pasan a ser convencionalizadas por los hablantes (Brinton y Closs, 2005), es decir que entra a ser parte del léxico de ese centro de interés y esto se evidencia además porque esas palabras no surgieron solamente en los grupos de aprendientes,

sino también en el grupo control. Además de eso, en el grupo control se encuentran unidades léxicas que no aparecen en los grupos de aprendientes y son medios de transporte que se relacionan con un contexto muy local como *mototaxi, chiva, escalera y chalupa*.

También, es posible encontrar palabras de comidas de otras regiones hispanohablantes o que son más usadas en otras variantes del español como *paella, antro, coche, empanadilla y zumo*, sin embargo, no es una cantidad muy grande frente al tamaño de la muestra.

Las características del léxico disponible encontrado, ponen en evidencia que el proceso de asociación léxica que más realizan los aprendientes de ELE y el grupo control participantes en esta investigación es un mecanismo semántico-cognitivo, es decir, que su léxico disponible se encuentra relacionado directamente con las representaciones mentales que tienen del contexto y los elementos que le rodean (Henríquez, et al, 2016), generando principalmente procesos de hiperonimia e hiponimia (*frutas, mango, guayaba*), de sinonimia (*carro, coche*), de coordinación (*fríjoles, garbanzos*), asociaciones espaciales (*casa, calle*) y asociaciones culturales (*bandeja paisa, sopa de mondongo*) (Henríquez, et al, 2016).

5.1.9. Unidades léxicas más disponibles y menos disponibles

Las diez palabras más disponibles en el centro de interés comidas y bebidas por cada uno de los niveles y las cinco palabras menos disponibles fueron las siguientes:

Tabla 7

Palabras con mayor disponibilidad: comidas y bebidas

A2	B2	Grupo control
Agua	Agua	Arroz
Cerveza	Pan	Fríjoles
Arepa	Carne	Jugo
Leche	Frutas	Agua
Pollo	Leche	Leche
Café	Arroz	Gaseosa
Pan	Jugo	Hamburguesa
Aguardiente	Manzana	Pan
Arroz	Fríjoles	Sopa
Huevos	Café	Carne

Tabla 8

Palabras con menos disponibilidad: comidas y bebidas

A2	B2	Grupo control
Jamón	Bretaña	Quesito
Sándwich	Papa criolla	Chócolo
Paella	Maracuyá	Salsa
Jugo de manzana	Granadilla	Pepperoni
Papitas	Kombucha	Soda

Como se evidencia en la tabla, en las primeras diez palabras se encuentran coincidencias entre los tres niveles, aunque en diferentes posiciones: agua, leche, pan y arroz. Además, se encuentra una coincidencia entre A2 y B2: café. Y finalmente se encuentran coincidencias entre B2 y el grupo control: carne y jugo. Al contrario, en las cinco palabras menos disponibles no se encuentran coincidencias. A continuación, se presentan las diez palabras más disponibles en el centro de interés lugares de la ciudad por cada uno de los niveles y las cinco palabras menos disponibles:

Tabla 9

Palabras con mayor disponibilidad: lugares de la ciudad

A2	B2	Grupo control
Parque	Iglesia	Universidad
Hospital	Parque	Jardín botánico
Universidad	Supermercado	Parque Berrío
Escuela	Escuela	Parque Explora
Restaurante	Universidad	Centro
Tienda	Centro comercial	Parque de los deseos
Supermercado	Calle	Parque Norte
Centro comercial	Museo	Museo
Mercado	Aeropuerto	Estadio
Hotel	Estadio	Centro comercial

Tabla 10

Palabras con menor disponibilidad: lugares de la ciudad

A2	B2	Grupo control
Peluquería	Antro	Parqueadero
Barrio	Puerto	Plaza de mercado
Parada de autobús	Barrio	Cementerio
Pricesmart	Pabellón del agua	Tienda
Bello	Parque Arví	Supermercado

En las primeras diez palabras es posible encontrar coincidencias entre los tres niveles: *universidad* y *centro comercial*. Sin embargo, se encuentran más coincidencias entre A2 y B2: *hospital*, *escuela* y *supermercado*. Es evidente además que en las diez palabras disponibles del grupo control se encuentran más nombres propios de lugares que en los otros niveles, ya que los hablantes nativos tienen una capacidad cognitiva mayor de realizar procesos de lexicalización en su propia lengua, y en este caso lexicalizan los nombres propios de lugares que hacen parte de una realidad inmediata que es la ciudad y de esta manera, hay una relación más estrecha entre el léxico de la ciudad y su propia ciudad, que la que puede haber para un extranjero en una ciudad ajena, por lo que puede haber una tendencia a generalizar y no lexicalizar los nombres propios. Por este mismo motivo puede explicarse que en las palabras menos disponibles en el grupo control

se encuentran tienda y supermercado, que estuvieron en las palabras más disponibles de los aprendientes de ELE.

En la siguiente tabla se presentan las diez palabras más disponibles en el centro de interés medios de transporte por cada uno de los niveles y las cinco palabras menos disponibles:

Tabla 11

Palabras con mayor disponibilidad: medios de transporte

A2	B2	Grupo control
Tren	Metro	Bus
Bicicleta	Carro	Metro
Avión	Taxi	Bicicleta
Metro	Bus	Carro
Taxi	Autobús	Taxi
Carro	Avión	Avión
Bus	Tren	Moto
Autobús	Bicicleta	Patines
Moto	Coche	Metrocable
Coche	Camión	Lancha

Tabla 12

Palabras con menos disponibilidad: medios de transporte

A2	B2	Grupo control
Didi	Auto	Paracaídas
Patines	Lancha	Auto
Llama	Nave	Circular
Chevy	Jet privado	Cicla
Ford	Mula	Locomotora

Como se muestra en la tabla, es posible encontrar coincidencias entre los tres niveles: *bicicleta, avión, metro, taxi, carro y bus*. Además, hay coincidencias entre A2 y B2: *autobús y coche*; que son variantes de *carro y bus*.

