



Forma y Función

ISSN: 0120-338X

formafun@bacata.usc.unal.edu.co

Universidad Nacional de Colombia

Colombia

Garzón Segura, Anni Marcela; Penagos López, Luis Alberto
DISPONIBILIDAD LÉXICA EN ESTUDIANTES DE PRIMER SEMESTRE DE
PREGRADO DE UNA INSTITUCIÓN UNIVERSITARIA DE VILLAVICENCIO, COLOMBIA
Forma y Función, vol. 29, núm. 2, enero-junio, 2016, pp. 63-84
Universidad Nacional de Colombia
Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=21947316004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

DISPONIBILIDAD LÉXICA EN ESTUDIANTES DE PRIMER SEMESTRE DE PREGRADO DE UNA INSTITUCIÓN UNIVERSITARIA DE VILLAVICENCIO, COLOMBIA*

*Anni Marcela Garzón Segura***

*Luis Alberto Penagos López****

Universidad Cooperativa de Colombia

Resumen

En este artículo, se presenta un estudio de disponibilidad léxica realizado sobre la muestra de alumnos de primer semestre de pregrado de una institución universitaria de Villavicencio, Colombia. Se usó la Encuesta de Disponibilidad Léxica, la cual recoge el léxico de los estudiantes a través de 16 centros de interés, más 5 añadidos por los investigadores. Se realizó el análisis de resultados a través del programa informático Dispolex. Los resultados presentan los índices de disponibilidad léxica y de cohesión, así como la totalidad de palabras y vocablos informados por los estudiantes. De esta manera, se hizo un análisis cuantitativo que mostró una disponibilidad léxica baja y difusa que es similar a la de otras investigaciones con poblaciones semejantes. Además, se encontró que las variables sociodemográficas no generan diferencias significativas en el léxico disponible, sino que son los estímulos o centros de interés los que explican las diferencias encontradas.

Palabras clave: *disponibilidad léxica; lexicón mental; lengua hablada; estudiantes universitarios; Villavicencio.*

Cómo citar este artículo:

Garzón Segura, A. M., & Penagos López, L. A. (2016). Disponibilidad léxica en estudiantes de primer semestre de pregrado de una institución universitaria de Villavicencio, Colombia. *Forma y Función*, 29(2), 63-84.

Artículo de investigación: Recibido: 29-08-2015, aceptado: 20-05-2016

-
- * Este artículo hace parte de la investigación «Diferencias léxicas entre hombres y mujeres estudiantes de primer semestre de carreras de pregrado de la Universidad Cooperativa de Colombia, sede Villavicencio» financiado como parte de la convocatoria del año 2013 por el Comité Nacional para el Desarrollo de la Investigación (CONADI) de la Universidad Cooperativa de Colombia.
 - ** Docente de tiempo completo en la Universidad Cooperativa de Colombia, sede Villavicencio. Candidata a Doctora en Estudios Interdisciplinarios de Género de la Universidad de Salamanca. anni.garzons@campusucc.edu.co
 - *** Docente de tiempo completo en la Universidad Cooperativa de Colombia, sede Villavicencio. Doctor en Lengua española: Investigación y Enseñanza, de la Universidad de Salamanca. luis.penagos@usal.es

FIRST SEMESTER UNDERGRADUATE STUDENTS' LEXICAL AVAILABILITY AT A UNIVERSITY IN VILLAVICENCIO, COLOMBIA

Abstract

This paper presents a study on lexical availability by first semester undergraduate students in a Villavicencio (Colombia) university. The Lexical Availability Survey was used in order to gather information on students' lexicon in 16 centers of interest, as well as 5 additional places added by the researchers. Dispolex was used in order to analyze the data obtained. Results display lexical availability and cohesion indexes, including all words provided by students. A quantitative analysis was also carried out that showed a low and diffused lexical availability, similar to the one found in other research on similar populations. It was also found that socio-demographic variables do not produce significant differences in lexical availability. These are explained on account of the stimuli or the centers of interest.

Keywords: *lexical availability; mental lexicon; spoken language; university students; Villavicencio.*

DISPONIBILIDADE LÉXICA EM ESTUDANTES DO PRIMEIRO SEMESTRE DE GRADUAÇÃO DE UMA INSTITUIÇÃO UNIVERSITÁRIA DE VILLAVICENCIO, COLÔMBIA

Resumo

Neste artigo, apresenta-se um estudo de disponibilidade léxica realizado sobre a amostra de estudantes do primeiro semestre de graduação de uma instituição universitária de Villavicencio (Colômbia). Utilizou-se a pesquisa de disponibilidade léxica, a qual coleta o léxico dos estudantes de 16 centros de interesse, mais cinco agregados pelos pesquisadores. Realizou-se a análise de resultados por meio do programa informático Dispolex. Os resultados apresentam os índices de disponibilidade léxica e de coesão, bem como a totalidade de palavras e vocábulos informados pelos estudantes. Dessa maneira, fez-se uma análise quantitativa que mostrou uma disponibilidade léxica baixa e difusa que é semelhante à de outras pesquisas com populações parecidas. Além disso, constatou-se que as variáveis sociodemográficas não geram diferenças significativas no léxico disponível, mas sim que são os estímulos ou os centros de interesse que explicam as diferenças encontradas.

Palavras-chave: *disponibilidade léxica; estudantes universitários; léxico mental; língua falada; Villavicencio.*

INTRODUCCIÓN

Los estudios en disponibilidad léxica tienen como objetivo recoger un corpus del léxico disponible en una comunidad o población, analizarlo y valorarlo, tanto cualitativamente como cuantitativamente. Se define al léxico disponible como el conjunto de palabras que los hablantes poseen en su lexicón mental (Bartol, 2006; Ortolano, 2005; Montenegro, 2010), es decir, el vocabulario personal del que dispone un hablante para desenvolverse exitosamente en los diferentes contextos comunicativos. Este se actualiza según la temática requerida de una situación comunicativa.

Los estudios en disponibilidad léxica son cuantitativos, ya que se basan principalmente en la cuantificación léxica, pero han dado cabida a un análisis cualitativo profundo del corpus desde diferentes disciplinas lingüísticas como la sociolingüística y la psicolingüística (Hernández, 2005a, 2005b). Además, brindan una herramienta para la lingüística aplicada y la enseñanza de la lengua materna y extranjera en el diseño de los currículos.

El término *disponibilidad léxica* surgió en los años cincuenta a partir del trabajo de los lingüistas Gougenheim, Michéa, Rivenc y Sauvageot (1964), quienes tenían como objetivo diseñar un diccionario con las palabras más importantes para los actos comunicativos cotidianos, es decir, las palabras de mayor uso o frecuencia de aparición en los discursos orales y escritos. Este diccionario estaría encaminado a la enseñanza de la lengua francesa a los habitantes de las colonias de la *Union Française*. Para la búsqueda del léxico necesario en este trabajo, se utilizaron las denominadas *pruebas de frecuencia*. Estas identificaban las palabras que más aparecían en los textos orales y escritos recogidos de una muestra de informantes y su uso permitía correlacionar la importancia de una palabra con su índice de frecuencia.

Surgió una problemática, ya que las palabras con mayor frecuencia de uso, según estas pruebas, eran de naturaleza gramatical (conjunciones, determinantes, preposiciones), mas no con contenido semántico. Las frecuencias de estas últimas eran mucho más bajas y, sin embargo, son necesarias para la enseñanza de un idioma. Los resultados arrojados por las pruebas de frecuencia evidenciaron una deficiencia en el método para hallar el corpus léxico necesario para definir el currículo de enseñanza de una lengua, ya que el inventario de palabras obtenido carecía de una carga semántica significativa (Bartol, 2006; Ortolano, 2005; Carcedo, 1998).

