

Disponibilidad léxica en estudiantes bogotanos

Bogotá, Students' Lexical Availability

Geral Matéus Ferro
Álvaro William Santiago Galvis*

Resumen

Este artículo presenta la investigación realizada por la Universidad Pedagógica Nacional y el Instituto Caro y Cuervo sobre el léxico disponible en estudiantes bogotanos. Primero se desarrollan el enfoque metodológico y algunos conceptos básicos (por ejemplo, léxico frecuente, léxico disponible e índice de disponibilidad léxica). En seguida se explica la relación entre los resultados cuantitativos y la productividad y la riqueza léxicas de los sujetos. Finalmente, se describe cómo la disponibilidad léxica determina la cohesión léxica de los centros de interés.

Palabras clave:

Léxico disponible, índice de disponibilidad léxica, índice de cohesión léxica, centro de interés, unidad léxica, productividad y riqueza léxica.

Abstract

This paper shows some results of the research *Léxico disponible en estudiantes bogotanos* carried out by the Universidad Pedagógica Nacional and the Instituto Caro y Cuervo. First, the methodological approach and some basic concepts (i. e. frequent vocabulary, available vocabulary and lexical availability index) of the research are developed. Then, the quantitative results are linked to the lexical productivity and lexical richness of the individuals. Finally, it is explained how lexical availability index determines the lexical cohesion of the interest focus.

Key words:

Available Vocabulary, lexical availability index, lexical cohesion index, interest focus, lexical unit, lexical productivity and lexical richness.

Artículo recibido el 25 de marzo de 2006 y aprobado el 3 de octubre de 2006.

* Profesores del Departamento de Lenguas de la Universidad Pedagógica Nacional.
asantiago@uni.pedagogica.edu.co y gmateus@pedagogica.edu.co

Presentación

La disponibilidad léxica es el campo de investigación que en lingüística, específicamente en el ámbito de los estudios fundamentados en léxico-estadística, tiene como objetivo la determinación y el análisis del léxico disponible de un sistema verbal; esto es, busca indagar sobre la norma léxica, ya sea de una lengua en general o de una comunidad de habla en particular.

Los estudios sobre disponibilidad léxica se iniciaron en Francia en la década de los años cincuenta del siglo pasado, con una finalidad concreta: la enseñanza del francés como lengua extranjera. Por aquella época, el trabajo se orientó hacia la elaboración de manuales para la enseñanza del francés, en los que la inclusión de un vocabulario básico y usual se constituía en un elemento central. Como resultado de esas investigaciones, en 1964, Georges Gougenheim, René Michéa, Paul Rivenc y Aurélien Sauvageot publicaron el trabajo *L'élaboration du français fondamental*, que ha sido considerado como el texto pionero y básico en el área de los estudios sobre disponibilidad léxica¹.

En el mundo hispánico, López Morales inició los trabajos sobre este tema con la investigación *Disponibilidad léxica de los escolares de San Juan* (1973), que sirvió de base teórica y metodológica para las investigaciones posteriores que se han dado en lengua española². Sin embargo, el desarrollo de este tipo de estudios en lengua española ha tenido lugar a partir de los años noventa, bajo la iniciativa y orientación del profesor López Morales³. Así, se han adelantado, entre otros, estudios de léxico disponible en España: Madrid, Castilla y León, Castilla-La

Mancha, Andalucía, Valencia, Castellón, Cataluña, Baleares, Aragón, País Vasco, Cantabria, Canarias, Asturias, Salamanca, Valencia. En América se han desarrollado trabajos en Puerto Rico, República Dominicana, México, Venezuela (Coro), Argentina (Buenos Aires) y Chile. Todos ellos encaminados hacia el análisis de la norma léxica del español con base en el léxico disponible de estudiantes que todavía no inician su formación universitaria.

Como parte de los estudios que se adelantan en el ámbito hispanohablante sobre disponibilidad léxica, la Universidad Pedagógica Nacional y el Instituto Caro y Cuervo desarrollaron el proyecto 526/2004 *Léxico Disponible en Bogotá*⁴; el cual centró su atención en el estudio de la disponibilidad léxica de los estudiantes bogotanos de educación básica primaria, grado quinto, y de educación media, grado once. Este artículo muestra, entonces, parte de los resultados del estudio que sobre disponibilidad léxica se adelantó con la población escolar de Bogotá.