De esta manera, en los tres centros de interés hay coincidencias en las palabras más disponibles entre A2, B2 y el grupo control. Es posible afirmar que esto se debe a que los mecanismos de asociación léxica que utilizan los estudiantes son los semánticos-cognitivos, realizando principalmente asociaciones culturales (Henríquez, et al, 2016), para activar así el léxico de los alimentos, los lugares de la ciudad y los medios de transporte que se relacionan directamente con la cotidianidad de la vida en Medellín. Además de eso, esto puede responder a aprendizajes unificados y realidades compartidas entre los aprendientes de ELE y el grupo control.

Finalmente, para analizar las coincidencias encontradas en el léxico disponible, es importante tener en cuenta la noción de categoría de las cosas (Lakoff, 1986), a través de las cuales

comprendemos el mundo, es decir que lo comprendemos no en términos de objetos individuales sino en agrupaciones de elementos. De esta manera, cuando se hace la categorización de un grupo de elementos, en este caso, una categorización léxica se está realizando una selección de prototipos, o sea, de elementos especialmente representativos en una categoría que actúan como puntos de referente cognitivo (Rosch, 1978). La selección de estos prototipos semánticos se basa en la experiencia humana y no en la lengua (Schmitt, 2000), así, los prototipos son los ejemplares idóneos comúnmente asociados a una categoría (Kleiber, 1995). En el caso de esta investigación, se puede decir que los prototipos de comidas y bebidas fueron *agua, leche, pan y arroz*, que corresponden a alimentos comunes en la alimentación occidental; en lugares de la ciudad fueron *hospital, universidad y supermercado* que corresponden a lugares frecuentados en la cotidianidad de los estudiantes (universidad y supermercado) y *hospital* corresponde a un lugar de primera necesidad; y en medios de transporte fueron *bicicleta, avión, taxi y bus*, que son medios de transporte comunes. Así, “a pesar de las diferencias que puedan existir entre los campos semánticos en diversas lenguas, las categorizaciones y las categorías constituyen un principio universal y esencial de la organización de la experiencia humana, que está en todas las lenguas” (Izquierdo, 2003, p. 401).

En ese mismo sentido, Briz (1996), en su obra *El español coloquial: situación y uso* plantea que tanto en el ámbito oral como en el escrito hay una influencia muy relevante del contexto social, así, cuando se evocan las palabras de un tema se establecen relaciones y asociaciones a partir de las percepciones del mundo que tenemos en nuestra cognición, creando así la posibilidad de una identificación rápida de un grupo de palabras pertenecientes a un centro de interés que se construyen a partir de la experiencia, la observación, la influencia cultural y la memoria (Aitchison, 1994). Un ejemplo de esto, es en el centro de interés de los medios de transporte, en

el que se encontraron unidades léxicas que evidencian asociaciones con el sistema de transporte integrado del Metro de Medellín con palabras como: *metro, tranvía, teleférico / Metrocable, metroplús*. O en el centro de interés de comidas y bebidas en donde es evidente la asociación de los elementos que componen la comida típica de la región o la bandeja paisa, con palabras como *fríjoles, arroz, aguacate, chicharrón, huevo, plátano* entre las más disponibles.

5.2. Riqueza léxica

5.2.1. Selección de los textos analizados

La segunda muestra escrita contó con los mismos participantes de la primera muestra, por lo tanto, se obtuvieron 18 textos de A2, 14 textos de B2 y 23 textos del grupo control. Sin embargo, se tuvo que descartar un texto del grupo control, ya que el estudiante solamente escribió un mensaje de 10 palabras que no cumplía con lo indicado en el ejercicio, por ello, se tomaron un total de 22 textos del grupo control.

5.2.2. Léxico no tenido en cuenta

En los textos escritos por los aprendientes de ELE se encontraron algunas unidades léxicas que no se tuvieron en cuenta para formar parte del análisis.

- Errores léxicos: en los textos de A2 se encontraron los siguientes errores léxicos *amablan, escriendo, aproximadamente, llages*. Mientras que en los textos de B2 se encontraron los siguientes errores *pares* (bares), *transvía, sendismo, Antioque, tramatica* (traumática). Es decir, que de un total de 4362 palabras escritas por los aprendientes de ELE se encontraron solamente 8 errores lexicales, lo que puede representar buenos niveles de uso lexical.

- Interferencias de otras lenguas: de la misma manera que en los resultados de disponibilidad léxica, en el nivel A2 se encontraron más casos de préstamos lingüísticos y de acuñaciones, que en el nivel B2, en el que en este caso no se encontró ninguna interferencia. Los préstamos

encontrados fueron los siguientes: *my, or, education, governor, avocados, ciao*. Las acuñaciones fueron: *enchanta, universidad, aplicaciones, chemicas*.

5.2.4. Extensión de los textos

En A2 los aprendientes escribieron un promedio de 120 palabras por texto, los de B2 un promedio de 158 palabras por texto y los estudiantes pertenecientes al grupo control escribieron un promedio de 159 palabras por texto. De esta manera, es posible afirmar que los aprendientes de B2 escribieron textos más largos que se acercan a la extensión de los textos escritos por el grupo control, lo que puede atender a que en el nivel B2 los estudiantes tienen la capacidad de escribir textos más complejos en su estructura y por lo tanto más extensos (Velandia et al, 2010).

5.2.5. Promedio, varianza, desviación y mediana

De cada uno de los grupos de datos arrojados por la fórmula de riqueza léxica, se calculó el promedio, la varianza, la desviación y la mediana. Sin embargo, para el análisis se tendrán en cuenta principalmente los datos correspondientes al porcentaje de vocablos (PV), ya que este valor ofrece una visión de la riqueza léxica y sirve como indicador grueso (López, 2010).