Como consecuencia de las limitaciones de estas pruebas, se buscó la creación de una prueba diferente, con la cual se pudiera obtener el léxico frecuente con una carga semántica significativa, es decir, las palabras disponibles a las que cualquier hablante acude al momento de hablar de un tema específico. Como lo refiere Michéa (1953), una palabra disponible es la que siempre está lista para ser usada, según sea necesaria en un

acto de habla. Esta palabra se emplea naturalmente y viene a la mente del hablante de manera inmediata, sin necesidad de un trabajo mental mayor. Es una palabra que, sin ser la más frecuente en los diferentes textos, siempre aparece cuando se habla de un tema o unos temas determinados, ya que hace parte de asociaciones de ideas comunes y viene al hablante cuando estas asociaciones entran en juego.

Se encontró solución al problema, diseñando pruebas asociativas en las que se pretendía recoger el léxico disponible de un hablante al proponerle un tema específico. Según Lewandowski (1992), las asociaciones de palabras se dan como respuesta a la reacción producida por una palabra estímulo, presentada por el investigador al informante. De esta manera, se empezaron a utilizar los denominados *centros de interés*, que serían los campos semánticos asociativos concretos que se les presentaría a los informantes para obtener de ellos las palabras relacionadas. Este método permitiría obtener un vocabulario más específico y lleno de significado.

Los estudios en disponibilidad léxica en la lengua española surgieron en los años setenta, con los trabajos del profesor H. López (López, 1999), quien tenía como objetivo la elaboración de un diccionario panhispánico, en el que se recogiera todo el vocabulario disponible de los estudiantes preuniversitarios de distintas regiones de habla hispana. En la actualidad, las investigaciones que se realizan en el campo de la disponibilidad léxica siguen unos parámetros básicos, en los que se presentan lineamientos predeterminados para las pruebas, como lo son los 16 centros de interés básicos, el tiempo de recogida de datos (dos minutos por cada centro de interés) y algunas variables determinadas (sexo, edad, nivel socioeconómico, área de procedencia y tipo de centro). Estos parámetros han sido estandarizados desde los noventa y hoy en día son trabajados a través de la plataforma Dispox de la Universidad de Salamanca, que forma parte del Proyecto Panhispánico de Disponibilidad Léxica.

Este proyecto, originado y coordinado por H. López, tiene como objetivo elaborar diccionarios de disponibilidad léxica que abarquen las diferentes normas del mundo hispano, para poder hacer comparaciones lingüísticas, dialectológicas, sociolingüísticas y culturales. También se ha fijado, como objetivo secundario, la implementación de estos en la lingüística aplicada a la enseñanza del español como lengua extranjera o segunda lengua (Guerra & Gómez, 2003; Grupo de investigación Dispox, 2003).

Para la implementación de este proyecto, se contó inicialmente con la herramienta LexiDisp, elaborada por José Enrique Moreno Fernández y Antonio García de las Heras, con el asesoramiento de Francisco Moreno y Pedro Benítez. Esta herramienta usaba la fórmula de López Chávez y Strassburger (Bartol, 2006). LexiDisp era administrado por la Asociación de Lingüistas y Filología de la América Latina, el Instituto Cervantes y la

Universidad de Alcalá de Henares (Montenegro, 2010). Luego, se llevó a cabo el cambio hacia el programa Dispolex, el cual trabaja con la misma fórmula matemática, es administrado por la Universidad de Salamanca y «da acceso gratuito a las herramientas necesarias para calcular los datos más habituales en disponibilidad léxica: índice de disponibilidad, frecuencia, porcentaje de aparición, número de palabras, número de vocablos, promedios por informante, índice de cohesión, comparaciones entre proyectos, etc.» (Grupo de investigación Dispolex, 2003).

Desde los años cincuenta, hasta la actualidad, se han realizado numerosos estudios en disponibilidad léxica en lengua española. Especialmente, como lo afirma Bartol (2006), a partir de los años noventa se ha visto un incremento y avance en este tipo de investigaciones, ya que han aparecido equipos de trabajo dedicados a estos estudios, particularmente en universidades españolas. Sin embargo, se pueden encontrar algunos grupos en países como Argentina, Puerto Rico, México, Chile, Panamá, Costa Rica, Guatemala, Cuba y Colombia. Teniendo esto en cuenta, a continuación se presentan algunas investigaciones que han tenido como muestra poblacional a estudiantes de primaria, secundaria y universidad.

En Finlandia, Carcedo (1998) realizó una investigación que buscó hacer una prueba diagnóstica para identificar cuál era la disponibilidad léxica de los estudiantes finlandeses de último año de bachillerato. Para esto, contó con una muestra de 78 alumnos pertenecientes a diez colegios de diferentes partes de Finlandia, quienes aprendían el español como lengua extranjera y habían recibido una media de 200 horas de clase en este idioma antes de realizarse la prueba. En esta investigación se tuvieron como variables el sexo, la edad, la lengua materna, el tipo de centro, el lugar de residencia habitual, las lenguas habladas y el nivel que se creía tener en cada una de estas. Para la realización de la prueba, se contó con los parámetros estándar de las pruebas de disponibilidad léxica, es decir, 16 centros de interés, con dos minutos para desarrollar cada centro.

Para la tabulación y análisis de la información, se usó el programa LexiDisp. Su análisis se centró en la identificación de las frecuencias, porcentajes y medias de uso de cada palabra. Se encontró que el promedio de palabras por alumno en cada centro de interés fue de 5.2; «comidas y bebidas» fue el centro de interés con el mayor número de palabras (12 palabras por alumno); y «trabajos del campo y del jardín», el centro con el menor número de palabras (una palabra de media por informante). En relación con la disponibilidad léxica, las diez palabras más disponibles fueron: *coche* (0.85), *perro* (0.80), *profesor* (0.76), *libro* (0.72), *autobús* (0.70), *cocina* (0.68), *tren* (0.68), *avión* (0.67) y *plato* (0.65). En relación con la variable de sexo, este estudio encontró que no existen diferencias cualitativas o cuantitativas importantes entre hombres y mujeres. Para el resto de variables sociodemográficas, esta investigación no presentó hallazgos sobre las diferencias o similitudes.

En el contexto polaco, A. López (2008) realizó una investigación que tuvo el objetivo de evaluar la disponibilidad léxica en español como lengua extranjera de estudiantes de secundaria y bachillerato de seis ciudades polacas. La muestra fue de 241 alumnos del programa de Secciones Bilingües del español y para la recolección de datos se siguieron las pautas del Proyecto Panhispánico de Disponibilidad Léxica: uso de los 16 centros de interés y dos minutos de respuesta por centro. El análisis de los datos se realizó a través de la plataforma Displex, lo cual permitió calcular los índices de disponibilidad, frecuencia relativa, porcentaje de aparición, número de palabras, promedio por informante e índice de cohesión. Como resultado, se encontró un promedio de palabras por informante de 12.05; el centro de interés con menor cantidad fue el de «objetos sobre la mesa para la comida» (5.92 palabras por informante); y el centro con mayor frecuencia, el de «comidas y bebidas» (19.99 palabras por informante).