Objetivo

El proyecto se planteó como meta dar cuenta de cuál es la *disponibilidad léxica* de los estudiantes bogotanos; de tal forma que se trazó como objetivo fundamental la realización del diagnóstico de la disponibilidad léxica de los estudiantes bogotanos pertenecientes a los grados quinto de educación básica primaria y once de educación media, a partir de una muestra representativa de los planteles educativos de la ciudad de Bogotá. El proyecto se propuso, además, establecer la progresión léxica de los estudiantes bogotanos, a través de la comparación de los índices de disponibilidad léxica obtenidos para cada grado; así como analizar la incidencia de factores extralingüísticos (edad, género, tipo de colegio y estrato socioeconómico) en los resulta-

¹ Otro de los trabajos que se considera básico en las primeras investigaciones sobre disponibilidad léxica corresponde al realizado por William F. Mackey (1971), *Le vocabulaire disponible du Français*.

² Entre los trabajos que pueden considerarse como pioneros en el contexto hispanohablante, se encuentran los desarrollados en la década de 1980 por Orlando Alba, en República Dominicana, y Pedro Benítez, en Madrid, España.

³ Cabe señalar que López Morales coordina en la actualidad el Proyecto Panhispánico de Estudio de Léxico Disponible, el cual agrupa los trabajos que al respecto están desarrollándose a ambos lados del Atlántico; este proyecto tiene como objetivo elaborar *Diccionarios de disponibilidad léxica* para las diversas zonas del mundo hispánico.

⁴ El equipo de investigación estuvo conformado por: Álvaro William Santiago Galvis, coordinador (Universidad Pedagógica Nacional); María Clara Henríquez y Nancy Roza, coordinadoras (Instituto Caro y Cuervo); Jaime Ruiz Vega, coinvestigador (Universidad Pedagógica Nacional); Geral Matéus Ferro, coinvestigador (Universidad Pedagógica Nacional); y Leonardo Ramírez Sarmiento, monitor de investigación (Universidad Pedagógica Nacional).

dos de la disponibilidad léxica de los estudiantes bogotanos⁵.

Muestra

Para la definición y selección de la muestra del estudio, se siguieron los parámetros determinados por el proyecto Panhispánico, los cuales, a su vez, siguen los lineamientos establecidos por los primeros estudios de disponibilidad léxica, en el sentido de trabajar con sujetos que aún no estén influidos por los tecnicismos adquiridos en escuelas profesionales y facultades universitarias o en la vida laboral, pero que ya hayan superado las etapas básicas de la adquisición de la lengua. Además, se consideró la necesidad de atender a los objetivos pedagógicos de comparación de los índices de disponibilidad léxica entre distintos niveles de escolaridad. De este modo, se eligieron dos grupos que permitieran cumplir con estos requerimientos en el momento de seleccionar la muestra: estudiantes de quinto y once grados de los colegios de la capital colombiana, como ya se señaló.

Sin embargo, a este criterio general fue preciso añadir algunos particulares para atender las características sociales específicas de la ciudad de Bogotá, pues ésta se encuentra organizada administrativamente en 20 localidades y en seis estratos socioeconómicos, definidos de acuerdo con ciertos condicionantes habitacionales. En consecuencia, se eligió una muestra representativa de instituciones educativas⁶ teniendo en cuenta el número de éstas y de estudiantes por localidad, así como el tipo de

colegio (oficial / no oficial) y su clasificación socioeconómica. De acuerdo con las cifras que ofrecía en 2004 la Secretaría de Educación de Bogotá, para el año 2003, en lo que respecta al número de instituciones por localidad al igual que la población en edad escolar en éstas, se estableció que se trabajaría con 11 de las 20 localidades (Ciudad Kennedy, Engativá, Suba, Ciudad Bolívar, San Cristóbal, Bosa, Usaquén, Rafael Uribe Uribe, Usme, Fontibón y Puente Aranda), por cuanto eran las que presentaban, en términos porcentuales, el mayor número de población en edad escolar.