En el caso de la varianza, la más alta la tuvo el grupo de A2 con un valor de 24,88 y la más baja la tuvo el grupo control con 20,29, lo que quiere decir que en la muestra control varió menos el porcentaje de la cantidad de palabras con contenido semántico nocional escritas por los estudiantes y los resultados son más homogéneos. Esto se evidencia, además, en la desviación estándar de cada uno de los grupos, ya que los datos de A2 están más dispersos con un valor de 4,99, mientras que el grupo control tiene una desviación de 4,5. Sin embargo, la desviación de A2 se acerca a la de B2 que tiene un valor de 4,81.

Tabla 13

Resultados arrojados por la fórmula de riqueza léxica

Estudiante	A2				B2				Control			
	N	V	PV	IA	N	V	PV	IA	N	V	PV	IA
1	204	105	51,47	1,94	284	130	45,77	2,18	302	136	45,03	2,22
2	184	77	41,85	2,39	228	103	45,18	2,21	278	117	42,09	2,38
3	151	74	49,01	2,04	200	103	45,18	2,21	273	123	45,05	2,22
4	147	70	47,62	2,1	165	64	38,79	2,58	249	102	40,96	2,44
5	147	61	41,5	2,41	163	75	46,01	2,17	168	77	45,83	2,18
6	142	65	45,77	2,18	159	81	50,94	1,96	167	78	46,71	2,14
7	128	54	42,19	2,37	153	76	49,67	2,01	166	76	45,78	2,18
8	120	51	42,5	2,35	150	69	46	2,17	162	86	53,09	1,88
9	111	42	37,84	2,64	141	75	53,19	1,88	158	75	47,47	2,11
10	110	60	54,55	1,83	136	65	47,79	2,09	156	67	42,95	2,33
11	107	45	42,06	2,38	136	62	45,59	2,19	145	62	42,76	2,34
12	104	48	46,15	2,17	120	62	51,67	1,94	133	65	48,87	2,05
13	100	50	50	2	91	39	42,86	2,33	130	62	47,69	2,1
14	97	46	47,42	2,11	86	43	50	2	130	61	46,92	2,13
15	97	45	46,39	2,16					130	60	46,15	2,17
16	70	37	52,86	1,89					127	71	55,91	1,79
17	66	37	56,06	1,78					124	69	55,65	1,8
18	65	32	49,23	2,03					118	56	47,46	2,11
19									113	57	50,44	1,98
20									92	42	45,65	2,19
21									88	44	50	2
22									85	49	57,65	1,73
Promedio	119,44	55,5	46,91	2,15	158	72,57	46,39	2,18	158,81	74,32	47,74	2,11
Varianza	1448,14	320,26	24,88	0,005	2698,30	514,42	23,2	0,06	3826,92	620,42	20,29	0,04
Desviación	38,05	17,90	4,99	0,23	51,94	22,68	4,81	0,25	61,86	24,91	4,5	0,19
Mediana	110,5	50,5	46,91	2,13	151,5	70,5	46,01	2,17	139	68	46,81	2,14

5.2.6. Intervalo de aparición de palabras con contenido semántico nocional (IA)

Con relación al intervalo con el que aparecen palabras con contenido semántico en los textos (IA), al redondear las cifras, se encuentra que en la gran mayoría de los textos de los tres grupos, se escribe un vocablo nuevo cada dos palabras, salvo una o dos excepciones.

5.2.7. Distribución normal

Al realizar el cálculo de la distribución normal, también conocida como *La campana de Gauss* se encontró que los datos no tenían una distribución normal en ninguno de los tres grupos analizados.

Esta distribución asimétrica de los datos se evidencia en la cantidad de vocablos (V) que se

escribieron, donde en el caso de A2 el texto con el número más alto de V tiene 105, mientras que el texto con menos vocablos tiene 32. De igual manera, en B2 el texto con más vocablos tiene 130 y el texto con menos vocablos tiene 43. Finalmente, en el grupo control el texto con el número más alto de V tiene 136, mientras que el texto con menos vocablos tiene 49.

En el caso de los aprendientes de ELE, la distribución asimétrica de los datos pudo ser ocasionada por las variables que no fueron controladas en esta investigación y en el caso del grupo control pudo ser ocasionada porque a pesar de que todos son hablantes nativos de español, es posible encontrar dificultades en los procesos escriturales o estudiantes más habilidosos para la producción escrita aunque se encuentren en el mismo nivel de educación y esto se relaciona con las experiencias de aprendizaje personales de cada uno de los estudiantes. Además, la producción escrita, teniendo también en cuenta el componente léxico, es una tarea que puede ser calificada según las nuevas experiencias adquiridas (Borja y González, 2005).

5.2.8. Porcentaje de vocablos (PV)

Si bien, se encontró diferencia entre la extensión de los textos de A2 y B2, no se encontró diferencia significativa entre ninguno de los tres grupos de muestras con relación al promedio de palabras con contenido semántico nocional o porcentaje de vocablos, es decir, sustantivos, verbos, adjetivos y adverbios (López, 2010), que son las palabras que permiten conocer la riqueza léxica. Como se evidencia en la tabla 10, el promedio del PV en A2 es de 46,91; en B2 es de 46,39 y en el grupo control es de 47,72.