En relación con la disponibilidad, la palabra más disponible fue *lámpara* del centro de interés «calefacción e iluminación»; y la menos disponible, *fregar* del centro «trabajos del campo y del jardín». Teniendo en cuenta el índice de disponibilidad, las diez palabras más disponibles halladas fueron: *lámpara* (0.8682), *coche* (0.8613), *perro* (0.8528), *gato* (0.8266), *plato* (0.7981), *cabeza* (0.7878), *mesa* (0.7719; 0.7087), *(auto)bus* (0.7522), *cocina* (0.7472), y *profesor* (0.7417). Con respecto a la variable de sexo, se encontró que, teniendo en cuenta el promedio de palabras por informante, las mujeres aportan una palabra y media más que los hombres; y en relación con el nivel sociocultural, los alumnos de clase alta responden una palabra y media más que los de clase media.

En España, Ortolano (2005) buscó identificar la disponibilidad léxica de los estudiantes de séptimo y noveno grado de bachillerato. Para esto, encuestó a estudiantes de los centros educativos de Ayamonte. En esta investigación, se tuvieron en cuenta las variables de sexo y tipo de centro. Por otro lado, en cuanto a la aplicación de la encuesta de disponibilidad léxica, se usaron las normas tradicionales (dos minutos por centro de interés y 16 centros de interés) y se realizó el análisis con el programa LexiDisp para determinar el grado de disponibilidad, la frecuencia absoluta, el porcentaje de aparición y la frecuencia acumulada. Se encontró que la edad es un factor que influye en el número de respuestas encontradas, pues a mayor edad, mayor número de respuestas. Con respecto al sexo, no se encontraron diferencias cuantitativas importantes, aunque a nivel cualitativo las mujeres usaron lenguaje más formal y contestaron más conscientemente. Por último, a nivel general, el centro de interés con mayor disponibilidad fue «animales», seguido por «alimentos y bebidas» y «la escuela». No se presentaron datos sobre índices de disponibilidad para las palabras analizadas.

Asimismo, en Madrid (España), Guerra y Gómez (2003) buscaron conocer cuál es el léxico más disponible acerca de los medios de comunicación. Para desarrollar esta investigación, tuvieron como variables el sexo y la carrera que cursaban los estudiantes encuestados. La muestra correspondió a 124 estudiantes de primer y cuarto curso de las carreras de Periodismo, Comunicación Audiovisual y Derecho. La encuesta de disponibilidad léxica usada en este estudio no tuvo en cuenta los 16 centros de interés, debido a sus objetivos, por lo que solo se preguntó por tres centros de interés: «prensa», «radio» y «televisión». Por otro lado, se siguieron las indicaciones estándar de la prueba, en cuanto al cuadernillo y el tiempo de respuesta por cada centro.

El análisis se realizó a través del programa Dispolex, considerando las palabras totales, las palabras por informante y el índice de cohesión. Se obtuvo un promedio de palabras por informante de 15.27. Del análisis de disponibilidad léxica, destacan la palabra *basura*, con el índice de disponibilidad más alto dentro del centro «televisión», y la palabra *música*, con el más alto en el centro «radio». A su vez, se destaca como resultado que en la variable de sexo no se hallaron diferencias significativas, ni cuantitativa ni cualitativamente entre hombres y mujeres, pues se observó que una mayor productividad está relacionada con una mayor presencia de uno u otro sexo en la muestra. En relación con la variable de carrera, los estudiantes de comunicación presentaron mayor disponibilidad léxica.

En Guatemala, Montenegro (2010) buscó crear un corpus de léxico disponible en estudiantes de tercer y sexto grado, a través de una muestra de 200 alumnos pertenecientes a ocho centros educativos (cuatro rurales y cuatro urbanos) de Guatemala. Para la aplicación de la prueba, se contó con el tiempo estándar, y la encuesta de disponibilidad léxica contó con cinco centros de interés de los 16 comúnmente usados (partes del cuerpo, animales, alimentos, medios de transporte, profesiones y oficios), y con uno nuevo (accidentes geográficos). Para la tabulación y análisis de los datos se usó el programa LexiDisp. Las variables tenidas en cuenta en esta investigación fueron la edad, el curso, el sexo y el centro de formación. Se encontró que el centro de interés con mayor número de palabras fue «animales» y el que reportó menor cantidad fue «accidentes geográficos». Además, los estudiantes de sexto presentaron mayor disponibilidad que los de tercero. Finalmente, el rango de disponibilidad léxica para los centros de interés estudiados estuvo entre 0.11 y 1.00, la palabra con mayor índice fue *río* (1.00), del centro de interés «accidentes geográficos»; y las que tuvieron menor disponibilidad fueron *cocodrilo* (0.14), del centro «animales», y *rodilla* (0.21), del centro «cuerpo humano».

Por otro lado, en Venezuela, la investigación de Rui y Carvajal (2005) tuvo como objetivo identificar y analizar la disponibilidad léxica de los estudiantes universitarios de Castellano y Literatura de la Universidad de los Andes «Pedro Rincón Gutiérrez» de

Venezuela. Para esto, tomaron como muestra 40 alumnos, de los cuales 20 eran de primer año y 20 de quinto año. Este número representaba el 10% de la totalidad del universo poblacional. Para esta investigación, se tuvo como variable el año cursante de los alumnos. En cuanto a la aplicación de la encuesta de disponibilidad léxica, se tuvo en cuenta el tiempo estándar de aplicación por campo de interés. Además, se usaron solo 10 de los 16 centros de interés comúnmente usados. Para la tabulación y análisis de los datos, se usó el programa LexiDisp, a través del cual se encontró que el centro de interés «partes del cuerpo» fue el que obtuvo menor disponibilidad y el centro «la ciudad» fue el que obtuvo mayor disponibilidad. También se encontró que los alumnos de quinto presentaron una disponibilidad mayor que los alumnos de primer año.

En Costa Rica, Ríos (2007) buscó identificar qué diferencias se podrían encontrar en la disponibilidad léxica de hombres y mujeres. Para esto, se encuestaron 512 estudiantes costarricenses (207 hombres y 305 mujeres) de último año de colegio, en los diferentes colegios a lo largo y ancho de Costa Rica. En esta investigación, se tuvo en cuenta tres centros de interés diferentes a los tradicionales: «saludos», «despedidas» y «temas de conversación». Las variables tenidas en cuenta fueron: sexo, tipo de centro (privado o público) y ubicación del centro (área rural o urbana). Sin embargo, el análisis se centró en las diferencias y similitudes en relación con la variable de sexo. En cuanto a los resultados, se pudo identificar que, para el centro «saludos y despedidas», ambos sexos comparten el léxico más disponible. En el centro «temas de conversación», a pesar de que la mayoría de palabras coinciden para ambos sexos, estos cuentan con diferentes índices de disponibilidad.

En el contexto colombiano, de acuerdo con Bartol (2006), se tiene conocimiento de que se han iniciado las siguientes investigaciones: léxico disponible de Bogotá y léxico disponible del Caribe, ejecutado por María Clara Henríquez Guarín, Álvaro William Santiago Galvis, Jaime Ruiz Vega y Geral Eduardo Matéus Ferro (Universidad Pedagógica Nacional) en el año 2004, como parte del Grupo de Investigación en Pedagogía, Lenguaje y Comunicación (GIPELEC). Por otra parte, se encuentra el trabajo de Nancy Rozo Melo e Ivonne Elizabeth Zambrano Gómez del Instituto Caro y Cuervo y la Universidad del Atlántico, en Barranquilla (Colombia), a través del Grupo de Investigación para el Estudio Sociolingüístico del Caribe (GIESCA), coordinado por la profesora Yolanda Rodríguez Cadena. Por último, se conoció el trabajo de grado presentado en la Universidad de Salamanca, de Sandra Liliana Bonilla Rengifo, sobre disponibilidad léxica en Popayán. Sin embargo, a pesar de tener conocimiento sobre la ejecución de estos proyectos, no se han encontrado publicaciones relacionadas, a través de las cuales se puedan conocer los resultados sobre el léxico disponible en las poblaciones colombianas analizadas.