Además del cálculo de promedios, se aplicaron las pruebas no paramétricas de Kruskal-Wallis y Mann-Whitney para verificar la existencia o no de diferencias estadísticamente significativas entre los porcentajes de vocablos de cada uno de los grupos. Estas pruebas, evidencian que no existían diferencias significativas entre los grupos. Esto quiere decir que los textos escritos por

los aprendientes de ELE de A2 y B2, así como los escritos por el grupo control tienen la misma riqueza léxica. Este resultado podría no ser esperado en la medida en que los estudiantes del grupo control deberían tener una mejor producción escrita que los aprendientes de ELE de B2 y que estos, a su vez, deberían contar con mayor riqueza léxica que los aprendientes de A2, tal y como lo indica el MER (2002). Para poder explicar este hallazgo, en el que no hay diferencias significativas entre la diversidad léxica de las tres muestras, a continuación, se presenta un texto de cada uno de los grupos:

A2	B2	Grupo control
<p>Medellin es una ciudad hermosa cerca de muchas montañas, tiene un historial de ser peligroso pero ahora es muy seguro. Medellin es muy diversa tambien. Tiene personas estan rico y pobres, y es rico en cultura de los españoles, Africa, y la gente de los Andes. Tengo clases de español en una zona muy bonita se llama El Poblado. Yo camino a la escuela y porque hay muchas colinas, siempre estoy sudando.</p> <p>Medellin tiene un sistema de transporte publico que es moderno, incluye un tren, un metrocable, y escaleras mecanicas, la gente usa los buses tambien. Transporte publico es muy importante en Medellin para las personas puedan ir a trabajar. Coches estan muy caro. La comida en Medellin es muy deliciosa. Una comida tipica es Bandeja Paisa, lo tiene frijoles, cerdo, arroz, chicharron, huevos, platano, y normalmente con arepa y aguacate. Me gusta mucho.</p>	<p>Querido amigo, me alegra que habías tenido la decisión de venir aquí en Medellín antes de que empezaba la crisis de salud. Ahora que estás atrapado conmigo y que no puedas visitar esta ciudad muy bonita, voy a explicarte lo que podrías hacer aquí. Bueno primero tienes que irte senderismo. Hay muchísimas caminatas muy cheveres alrededor de la ciudad (una muy sencilla es la caminata de la 3 cruces, la cual te ofrecerá un reflejo muy completo de toda la gente de la ciudad). También a mi me parece importante de salir y descubrir la ciudad afuera de las zonas turísticas, es decir por ejemplo el centro como el barrio Bostón donde hay muchos cafés y bares muy bacanos. El clima es muy bueno aunque el taco y la contaminación son enorme; pero a mi me di cuenta que hay tacos aquí porque muchísimas personas están trabajando todos los días.</p>	<p>Querido Juan,</p> <p>Cuando vengas a Medellín tendremos muchos planes por realizar, por estos días el clima está bastante cambiante pero eso no es impedimento para visitar los museos de la ciudad, los restaurantes o los cines. En los días soleados te puedo llevar a conocer el centro, tiene zonas que quiero que conozcas, podemos ir caminando o conocerlo en bicicleta, conocer la plaza Botero, el teatro Pablo Tobón, La playa y lugares que caracterizan el centro. También podemos ir a lugares naturales cercanos como Santa Elena, es uno de mis lugares favoritos o si quieres un poco de fiesta puedes ir al Poblado. En Medellín las personas son amables, es fácil hacer amigos, una tarde podrías conocer mi universidad y sus alrededores. En esta ciudad encuentras todo tipo de comida, lo típico es la bandeja paisa y aunque no me gusta, no te puedes ir sin probarla. En esta ciudad siempre encuentras algo que hacer y que conocer, espero vengas pronto.</p>

Figura 6: *Textos escritos por un estudiante de cada nivel*

Al realizar la lectura detallada de los textos escritos por cada uno de los grupos de la muestra, si bien son evidentes las diferencias en cuanto a estructuras sintácticas y gramaticales, en el caso del léxico no se encuentran diferencias y se evidencia un uso acertado de las palabras en su contexto y una riqueza léxica amplia en los tres textos, por lo tanto, es posible afirmar que la riqueza léxica no depende de la extensión de los textos. En un estudio de riqueza léxica realizado por Linnarud (1986) entre un grupo de suecos aprendientes de ELE de A2 y un grupo de hablantes nativos de España, no encontró diferencias en la riqueza léxica de los textos en cuanto a la diversidad, pero

sí en cuanto a la sofisticación léxica, es decir palabras “raras”, infrecuentes o de un mayor grado de dificultad.

5.2.9. Cuestionamiento a la fórmula

Al revisar la fórmula empleada para hacer el análisis de la RL de los textos (López, 2010) y de los datos obtenidos a través de dicha fórmula, se encuentra que esta parece favorecer a los textos en los que hay menos cantidad de palabras (N), ya que, al descontar la cantidad de palabras repetidas, es posible que haya un descuento mayor en los textos más extensos y por lo tanto parece que tienen menos porcentaje de palabras con contenido semántico nocional. En este mismo sentido, Haché (1991), en su investigación titulada *Aportes de la riqueza léxica a la enseñanza de la lengua materna* en la que midió la riqueza léxica de un grupo de estudiantes con la fórmula propuesta por López (2010) encontró que esta parece penalizar a quienes puedan tener un discurso más elocuente y la capacidad de escribir extensamente más desarrollada, puesto que a menor cantidad de palabras se encontró mayor PV. Asimismo, (Berton, 2014) plantea que la mayoría de las fórmulas para analizar la riqueza léxica dependen de la extensión de los textos por lo que no es posible establecer comparaciones entre grupos de manera clara y no existe un modelo completamente satisfactorio para realizar estos análisis, por lo que en algunos estudios se han hecho estandarizaciones, es decir, tener en cuenta un número determinado de oraciones (Miller, 1991; Klee, 1992) o de palabras (Laufer, 1991).