La única investigación colombiana de la que se tiene referencia es la de Matéus Ferro y Santiago Galvis (2006), la cual buscó hacer un diagnóstico de la disponibilidad léxica de los estudiantes bogotanos de quinto de primaria y once de bachillerato en una muestra de instituciones educativas de Bogotá. La muestra estuvo conformada por 911 informantes y los datos se recolectaron, teniendo en cuenta los 16 centros de interés del Proyecto Panhispánico de Disponibilidad y siguiendo las pautas del mismo. El análisis se centró en la identificación del índice de disponibilidad léxica para cada palabra, el índice de cohesión y las frecuencias léxicas por centro de interés. Se encontró que los estudiantes de once presentaron mayor disponibilidad que los de quinto: a medida que aumenta la edad, hay una mayor riqueza cultural, cognitiva y lingüística. En relación con los centros de interés, el centro con mayor productividad léxica fue «animales» y el de menor productividad fue «iluminación, calefacción y medios de airear un recinto». La palabra con mayor disponibilidad fue *perro* (0.84), del centro de interés «animales»; y la que tuvo menor disponibilidad fue *alimentar* (0.04), del centro «trabajos del campo y el jardín». Finalmente, el promedio de palabras por informante en esta investigación fue de 15.6.

METODOLOGÍA

Para la realización de esta investigación, se han seguido los parámetros del Proyecto Panhispánico de Léxico Disponible (PPLD), los cuales son estandarizados para este tipo de estudio y, por ende, los más utilizados por los investigadores de este campo en los diferentes lugares del mundo, ya que la homogeneidad en las pautas de elaboración de los estudios es importante para que los resultados puedan ser contrastables entre sí.

En primer lugar, se seleccionó el centro educativo en el que se realizó el estudio, que en este caso fue una institución universitaria de Villavicencio. Luego, se eligió una muestra representativa de 184 estudiantes de un universo de 353 matriculados en el primer semestre de pregrado. Para elegir esta muestra, se usó la fórmula expuesta por Suárez y Tapia (2012) y Suárez (2011), a través de la cual se buscó una confiabilidad del 95% y un margen de error de un 5%.

Se verificó que los estudiantes seleccionados no tuvieran experiencia académica previa en programas universitarios. Con esto, se evitó la disparidad con respecto a los alumnos de último grado de secundaria, lo cual ayudó a homogenizar la muestra de este estudio con respecto a la de otros estudios en disponibilidad léxica. En este orden de ideas, la encuesta fue aplicada durante la semana de inducción a la vida universitaria, es decir, hasta un momento en que los estudiantes no habían empezado sus clases en las carreras de pregrado. Los estudiantes accedieron a participar en el estudio voluntariamente, mediante un consentimiento informado, y su anonimato se respetó durante toda la investigación.

La encuesta aplicada es común a la mayoría, sino a todos, los trabajos realizados en este campo, y consiste en un cuadernillo de hojas en blanco que se usa para que los informantes puedan escribir sus respuestas. Se contaba con dos páginas para cada centro de interés. Los informantes contaban con un límite de tiempo de dos minutos, cronometrados por el encuestador, para escribir las respuestas en cada centro de interés. De acuerdo con Ríos (2007), el límite de tiempo se justifica en la acción de buscar el léxico más disponible, es decir, el más natural y espontáneo, ya que el léxico del que verdaderamente dispone un hablante es el que acude a su mente de manera más directa cuando aborda un tema concreto, mas no el que recuerda tras un largo proceso de meditación, pues este es más rebuscado, artificioso y poco natural.

Para la sistematización de los datos, se tomó en cuenta el orden de escritura de los participantes, ya que las palabras anotadas en los primeros lugares se consideran más disponibles en el léxico de los informantes, pues son las que han surgido primero en su mente al plantearles un tema concreto. Para evitar que los informantes supieran con antelación los temas, y así poder obtener el léxico disponible real, el cuadernillo fue entregado totalmente en blanco y el encuestador indicaba, en voz alta, el tema o centro de interés. Con esto, también se evitaba que los encuestados se concentraran en otros temas futuros y se lograba que enfocaran su concentración en el centro de interés pertinente.

En este estudio, se utilizaron los dieciséis centros de interés que se han utilizado en el PPD, los cuales fueron propuestos desde el trabajo de Gougenheim et al (1964), y se añadieron dos centros de interés retomados de la investigación de Ríos (2007): «saludos y despedidas» y «temas de conversación», puesto que esta autora encontró diferencias léxicas en los mismos en función del sexo de los participantes. Además, se añadieron los centros «los deportes», «los colores» y «la higiene y el cuidado personal», debido al interés en conocer posibles diferencias en relación con la variable de sexo en respuesta a la creencia estereotipada sobre los géneros y al análisis constante de esta variable en los diferentes estudios de disponibilidad léxica. Como resultado los 21 centros de interés fueron: 1) partes del cuerpo, 2) la ropa, 3) partes de la casa sin los muebles, 4) los muebles de la casa, 5) alimentos y bebidas, 6) objetos colocados en la mesa para la comida, 7) la cocina y sus utensilios, 8) la escuela, 9) calefacción e iluminación, 10) la ciudad, 11) el campo, 12) medios de transporte, 13) trabajos del campo y del jardín, 14) los animales, 15) juegos y diversiones, 16) profesiones y oficios, 17) saludos y despedidas, 18) temas de conversación, 19) los colores, 20) los deportes, 21) la higiene y el cuidado personal.

Las encuestas fueron anónimas, aunque se completaron con algunos datos de los encuestados que aportaron variables sociodemográficas y de interés sociolingüístico:

edad, sexo, nivel socioeconómico, tipo de centro educativo en el cual realizó sus estudios de secundaria (privado o público), y área de procedencia (rural, urbana o mixta). Estas permiten un análisis más profundo de la relación entre las variables y el uso del léxico por parte de los hablantes de la muestra.

Después de la recolección de datos, se procedió a la estandarización de estos (basada en Valencia, 1997; Ortolano, 2005; Ríos, 2007), es decir, se establecieron los criterios para seleccionar las palabras dadas por los informantes y poder cargarlas en el programa de análisis Dispalex. Entre estos criterios estaban: la eliminación de los términos repetidos en un mismo centro de interés de la encuesta, el respeto por el orden en el que el informante haya anotado sus respuestas, la escritura de todas las palabras en minúscula con la letra inicial en mayúscula; además, se adoptó la norma ortográfica seguida por la RAE y por los principales diccionarios del español; en los extranjerismos, si no aparecían en los diccionarios españoles, se adoptó la grafía original; se aceptaron todas las marcas comerciales escritas correctamente; cada marca se incluyó como una entrada diferente; se aceptó la variación de género para palabras que pudieran tener significados diferentes en su acepción masculina o femenina; se adoptó la forma no marcada del paradigma, es decir, los verbos en infinitivo, y los sustantivos y adjetivos en singular; no se escribieron las palabras ilegibles; para cualquier centro, se aceptaron adjetivos, debido al posible análisis cualitativo que pudiera surgir de los mismos; y se aceptó lo máximo posible las palabras dichas por los encuestados, incluyendo las groserías.

Para el análisis de los resultados, se utilizó el programa de acceso gratuito en línea Dispalex, el cual permitió hacer análisis cuantitativos que arrojaron los índices de disponibilidad léxica, de cohesión, el número de palabras totales, el número de vocablos y el promedio de palabras por informante, así como las comparaciones, de acuerdo con las variables seleccionadas para la investigación.

ANÁLISIS DE RESULTADOS

Para empezar, es importante conocer las principales características de la población encuestada. Estas se muestran en la Tabla 1: el número de estudiantes hombres y mujeres es equivalente (92 estudiantes para cada sexo); aunque hay estudiantes en todos los niveles socioeconómicos, prevalecen en frecuencia los estratos bajo (60 estudiantes, es decir 32.6%) y medio bajo (94 estudiantes, es decir 51.1%); la mayoría de estudiantes proviene del área urbana (156 estudiantes, es decir 84.7%) y de colegios públicos (122 estudiantes, es decir 66.3%); y aunque la muestra de estudiantes se encuentra en el rango de los 14 a los 28 años, la mayoría se encuentra en el rango de los 16 a los 19 años (157 estudiantes, es decir el 85.3%).

Tabla1. Caracterización sociodemográfica

Variable	Frecuencia
Sexo	Hombres: 92 Mujeres: 92
Niveles socioeconómicos ¹	Bajo: 60 Medio bajo: 94 Medio alto: 24 Alto: 6
Tipo de centro donde realizó sus estudios secundarios	Público: 122 Privado: 58
Área de procedencia	Rural: 24 Urbano: 156 Mixto ² : 4
Edad	14 a 19 años: 160 20 a 25 años: 21 26 en adelante: 3

A continuación, se presentará el análisis por índices cuantitativos, por índice de disponibilidad, y, finalmente, el análisis comparativo acerca de algunas de las variables de estudio.

Análisis de índices cuantitativos

Los índices cuantitativos que se analizan a través del programa Dispoplex corresponden a las palabras totales, las palabras diferentes, las palabras por informante y el índice de cohesión.

El índice de palabras totales (PT) se refiere al total de palabras obtenidas para cada centro de interés. En este caso, se cuenta cada palabra individualmente, incluso si está repetida una o varias veces. El índice de palabras diferentes (PD) se refiere al total de palabras obtenidas para cada centro de interés, sin tener en cuenta las palabras repetidas, es decir, cada palabra diferente se cuenta una única vez. Este da cuenta de los vocablos más conocidos por el grupo muestral. El índice de palabras por informante (PI) corresponde al promedio de palabras por informante para cada centro de interés. Se obtiene dividiendo PT, de un centro de interés, entre el total de informantes de la muestra. El índice de cohesión (IC) corresponde al grado de coincidencias entre las palabras de un centro de interés, lo

1 El nivel socioeconómico se midió en función de los estratos socioeconómicos, clasificados del 1 al 6, informados por cada participante en la hoja inicial de la encuesta. Sin embargo, la plataforma Dispoplex tiene predefinidos los niveles socioeconómicos como bajo, medio bajo, medio alto y alto. Por tanto, en función de estas categorías, los estratos socioeconómicos se organizaron así: estrato 1, nivel bajo; estrato 2 y 3, nivel medio bajo; estrato 4 y 5, nivel medio alto y estrato 6, nivel alto.

2 El área de procedencia mixta se refiere a personas que viven en zonas rurales, pero dentro del perímetro urbano de la ciudad, es decir, en barrios aledaños o fincas.

cual permite saber si un grupo es compacto (es decir, si tiene muchas coincidencias), o difuso (si tiene pocas coincidencias). Este índice toma un valor entre 0 y 1; si tiende a 1, indica que el grupo es más cerrado o compacto.

De acuerdo con los índices cuantitativos (Tabla 2), el total de palabras dadas por los informantes es de 50 444, mientras que el total de palabras diferentes es de 9 423, con un promedio de 13.05 palabras por informante. Discriminando estos índices cuantitativos por áreas de interés, se puede encontrar que el centro con el mayor número de palabras es «los animales» (3 792), seguido por «comidas y bebidas» (3 618), «el cuerpo humano» (3 347) y «la ciudad» (3 124); mientras que los centros de interés con el menor número de entradas son «calefacción e iluminación» (1 210), «trabajos del campo y del jardín» (1 320), «objetos colocados en la mesa para la comida» (1 720), «los muebles de la casa» (1 742) e «higiene y cuidado personal» (1 814). En los otros centros de interés, se encuentra un promedio de 2 343 palabras.

Tabla 2. Índices cuantitativos

Nº	Centro de interés	PT	PD	PI	IC
1	El cuerpo humano	3 347	305	18.19	0.06
2	La ropa	2 703	413	14.69	0.04
3	Partes de la casa	2 110	279	11.47	0.04
4	Los muebles de la casa	1 742	329	9.47	0.03
5	Comidas y bebidas	3 618	633	19.66	0.03
6	Objetos colocados en la mesa para la comida	1 720	239	9.35	0.04
7	La cocina y sus utensilios	2 188	350	11.89	0.03
8	La escuela (muebles y materiales)	2 781	537	15.11	0.03
9	Calefacción, iluminación y medios de airear un recinto	1 210	224	6.58	0.03
10	La ciudad	3 124	699	16.98	0.02
11	El campo	2 679	668	14.56	0.02
12	Medios de transporte	2 339	269	12.71	0.05
13	Trabajos del campo y del jardín	1 320	430	7.17	0.02
14	Los animales	3 792	556	20.61	0.04
15	Juegos y distracciones	2 201	618	11.96	0.02
16	Profesiones y oficios	2 547	653	13.84	0.02
17	Saludos y despedidas	2 402	466	13.05	0.02
18	Temas de conversación	2 021	708	10.98	0.02
19	Los colores	2 733	183	14.85	0.08
20	Los deportes	2 053	357	11.16	0.03
21	La higiene y el cuidado personal	1 814	507	9.86	0.02

Por otro lado, en relación con el número de palabras diferentes, se encuentra que las mayores frecuencias están en los centros «temas de conversación» (708), «la ciudad» (699), «el campo» (668), «comidas y bebidas» (633) y «juegos y distracciones» (618); y las menores frecuencias, en los centros «partes de la casa» (279), «medios de transporte» (269), «objetos colocados en la mesa para la comida» (239), «calefacción e iluminación» (224) y «los colores» (183). Se puede decir que no es directamente proporcional la totalidad de palabras a la cantidad de palabras diferentes, ya que, como se puede observar en los datos presentados anteriormente, los centros de interés con mayor número de palabras totales no se corresponden con los centros con mayor número de palabras diferentes. Esto coincide con la investigación de A. López (2008) y la de Carcedo (1998), en las que un número mayor de palabras no coincide necesariamente con un número mayor de vocablos.

Teniendo en cuenta esta diferencia, es importante observar el índice de cohesión. De este modo, se puede notar que los centros de interés más compactos son «los colores» (0.08) y «el cuerpo humano» (0.06). Sin embargo, priman los centros de interés abiertos o con mayor variedad en las respuestas de los encuestados, con un índice de cohesión de 0.02 («la ciudad», «el campo», «trabajos del campo y del jardín», «juegos y distracciones», «profesiones y oficios», «saludos y despedidas», «temas de conversación» y «la higiene y el cuidado personal»).