5.3. Relación entre la disponibilidad léxica y la riqueza léxica

Se encontró una coincidencia entre los aprendientes de ELE que tuvieron mayor disponibilidad léxica en los tres centros de interés analizados y los que tuvieron mayor número de vocablos en los textos de la segunda muestra escrita. Tanto en A2 como en B2 se encontró que tres de los

cinco aprendientes que tuvieron mayor número de palabras disponibles tuvieron a su vez mayor número de vocablos en el texto escrito:

A2				B2			
Estudiante	Palabras disponibles (Muestra 1)	Estudiante	Número de vocablos (Muestra 2)	Estudiante	Palabras disponibles (Muestra 1)	Estudiante	Número de vocablos (Muestra 2)
Est. 1	52	Est.3	105	Est. 1	61	Est. 2	130
Est. 2	52	Est.4	77	Est. 2	60	Est.6	103
Est. 3	49	Est.6	74	Est. 3	58	Est.7	81
Est. 4	49	Est.7	70	Est. 4	56	Est. 3	76
Est. 5	47	Est.2	65	Est. 5	52	Est. 5	75

Tabla 14

Estudiantes con mayor disponibilidad léxica y mayor riqueza léxica

De la misma manera, al contrastar los resultados de los aprendientes que tuvieron menores resultados en las pruebas de disponibilidad léxica, se evidenció que la mayoría de ellos también fueron los que tuvieron menor cantidad de palabras de contenido semántico en los textos escritos como se muestra a continuación:

Tabla 15

Estudiantes con menor disponibilidad léxica y menor riqueza léxica

A2				B2			
Estudiante	Palabras disponibles (Muestra 1)	Estudiante	Número de vocablos (Muestra 2)	Estudiante	Palabras disponibles (Muestra 1)	Estudiante	Número de vocablos (Muestra 2)
Est. 1	10	Est. 4	32	Est. 1	27	Est. 2	39
Est. 2	15	Est. 1	37	Est. 2	38	Est. 5	43
Est. 3	19	Est. 6	37	Est. 3	40	Est. 3	62
Est. 4	24	Est. 3	42	Est. 4	41	Est.6	62
Est. 5	25	Est. 7	45	Est. 5	43	Est. 1	64

De esta manera, es posible afirmar que existe una relación entre la disponibilidad léxica y la riqueza léxica, ya que, al contar con un número mayor de palabras disponibles en el lexicón mental, se puede lograr la producción escrita de textos con mayor número de palabras con contenido semántico nocional y con mayor elocuencia (Vargas, 2015).

Asimismo, según Nunan (2001), el lenguaje escrito tiene un alto grado de complejidad, el cual viene determinado por la densidad léxica, es decir, por el número de palabras de contenido que se necesitan tener en la producción escrita para que el texto sea más elocuente y pertenezca a un código más formal, por lo que es fortalecer el léxico para mejorar la producción escrita (Alcaraz, 2000). Finalmente, el plano de la producción escrita no representa una dificultad para los aprendientes de una lengua extranjera, ya que la mayoría de ellos saben escribir de manera formal en su lengua materna, sino que la dificultad puede residir en la elección de la forma o del léxico adecuado en la otra lengua para escribir un texto (Tribble, 1996). Estas dificultades no se aprecian en los aprendientes de ELE participantes en esta investigación, quienes emplean y seleccionan adecuadamente el léxico en los textos que escribieron, y probablemente esto se deba a la metodología de enseñanza en el aula de clase en la que posiblemente se fortalecen los procesos escriturales y lexicales.

Capítulo 7: Conclusiones

En este trabajo se presentó un acercamiento metodológico para conocer la disponibilidad y riqueza léxicas de un grupo de aprendientes de ELE de A2 y B2 y de un grupo control de hablantes nativos del español. Los campos de estudio de la disponibilidad y riqueza léxicas de aprendientes de ELE han sido poco explorados en Colombia, por lo que este trabajo puede significar un precedente importante para continuar la investigación en estos ámbitos.

Inicialmente, es pertinente resaltar que tanto en DL, como en RL se encontraron errores léxicos y también interferencias de otras lenguas, mayormente del inglés, ya que esta representaba la L1 o la L2 de la mayoría de los aprendientes, sin embargo, para el tamaño de los datos recogidos, estos no representaron mucha cantidad, lo que puede significar que los aprendientes de ELE tienen buenos niveles léxicos.

En el caso de la disponibilidad léxica, esta se trabajó a partir de tres centros de interés que fueron comidas y bebidas, lugares de la ciudad y medios de transporte. Con respecto a esto, se encontró que a pesar de que los tres temas hacen parte del léxico básico y de uso cotidiano, hubo una mayor cantidad de palabras disponibles en el centro de las comidas y bebidas, y una menor cantidad de palabras disponibles en los medios de transporte, posiblemente porque las comidas hacen parte de los elementos de primera necesidad.

En los tres centros de interés se encontraron vocablos que corresponden a un léxico muy contextualizado en la ciudad de Medellín, especialmente en las comidas y bebidas donde entre las palabras más disponibles se encontraron comidas típicas de la región, esto puede ser debido a que los estudiantes llevan a cabo un mecanismo semántico-cognitivo, en el que activan el léxico disponible a partir de las representaciones mentales que tienen del contexto y los elementos que les rodean, lo que se evidencia además en la cantidad de coincidencias que se encuentran en las palabras más disponibles en los tres grupos, también, en el caso de los aprendientes de ELE, en este aspecto pueden tener influencia los métodos de enseñanza aplicados por los docentes en el aula de clase. Además de eso, se encontró que los estudiantes de los tres grupos crean redes semánticas y se generan activaciones de vocablos que no pertenecen a los centros de interés pero que se relacionan con cada uno de ellos.

Finalmente, en cuanto a la disponibilidad léxica, se encontraron diferencias significativas entre los grupos de A2 y B2, donde los aprendientes de B2 tienen un mayor léxico disponible, lo que

concuera con el Marco Común de Referencia Europeo, pero no se encontraron diferencias significativas entre B2 y el grupo control, lo que posiblemente se debe a que los aprendientes de B2 tienen un nivel más alto que se acerca al conocimiento del léxico de los hablantes nativos en el caso específico de los tres centros de interés analizados.