Análisis por índices de disponibilidad léxica

Una vez realizado el análisis por índices cuantitativos, resulta importante e interesante realizar el análisis por índices de disponibilidad léxica, el cual permite conocer más profundamente los resultados al interior de cada centro de interés. El índice de disponibilidad léxica se mide, siguiendo la fórmula creada por López Chávez y Strassburger Frías (1987), que se usa para el Proyecto Panhispánico y se puede ver en la Figura 1:

$$D(P_j) = \sum_{i=1}^n \left(\frac{f_{ij}}{I} \right) e^{-2.3 \left(\frac{i-1}{n-1} \right)}$$

En donde:

- | | |
|---|---|
| $D(P_j)$ =disponibilidad de la palabra j. | e =número natural (2.718181818459045). |
| n =máxima posición alcanzada en el centro de interés. | f_{ij} =Frecuencia absoluta de la palabra j en la posición i. |
| I =número de posición de que se trata. | i =número de informantes que participaron en la encuesta. |
| j =índice de la palabra en cuestión. | |

Figura 1. Fórmula de disponibilidad léxica

Por tanto, como se mencionaba antes, la medición de la disponibilidad léxica, a través del índice de disponibilidad, permite analizar cada palabra más allá de su frecuencia de aparición, considerando también su posición y el número de informantes que participan en la encuesta. Así, se dispone de una respuesta más clara sobre cuál es la disponibilidad de las palabras analizadas con un valor entre 0 y 1 (donde 1 es la disponibilidad más alta; y 0, la más baja). Entonces, «la palabra más disponible es aquella que el individuo actualiza de inmediato, es decir, la que acude a su mente en forma instantánea ante el estímulo temático dado por la situación comunicativa» (Valencia, 1997, p. 198).

De acuerdo con el análisis realizado, las 10 palabras más disponibles son: *hola* (0.85368), *fútbol* (0.83171), *azul* (0.81186), *amarillo* (0.75153), *rojo* (0.70809), *cabeza* (0.67526), *cocina* (0.63351), *ojo(s)* (0.62715), *verde* (0.62144) y *ventilador* (0.60227).

De esta información, llama la atención la palabra *ventilador*, que forma parte del centro de interés «calefacción e iluminación», así como las palabras *azul*, *amarillo* y *rojo*, del centro de interés «los colores», que resultaron ser de los centros de interés con menor número de palabras diferentes. Sin embargo, el resto de palabras sí formó parte de los centros de interés con mayor número de palabras. Para una visión más detallada de la disponibilidad, en la Tabla 3 se observan las primeras cinco palabras más disponibles por cada centro de interés.

Tabla 3. Palabras con mayor índice de disponibilidad léxica por cada centro de interés

Centro de interés	Palabras con mayor índice de disponibilidad		Centro de interés	Palabras con mayor índice de disponibilidad	
El cuerpo humano	Cabeza	0.67526	Medios de transporte	Moto	0.42789
	Ojo(s)	0.62715		Carro	0.37798
	Nariz	0.57718		Avión	0.37158
	Dedo(s)	0.52413		Carro(s)	0.29761
	Pie(s)	0.47216		Bus	0.29514
La ropa	Camisa	0.50630	Trabajos del campo y del jardín	Sembrar	0.31331
	Pantalón	0.44191		Cultivar	0.31099
	Shorts	0.29700		Ordeñar	0.28757
	Media(s)	0.29516		Podar	0.27825
	Zapato(s)	0.23422		Cosechar	0.17625
Partes de la casa	Cocina	0.63351	Los animales	Perro	0.43051
	Baño	0.46630		Gato	0.36349
	Sala	0.42822		Vaca	0.27576
	Patio	0.39885		León	0.25430
	Garaje	0.28746		Caballo(s)	0.23510

Centro de interés	Palabras con mayor índice de disponibilidad		Centro de interés	Palabras con mayor índice de disponibilidad	
Los muebles de la casa	Comedor	0.45633	Juegos y distracciones	Fútbol	0.42406
	Cama	0.43941		Xbox	0.29457
	Silla(s)	0.36360		Baloncesto	0.19879
	Sofá	0.33006		Parqués	0.17742
	Televisor	0.22160		Ajedrez	0.17230
Comidas y bebidas	Gaseosa	0.47172	Profesiones y oficios	Abogado	0.32257
	Arroz	0.47012		Médico	0.24927
	Carne	0.35032		Psicólogo	0.24667
	Agua	0.34750		Profesor	0.22054
	Pollo	0.32601		Ingeniero	0.21894
Objetos colocados en la mesa para la comida	Plato(s)	0.50234	Saludos y despedidas	Hola	0.85368
	Tenedor	0.45056		Chao	0.57657
	Vaso(s)	0.40036		Ole	0.27831
	Cuchara	0.38959		Hasta luego	0.20718
	Servilleta(s)	0.33677		Adiós	0.20428
La cocina y sus utensilios	Olla(s)	0.58201	Temas de conversación	Estudio	0.35238
	Estufa	0.46063		Política	0.20373
	Nevera	0.34355		Sexo	0.20193
	Plato(s)	0.34010		Trabajo	0.14246
	Licuada	0.29539		Familia	0.13447
La escuela (muebles y materiales)	Silla(s)	0.41603	Los colores	Azul	0.81186
	Tablero	0.40451		Amarillo	0.75153
	Pupitre(s)	0.39197		Rojo	0.70809
	Cuaderno(s)	0.30290		Verde	0.62144
	Lápiz	0.25750		Negro	0.58306
Calefacción, iluminación y medios de airear un recinto	Ventilador	0.60227	Los deportes	Fútbol	0.83171
	Aire acondicionado	0.42551		Natación	0.58380
	Bombillo(s)	0.34718		Baloncesto	0.47587
	Lámpara(s)	0.28906		Voleibol	0.40169
	Ventana(s)	0.22699		Patinaje	0.33163
La ciudad	Carro(s)	0.55017	La higiene y el cuidado personal	Jabón	0.39862
	Casa(s)	0.53519		Bañar	0.32277
	Edificio(s)	0.41356		Desodorante	0.28829
	Moto(s)	0.36316		Cepillar	0.22261
	Parque(s)	0.31367		Cepillo	0.18853
El campo	Vaca(s)	0.50323			
	Caballo(s)	0.38940			
	Árbol(es)	0.38184			
	Pasto	0.29643			
	Animal(es)	0.26403			

Análisis por variables sociodemográficas

Para el análisis por variables sociodemográficas se tuvieron en cuenta los promedios por informante, debido a la disparidad en la cantidad de informantes para cada una de las variables (la única variable en la que se contó con igual número de participantes fue la de sexo). Las variables analizadas fueron: sexo, nivel socioeconómico, edad, área de procedencia y tipo de centro donde se culminaron los estudios de bachillerato.

Por el tipo de información tabulada y analizada en el programa Dispox, no se hace posible realizar análisis estadísticos que permitan decir con certeza si las diferencias halladas entre los grupos analizados son significativas. Esto se debe a que el programa no hace un registro sobre lo dicho por cada informante en relación con las variables sociodemográficas usadas, sino que hace análisis globales, tomando como unidad de análisis básica cada palabra informada.

Ahora bien, para sortear esta situación, en la literatura previamente mencionada se suele hacer descripciones teniendo en cuenta el promedio de palabras por informante, lo cual da una visión general de lo que ocurre con cada variable sociodemográfica, por medio de comparaciones sobre las diferencias y similitudes encontradas, en función de la cantidad de palabras mayor o menor reportada por los informantes. Por tanto, es importante tener claridad sobre el hecho de que las comparaciones realizadas a continuación son de tipo descriptivo, y que no necesariamente llegan a ser significativas estadísticamente hablando.