En cuanto a la riqueza léxica, se encontraron diferencias en la extensión, donde en B2 se escribieron textos más largos, posiblemente por la capacidad de los aprendientes de escribir con estructuras gramaticales más complejas y por lo tanto textos más extensos. Sin embargo, no se encontraron diferencias significativas en RL entre ninguno de los tres grupos, y esto se evidencia al realizar la lectura de los textos, ya que si bien hay diferencias en las estructuras gramaticales y sintácticas, en cada uno de los niveles hay una selección y un uso adecuado del léxico, esto se debe posiblemente a los métodos de enseñanza en las instituciones de ELE, en las que se fortalece la producción escrita y, a que la riqueza léxica, no depende de la extensión de los textos.

En cuanto a la relación entre la DL y la RL se encontró que existe una relación en la que los aprendientes con mayor número de palabras disponibles tenían a su vez mayor riqueza léxica.

Tanto en disponibilidad, como en riqueza léxica se encontró que los datos no tenían una distribución normal, ya que entre los mismos grupos de datos había mucha variación, lo que puede atender a la influencia de las variables que no fueron consideradas, la influencia de estas variables puede ser analizada en futuros estudios. Así como la puesta en relación de los datos recogidos con otros campos de la lingüística como la sociolingüística.

Por otra parte, se puede concluir que es importante la elaboración de herramientas de lingüística computacional que permitan el análisis de los datos alrededor de los temas de DL y RL como el que se realizó para esta investigación y que puede ser mejorado con otros elementos necesarios.

Finalmente, se puede afirmar que los estudios del léxico son importantes, ya que desarrollar la competencia léxica es crucial para el aprendizaje de una lengua extranjera (Llach, 2017), y a través de estos se puede conocer el vocabulario disponible de los estudiantes, conocer la influencia de factores socioculturales en el aprendizaje del léxico, y así desarrollar aplicaciones metodológicas que ayuden a potenciar el conocimiento léxico de los estudiantes seleccionando y graduando el léxico para su enseñanza.

Referencias bibliográficas

- Agray, N; Baralo, M. (2020). Enseñanza virtual de ELE: del estudio empírico a la aplicación didáctica. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*.
- Agustín, M. (2017). Aprendizaje de vocabulario en la Lengua Extranjera. Influencia interlingüística, modo de aprendizaje y otras variables moldeadoras. *Palabras Vocabulario Léxico*, (1), pp.17-34.
- Aitchison, J. (1994). *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford - New York: Basil Blackwell.
- Alcaraz, E. (2000). El inglés profesional y académico. Madrid: Alianza.
- Alvar, M. (2004). *La frecuencia léxica y su utilidad en la enseñanza del español como lengua extranjera*. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=1420379>
- Amat, J. (2016). *Kruskal-Wallis test*. Recuperado de: https://rpubs.com/Joaquin_AR/219504
- Amat, J (2017). Test de Wilcoxon-Mann-Whitney como alternativa al t-test. Recuperado de: https://www.cienciadedatos.net/documentos/17_mann%E2%80%93whitney_u_test
- Areiza, R; Flórez, M. (2015). Variantes del español colombiano y su efecto en la enseñanza del español como lengua extranjera. Marco Sociolingüístico. *Cuadernos de Lingüística Hispánica* (27).
- Ávila, A. (2006). *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Universidad de Málaga.
- Ávila, R. (1986). Léxico infantil de México: palabras, tipos, vocablos. En: *Actas del II Congreso Internacional sobre el español de América*. México: Universidad Nacional Autónoma de México, pp. 510-517.
- Baralo, M. (1997). La organización del léxico en lengua extranjera. *Revista de filología románica*. 14, pp. 59-72.
- Baralo, M. (2005). La competencia léxica en el Marco Común Europeo de Referencia. Centro Virtual Cervantes. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/carabela/58.htm
- Berton, M. (2014). *La riqueza léxica en la producción escrita de estudiantes suecos de ELE*. Tesis doctoral. Estocolmo, Suecia: Universidad de Estocolmo.
- Bonilla, E. & Rodríguez, P. (2005). *Más allá del dilema de los métodos*. Bogotá, Colombia: Nomos.
- Blanco, L; Ferreira, A. (2018). Léxico disponible en tres centros de interés de aprendices de Español como Lengua Extranjera. *Íkala, Revista de Lenguaje y Cultura*. 23 (3), pp. 505-517.
- Borja, M; González, C. (2005). Una experiencia de investigación: dificultades en la producción y comprensión de textos. *Centro de investigaciones y desarrollo científico*, pp. 467-482.
- Butrón, G. (1987). *El léxico disponible: índices de disponibilidad*. Tesis de doctorado. Universidad de Puerto Rico, San Juan.
- Brinton, L; Closs, E. (2005). *Lexicalization and Language Change*. Universidad de Cambridge.
- Briz, A. (1996). El español coloquial: situación y uso. *Cuadernos de lengua española* (38).
- Cabrera C. (2011) *Disponibilidad léxica en estudiantes de preparatoria, de la ciudad de Jerez, Zacatecas*. Tesis de Maestría. Jerez, México: Universidad Autónoma de Zacatecas.
- Callealta, F; Gallego, D. (2016). Medidas de disponibilidad léxica: comparabilidad y normalización. *Boletín de Filología*, (1), pp. 39-92.