En relación con la variable de sexo, se encontró que, en promedio, las mujeres contestaron 12.5 palabras y los hombres 12.3 (hubo una pequeña diferencia de 0.16 palabras). Las mayores diferencias se encontraron en los centros «la escuela (muebles y materiales)» (con promedios por informante de 13.8 —hombres—, 16.4 —mujeres— y una diferencia de 2.6 a favor de las mujeres); «comidas y bebidas» (18.4 —hombres—, 20.9 —mujeres— y una diferencia de 2.5 a favor de las mujeres); «los deportes» (12 —hombres—, 10.3 —mujeres— y una diferencia de 1.7 a favor de los hombres); «los colores» (14 —hombres—, 15.7 —mujeres— y una diferencia de 1.7 a favor de las mujeres); y «partes de la casa» (12 —hombres—, 10.9 —mujeres— y una diferencia de 1.1 a favor de los hombres). Para el resto de los centros de interés, se encontraron diferencias pequeñas (en promedio de 0.36). De esta variable, se concluye que priman más las similitudes que las diferencias, puesto que solamente se pueden observar diferencias de entre una y tres palabras en 5 de los 21 centros estudiados.

En cuanto al nivel socioeconómico, se encontró que el nivel bajo tuvo un promedio de palabras por informante de 13.9; el nivel medio bajo, de 13.3; el nivel medio alto, de 14.6; y el nivel alto, de 14.3. En general, estos promedios priman en la mayoría de

centros de interés, con lo que se da lugar a que el nivel socioeconómico medio alto ocupe el primer lugar en cantidad de palabras por informante.

Ahora bien, los centros de interés en donde se encontraron mayores diferencias para esta variable fueron «comidas y bebidas» (con promedios por informante de 20 —bajo—, 19.3 —medio bajo—, 20.2 —medio alto— y 18.8 —alto—); «objetos colocados en la mesa para la comida» (9.7 —bajo—, 9 —medio bajo—, 8.9 —medio alto— y 13 —alto—); «el campo» (14.4 —bajo—, 14.1 —medio bajo—, 15.9 —medio alto— y 17.1 —alto—); «trabajos del campo y del jardín» (7.5 —bajo—, 6.7 —medio bajo—, 8.2 —medio alto— y 6.5 —alto—); «los animales» (20.9 —bajo—, 20.1 —medio bajo—, 22.2 —medio alto— y 19.7 —alto—); y «temas de conversación» (11 —bajo—, 10.7 —medio bajo—, 12.4 —medio alto— y 10.5 —alto—). Como se puede observar, no necesariamente el nivel socioeconómico alto obtiene los mayores promedios; así como el bajo, los menores promedios. Esto significaría que las diferencias estarían más asociadas al centro de interés, que al nivel socioeconómico del informante.

Por otro lado, los resultados de la variable de área de procedencia muestran un promedio de palabras por informante de 15.6 palabras para el área urbana, 15.3 para el área rural y 15.2 para el área mixta, por lo que, a simple vista, no hay diferencias entre las áreas. Sin embargo, se destacan algunos centros de interés, como «comidas y bebidas» (con promedios de palabras por informante de 19.8 —urbano—, 17.6 —rural— y 19.3 —mixto—); «la higiene y el cuidado personal» (10.1 —urbano— y 8.7 —rural y mixto—); «la ciudad» (17.1 —urbano—, 16 —rural— y 16.4 —mixto—); «el campo» (14.5 —urbano—, 14.9 —rural— y 14.7 —mixto—); «la cocina y sus utensilios» (12 —urbano—, 12.4 —rural— y 11 —mixto—); y «los animales» (20.5 —urbano—, 22 —rural— y 21.2 —mixto—).

Se puede observar que, aunque las diferencias en el promedio de palabra por informante no son altas, sí aparecen diferencias asociadas a determinados centros de interés, las cuales están en relación con el área de procedencia. Por ejemplo, el área rural se destaca sobre las otras áreas en el centro «los animales» (en donde hay una diferencia de 1.5 palabras del área rural sobre la urbana, y de 0.75 palabras sobre la mixta), y en «el campo» (en donde hay una diferencia de 0.35 palabras del área rural sobre la urbana, y de 0.18 palabras sobre la mixta); mientras que el área urbana se destaca sobre la rural en el centro «la higiene y el cuidado personal» (con una diferencia de 1.3 palabras del área urbana sobre la rural y la mixta), o en «la ciudad» (con una diferencia de 1.1 palabras del área urbana sobre el área rural, y de 0.65 palabras sobre el área mixta).

Acerca de la variable de tipo de centro donde se culminaron los estudios de bachillerato, se encontró que los promedios de palabras por informante fueron de 12 palabras

para colegios privados y de 13.3 para colegios públicos, con lo que se destacó el colegio público con 1.3 palabras de diferencia. En esta variable, los centros de interés en donde se hallaron mayores diferencias fueron «la cocina y sus utensilios» (con promedios de palabras por informante de 11.2 —privado—, 12.2 —público— y una diferencia de 1 palabra a favor de los colegios públicos); «medios de transporte» (11.8 —privado—, 13.1 —público— y una diferencia de 1.3 palabras a favor de los colegios públicos); «trabajos del campo y del jardín» (6.5 —privado—, 7.5 —público— y una diferencia de 1 palabra a favor de los colegios públicos); «los animales» (19.8 —privado—, 21 —público— y una diferencia de 1.2 palabras a favor de los colegios públicos); y «juegos y distracciones» (10.8 —privado—, 12.6 —público— y una diferencia de 1.8 palabras a favor de los colegios públicos). Para el resto de centros de interés, la diferencia promedio es de 0.4 palabras, mayoritariamente a favor de los colegios públicos. De esta variable se infiere que los estudiantes con mayores promedios de palabras fueron los que estudiaron y culminaron sus estudios secundarios en colegios públicos.

Finalmente, para la variable de edad, se encontró que los promedios de palabras por informantes fueron de 12.4 para las edades de 14 a 19 años, 12.8 para las edades de 20 a 25 años y de 15 para las edades de 26 años en adelante. Se evidencia que el número de palabras es directamente proporcional a las edades de los informantes; es decir, a mayor edad, mayor cantidad de palabras dadas por estos. Un ejemplo de esto se observa en los siguientes centros de interés: «la ropa» (con promedios de 13.8 —14 a 19 años—, 14.7 —20 a 25 años— y 17.5 —26 años en adelante—); «comidas y bebidas» (18 —14 a 19 años—, 19.3 —20 a 25 años— y 25.2 —26 años en adelante—); «la cocina y sus utensilios» (11.6 —14 a 19 años—, 12.8 —20 a 25 años— y 14.7 —26 años en adelante—); y «profesiones y oficios» (14.3 —14 a 19 años—, 15.4 —20 a 25 años— y 16.2 —26 años en adelante—).

DISCUSIÓN Y CONCLUSIONES

En relación con los centros de interés destacados en otras investigaciones, según el número de vocablos, Rui y Carvajal (2005) encontraron que el centro con mayor número de vocablos fue «la ciudad»; y el de menor cantidad, «partes del cuerpo». Por otro lado, Carcedo (1998) identificó que los centros de interés más productivos fueron: «comidas y bebidas», «la ciudad», «profesiones y oficios» y «partes del cuerpo»; y los menos productivos: «calefacción, iluminación y medios de airear un recinto», «la cocina» y «trabajos del campo y del jardín». En el presente estudio, los centros de interés con mayor número de vocablos fueron: «los animales», «comidas y bebidas», «el cuerpo humano» y «la ciudad»; mientras que los centros de interés con menor número de entradas fueron: «calefacción e iluminación», «trabajos del campo y del jardín»,

«objetos colocados en la mesa para la comida» y «los muebles de la casa». De tal forma, los resultados hallados en esta investigación coincidieron más con los encontrados por Carcedo (1998), como se puede constatar también en sus análisis por variables. Por otra parte, mientras que, para Rui y Carvajal (2005), el centro de menor producción fue «partes del cuerpo», en este trabajo fue uno de los más productivos.