- Celaya, M; Torras, M. (2001). L1 Influence and EFL Vocabulary. Do Children Rely More on L1 Than Adult Learners?. *Actas del XXV Congreso AEDEAN*. Granada: Universidad de Granada, pp. 1-14.
- Clouet, R. (2010). El enfoque del Marco Común Europeo de Referencia para las Lenguas: unas reflexiones sobre su puesta en práctica en las facultades de traducción e interpretación en España. *Revista de Lingüística teórica y aplicada*, 48, (2), pp.71-92.
- Cintrón, F. (1993). *Nuevos índices de riqueza léxica en escolares de Barranquitas*. Tesis de pregrado. Río Piedras: Universidad de Puerto Rico.
- Dulay, H., Burt, M. & Krashen, S. (1982). *Language two*. Oxford: Oxford University Press.
- Espejo, M; Flórez, M; Zambrano, I. (2011). Tendencias de los estudios de español como lengua extranjera (ELE) en Bogotá. *Lenguas en contacto y bilingüismo*, (1), pp. 55-91.
- Ethnologue (2018). *Sistema estadístico sobre población y lenguas*. Instituto Etnográfico de Suiza. Recuperado de <https://www.ethnologue.com/browse/names>
- García, A. (2003). *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años*. Tesis doctoral. Almería, España: Universidad de Almería.
- García, D; García, J; Buitrago, Y. (2011). Estado del arte de ELE en Colombia, una mirada holística: metodologías y enfoques, material didáctico, variedades lingüísticas y dificultades en el área. *marcoELE, Revista de Didáctica del Español como Lengua Extranjera*.
- García Hoz, V. (1953). *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid: C.S.I.C.
- Garzón, A; Penagos, L. (2016). Disponibilidad léxica en estudiantes de primer semestre de pregrado en una institución universitaria de Villavicencio, Colombia. *Forma y Función*, 29 (1), pp. 63-84.
- Girón, M. (1998). Sobre los estudios literarios en la Universidad de Antioquia. *Lingüística y Literatura*. 34,35. pp. 172-202.
- Gómez, M. (2004). *La Disponibilidad Léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Tesis de doctorado. Valencia, España: Servei de Publicacions.
- Grupo de investigación DispoLex. (2003-2019). *¿Qué es el proyecto Panhispánico?*. Recuperado de: <http://www.dispoplex.com/info/el-proyecto-panhispanico>
- Hache, A. (1991). *Aportes de la riqueza léxica a la enseñanza de la lengua materna*. En H. López Morales (ed.), *La enseñanza del español como lengua materna*, Río Piedras: Universidad de Puerto Rico, pp. 47-60.
- Henríquez, M; Mahecha, V; Mateur, G. (2016). Análisis de los mecanismos cognitivos del léxico disponible de cuerpo humano a través de grafos. *Lingüística y literatura* (69), pp.229-251.
- Hernández, R., Fernández, C. & Baptista, P. (2006). *Metodología de la investigación*. México: McGraw-Hill.
- Higuaita, D; Martínez, F; Patiño, L. (2012). *Estado del arte de la enseñanza del español como lengua extranjera: una aproximación a las experiencias didácticas en la enseñanza del español como lengua extranjera en la ciudad de Medellín*. Recuperado de: <http://bibliotecadigital.usbcali.edu.co/handle/10819/1557>
- Izquierdo, M. (2003). *La selección de léxico en la enseñanza del español como lengua extranjera. Su aplicación al nivel elemental en estudiantes francófonos*. Tesis de doctorado. España: Universidad de Valencia.

- Jarvis, S. (2009). *Lexical Transfer*. Pavlenko, Aneta (ed.), *Bilingual mental lexicon: Interdisciplinary approaches*. Clevedon: Multilingual Matters, pp. 99-124.
- Jing, L. (2012). El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera. *marcoELE, Revista de Didáctica del Español como Lengua Extranjera*.
- Justicia, F. (1995). *El desarrollo del vocabulario: diccionario de frecuencias*. Universidad de Granada.
- Kerlinger, F. (2002). *Investigación del comportamiento: técnicas y comportamiento*. México: Editorial Interamericana.
- Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, (12), pp. 28-41.
- Kleiber, G. (1995). *La semántica de los prototipos. Categoría y sentido léxico*. Madrid: Visor.
- Kumar, V. (2014). El uso de las nuevas tecnologías en la enseñanza a distancia de español como lengua extranjera en la India. *Centro Virtual Cervantes*.
- Lakoff, G. (1986). *Classifiers as a reflection of mind*. En C. Craig (ed.), pp. 13-51.
- Laufer, B. (1991). The development of L2 lexis in the expression of advanced language learner. *Modern Language Journal*, (75), pp. 440-448.
- Lehmann, F. (1992). Semantic networks. *Computers & Mathematics with Applications*, 23,(2), pp. 1-50.
- Linnarud, M. (1986) *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*. Malmö: CWK Gleerup.
- Llach, M. (2017). Aprendizaje de vocabulario en la Lengua Extranjera. *Palabras vocabulario léxico*, (1), pp.17-33.
- López, H; Lorán, R. (1983). *Nouveau calcul de l'indice de disponibilité*. Universidad de Puerto Rico: San Juan.
- López, H. (1995). Disponibilidad léxica en Andalucía proyecto de investigación. *Revista de estudios de adquisición de la lengua española*, 3, pp. 65-76.
- López, H. (1996). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología Universidad de Chile*, 35, pp. 245 – 259.
- López, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros.
- López, B. (2008). *Léxico disponible en el español de Galicia*. Santiago de Compostela: Universidad de Santiago de Compostela.
- López, H. (2010). *Los índices de riqueza léxica y la enseñanza de lenguas*. En: *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE coord. por Javier de Santiago Guervós*, pp. 15-28.
- López, A. (2014). *Desarrollo de los estudios de disponibilidad léxica en Español Lengua Extranjera (ELE)*. En N. Contreras. *La enseñanza del Español como LE/L2 en el siglo XXI* (397-408).
- Luna, E; Viguera, A & Báez, G. (2005). *Diccionario básico de lingüística*. Universidad Autónoma de México: México.
- Martínez, M; Marí, M. (2016). La distribución normal. *Repositorio Institucional de la Universidad Politécnica de Valencia*.
- Mateus, G; Santiago, W. (2006). Disponibilidad léxica en estudiantes bogotanos. *Folios*, 1(24), pp. 3-26.
- Mayén, S. (2009). *Comprensión de las medidas de tendencia central en estudiantes mexicanos de educación secundaria y bachillerato*. Granada: Universidad de Granada.

- Méndez, E. (2005). Problemas de los estudiantes chinos de español. Tesis de maestría. Madrid, España: Universidad Nacional de Educación a Distancia.
- Mendoza, J. (2018). El léxico disponible de 82 estudiantes coreanos de Español como Lengua Extranjera. *marcoELE: Revista Didáctica del Español como Lengua Extranjera*.
- Miller, J. (1991). *Quantifying productive language disorders*. En J. F. Miller (ed.), *Research on child language disorders: a decade of progress* (pp. 211–220). Austin, TX: Pro-Ed.
- Ministerio de Educación, Cultura y Deporte. (2002) Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación.
- Mitkova, A. (2017). La contribución de la sociolingüística a la enseñanza de lenguas extranjeras. *Centro Virtual Cervantes*.
- Nielsen, J. (2018). *Los estudiantes extranjeros y su proceso de aprendizaje del español en Colombia: primeros indicios del translanguajeo*. Recuperado de: <https://www.scielo.br/pdf/alfa/v62n1/1981-5794-alfa-62-1-0053.pdf>
- Nunan, D. (2001) *Second Language Acquisition*. En R. Carter, & D. Nunan (Eds.), *The Cambridge Guide to Teaching English to Speakers of Other Languages*, pp. 87-92). Cambridge: Cambridge University Press.
- Paolini, A. (2017). *Estudio sobre disponibilidad léxica en alumnos de ELE en la Universidad de Padua*. Tesis de pregrado. Padua, Italia: Universidad de Estudios de Padua.
- Pérez, M. (2011). *Disponibilidad léxica de los escolares de Guamá en tres centros de interés*. En *Comunicación Social en el siglo XXI, Vol. I*, Santiago de Cuba: Ediciones Centro de Lingüística Aplicada, pp. 80-84.
- Piñeros, H. (2014). *El registro lingüístico en el discurso literario colombiano para el fomento de la competencia sociolingüística en la enseñanza del español como lengua extranjera*. Repositorio Universidad Pedagógica Nacional.
- Prundeanu, R. (2013). *El aprendizaje de locuciones verbales por parte de alumnos ELE rumanos. Efectos de proximidad semántica y riqueza léxica en actividades escritas*. Tesis doctoral. Madrid, España: Universidad Nebrija.
- Quintero, L; Avilés, L; Suárez, K. (2013). Estado de la enseñanza del español como lengua extranjera en Santander. *Colombian Applied Linguistics Journal*.
- Ringbom, H. (2001). *Lexical transfer L3 production*. En J. Cenoz, B. Hubiesen y U. Jessner (eds). *Cross - linguistic in third language acquisition: Psycholinguistic perspectives*, Clevedon: Multilingual Matters, p. 59-68.
- Ringbom, Hakan (2006). *The Importance of Different Types of Similarity in Transfer Studies*. Arabski, Janusz (ed.), *Cross-linguistic Influences in the Second Language Lexicon*. Clevedon: Multilingual Matters, pp. 36-45.
- Rodríguez, G., Gil, J. & García, E. (1999). *Metodología de la investigación cualitativa*. Málaga: Aljibe.
- Rosch, E. (1978). *Principles of Categorization*. En E. Rosch & B. Lloyd (eds.): *Cognition and Categorization*, pp. 27-48.
- Rubio, M. (2018). Disponibilidad léxica de 52 estudiantes taiwaneses universitarios de ELE A1. *marcoELE, Revista de Didáctica del Español como Lengua Extranjera*.
- Sánchez, J. (2016). El potencial de la realidad aumentada en la enseñanza de español como lengua extranjera. *Revista de Educación Mediática y TIC*.
- Serfati, M. (2016). La Disponibilidad Léxica de estudiantes marroquíes de nivel universitario: resultados generales. *Philologica Canariensia*, 22, pp. 105-116.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

- Serrano, C. (1977). Historia de la lexicografía griega, antigua y medieval. En Rodríguez, F (comp.). Introducción a la Lexicografía Griega (61-107). CSIC: Madrid.
- Sierra, L; Pérez, C. (2016). Disponibilidad léxica de estudiantes de español como lengua extranjera en Santiago de Cuba. *Revista Universidad de Oriente*, 141, pp. 770-782.
- Sifrar, M. (2012). Análisis comparativo de la disponibilidad léxica en Español Como Lengua Extranjera (ELE) y Lengua Materna (ELM). *marcoELE: Revista Didáctica del Español como Lengua Extranjera*.
- Strassburger, C; López, J. (2000). El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable. *Anuario de letras: lingüística y filología*, 38, pp. 227 – 251.
- Tesitelová, J. (1992). *The main areas of quantitative linguistics*. Nueva York: Planum Press.
- Tribble, C. (1996). *Writing*. Oxford: Oxford University Press.
- Vargas, E. (2015). Uso del léxico en textos escritos por estudiantes de nivel B2 de español como segunda lengua. *Filología y lingüística*, (41), pp. 81-92.
- Villanueva, J & Perdrix, M. (2011). El aula virtual en la enseñanza online de una lengua extranjera: tipos y materiales. *Universidad de Granada*. Recuperado de: <https://pdfs.semanticscholar.org/143b/023269fa48d86751f294b1893543f0bc2229.pdf>
- Velandia, D; Ussa, M; Waked, M. (2010). Comprensión y producción textual del español como lengua extranjera. *Cuadernos de Lingüística Hispánica*, (15), pp.109-120.
- Vieytes, R. (2004). Metodología de la investigación en organizaciones, mercado y sociedad. Epistemología y técnicas. Buenos Aires: Editorial de las Ciencias.