Ahora bien, es importante decir que, al realizar el análisis comparativo por variables sociodemográficas, se encontraron algunas diferencias en variables como el área de procedencia, el colegio en donde se estudió, el estrato socioeconómico y la edad. Sin embargo, estas diferencias, más que explicarse por las variables mismas, se evidencian en determinados centros de interés. Como lo afirma Carcedo (1998, p. 221): «El volumen medio de palabras que un informante produce depende del estímulo (centro de interés), ya que unos centros pueden mostrar mucha riqueza léxica y otros muy pobre desarrollo léxico, esto se encontró en las siguientes variables: área de procedencia, estrato socioeconómico».

Esto muestra otra coincidencia con este mismo estudio, ya que en las mismas variables se pudo observar que las diferencias eran explicadas por el centro de interés (como en el caso de la variable de área de procedencia rural, la cual presentó más producción en los centros de interés «el campo» y «los animales», por encima de las otras dos áreas de esta misma variable). Por lo tanto, dada la cercanía del centro de interés con la variable, el estímulo es más significativo y genera una mayor productividad léxica. Esto es confirmado por Ortolano (2005), pues, de acuerdo con sus resultados de investigación, las variables sociales no son totalmente determinantes, es decir, no parecen incidir en el tipo de respuestas de los individuos de manera significativa.

Pasando ahora a la variable de sexo, una de las más investigadas en los estudios de disponibilidad léxica, también se pueden destacar los resultados de Ortolano (2005), quien encontró que en esta no hay diferencias significativas cuantitativamente hablando, sino que las diferencias son cualitativas. Por ejemplo, los hombres usan más un lenguaje soez. Esto es similar a lo encontrado en este estudio, en el centro de interés «saludos y despedidas», en el que los hombres usan groserías en mayor cantidad y con mayor disponibilidad que las mujeres.

En esta misma variable, López (2008) encontró que las mujeres demuestran conocer más vocabulario que los hombres en la mayoría de centros de interés, con la salvedad del centro «trabajos del campo y del jardín». Este resultado contrasta con la presente investigación, en la que se encontró que, en promedio, las mujeres contestaron 0.16 palabras más que los hombres, es decir, menos de una palabra de diferencia, por lo que hay mayores similitudes que diferencias entre hombres y mujeres. De este mismo modo,

Guerra y Gómez (2003) y Carcedo (1998) no encontraron diferencias cuantitativas en la variable de sexo, por lo que prima un grado de similitud entre estos.

Por otro lado, respecto a la variable edad, se pudo ver que, al igual que en los estudios de Montenegro (2010), Ortolano (2005) y Matéus Ferro y Santiago Galvis (2006), a mayor edad, mayor productividad léxica. Sin embargo, resultaría interesante discutir esta conclusión con otros grupos poblacionales que no sean estudiantiles o, incluso, con estudios longitudinales en los grupos ya estudiados.

Por último, de este trabajo de investigación se puede concluir que son necesarios los estudios en disponibilidad léxica en Colombia, debido a la falta de información e investigación en esta área, la cual puede dar herramientas para la descripción sincrónica del lenguaje colombiano, por un lado, de cara al conocimiento sociolingüístico del español en este contexto, y por otro lado, como base para los currículos y las listas lexicográficas a tener en cuenta en la enseñanza del español de Colombia como lengua extranjera.

Además, en el contexto educativo, los estudios de disponibilidad léxica permiten evaluar el estado léxico de los estudiantes, con lo que dejan ver el aprendizaje que los estudiantes tienen en sus áreas de estudio. Esto resulta clave, pues da pie para pensar en una futura investigación longitudinal con la misma población de este estudio, para conocer los cambios en el léxico disponible una vez que finalice el pregrado.

REFERENCIAS

- Bartol, J. (2006). La disponibilidad léxica. *Revista Española de Lingüística*, 36(1), 379-396.
- Carcedo, A. (1998). Sobre las pruebas de disponibilidad léxica para los estudiantes de español/LE. *RILCE*, 14(2), 205-224.
- Gougenheim, G., Michéa, R., Rivenc, P., & Savageot, A (1964). *L'elaboration du Français fondamental (I Degré). Étude sur l'elaboration d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Grupo de investigación Dispolex (2003). *DispoLex. Investigación léxica*. Recuperado de <http://www.dispolex.com/>
- Guerra, L., & Gómez, M. (2003). *Español de los medios de comunicación: aspectos de disponibilidad léxica*. Trabajo presentado en el XIV Congreso Internacional de ASELE, Burgos, España. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/14/14_0357.pdf
- Hernández, N. (2005a). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en lingüística y psicología. *IV Congreso Internacional sobre*

- la Adquisición de las Lenguas del Estado* (págs. 1-16). Salamanca: Ediciones Universidad de Salamanca.
- Hernández, N. (2005b). *Hacia una Teoría Cognitiva Integrada de la Disponibilidad Léxica: El Léxico Disponible de los Estudiantes Castellano–Manchegos* (Tesis doctoral). Salamanca: Ediciones Universidad de Salamanca.
- Lewandowsky, T. (1992). *Diccionario de lingüística*. Madrid: Cátedra.
- López, A. (2008). *La disponibilidad léxica en las secciones bilingües de español de Polonia*. Polonia: Centro Virtual Cervantes.
- López Chávez, J., & Strassburger Frías, C. (1987). Otro cálculo del índice de disponibilidad léxica. En *Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. México D. F.: UNAM.
- López, H. (1999). *Léxico disponible en Puerto Rico*. Madrid: Arco Libros.
- Mateus Ferro, G. E., & Santiago Galvis, A. W. (2006). Disponibilidad léxica en estudiantes bogotanos. *Folios*, 1(24), 3-26
- Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les langues modernes*, 47, 334-338.
- Montenegro, R. (2010). Estudio de disponibilidad léxica en estudiantes de tercero y sexto primaria de escuelas públicas del departamento de Guatemala. *Informes breves de investigación*, 1. USAID–Reforma Educativa en el Aula.
- Ortolano, B. (2005). Disponibilidad léxica de los estudiantes de Ayamonte. *Interlingüística*, 16(2), 847-857.
- Ríos, G. (2007). Diferencias léxicas entre el hombre y la mujer en tres centros de interés: saludos, temas de conversación y despedidas. *Filología y Lingüística*, 33(1), 151-166.
- Rui, G., & Carvajal, J. (2005). *Disponibilidad léxica en alumnos de la carrera de Castellano y Literatura de la ULA Táchira*. Trabajo presentado en la I Jornada Estudiantil de Investigación Lingüística, Universidad de Los Andes, Táchira, Venezuela. Recuperado de http://servidor-opsu.tach.ula.ve/profeso/garcia_mar/sil/RUIYCARVAJARINFORMEFINAL.pdf.
- Suárez, M. (2011). *Interaprendizaje holístico de matemática*. Ecuador: Editorial Planeta.
- Suárez, M., & Tapia, A. (2012). *Interaprendizaje de estadística básica*. Ecuador: Ibarra.
- Valencia, A. (1997). Disponibilidad léxica, muestreo y estadísticos. *Revista semestral de Lingüística, Filología y Traducción*, 2, 197-226.