Una vez realizado el respectivo análisis estadístico, con base en los datos de la Secretaría de Educación, se estableció que se trabajaría en 46 instituciones escolares (23 oficiales y 23 no oficiales), en las que se les aplicarían pruebas a 20 estudiantes (10 de grado once y 10 de grado quinto; de estos 10 informantes por grado, cinco corresponderían a sexo femenino y, obviamente, los otros cinco a sexo masculino). Según esto, el universo poblacional estaría conformado por 920 informantes: 460 de grado once y 460 de grado quinto. Sin embargo, debido a factores de diverso orden, la muestra real estuvo conformada por 911 informantes. De éstos, 459 (50,4%) correspondieron a género femenino y 452 (49,6%), a género masculino; 439 (48,2%) informantes de colegios oficiales y 472 (51,8%) de colegios no oficiales; 455 (49,9%) de grado 11° y 456 (51,1%) de grado 5°; 135 (14,8%) informantes pertenecían a estrato alto, 357 (39,2%) a estrato medio y 419 (45,9%) a estrato bajo. La tabla 1 presenta una síntesis de la distribución de la muestra según la localidad, el estrato, el tipo de colegio y el género, este último aspecto de acuerdo con el grado:

Al discriminar la población por grado, se tiene que de los 455 informantes de grado once, 227 (49,8%) fueron de género masculino y 228 (50,1%) de género femenino; 220 (48,4%) pertenecían a colegios oficiales y 235 (51,6%) a colegios no oficiales; 67 (14,7%) informantes de estrato alto, 178 (39,1%) de estrato medio y 210 (46,2%) de estrato bajo. Por su parte, con la población del grado quinto, se trabajó con 225 (49,1%) informantes de género masculino y 231 (50,9%) de género femenino; 219

⁵ En este artículo, por limitaciones de espacio, sólo se presentan los resultados referidos al diagnóstico de la disponibilidad léxica de los estudiantes bogotanos, así como el índice de cohesión léxica, de acuerdo con los grados establecidos, esto es, atendiendo únicamente a la variable extralingüística *edad*.

⁶ Para efectos de la definición de la muestra, se trabajó con base en los datos estadísticos que ofrecía en el año 2004 la Secretaría de Educación de Bogotá en su página web. Según estos datos, en 2003, en la ciudad de Bogotá había 386 colegios oficiales y 2.549 no oficiales. Sin embargo, esta diferencia no era directamente proporcional a la población estudiantil, ya que el 47,5% estudiaba en planteles oficiales y el 52,5% lo hacía en no oficiales, de ahí que en el proceso de definición de la muestra se determinó que el número de colegios oficiales y no oficiales debería ser igual, 23 por cada uno.

Tabla 1

Localidad	Estrato			Tipo de colegio		Informantes 11		Informantes 5	
	Alto	Medio	Bajo	Of	Nof	Masc.	Fem.	Masc.	Fem.
Ciudad Kennedy	18	59	40	3	3	29	29	29	30
Engativá	20	80	20	2	4	30	30	29	31
Suba	40	40	40	2	4	29	31	30	30
Ciudad Bolívar	-	20	80	2	3	26	24	25	25
San Cristóbal	-	20	60	2	2	20	20	19	21
Bosa	-	20	60	2	2	20	20	20	20
Usaquén	37	18	20	1	3	19	18	19	19
Rafael Uribe Uribe	-	20	20	2	-	9	11	10	10
Usme	-	-	40	2	-	10	10	10	10
Fontibón	20	40	19	3	1	20	20	19	20
Puente Aranda	-	40	20	2	1	15	15	15	15
Totales	135	357	419	23	23	227	228	225	231

(48%) de colegios oficiales y 237 (52%) de colegios no oficiales; 68 (14,9%) de estrato alto, 179 (39,3%) de estrato medio y 209 (45,8%) de estrato bajo.

Instrumento y recolección de la información

Teniendo en cuenta que el trabajo realizado fue de carácter descriptivo, la técnica que se utilizó para desarrollar el proyecto fue la de la prueba escrita⁷. De esta forma, el instrumento básico para la recolección de los datos léxicos consistió en la aplicación de una prueba escrita de carácter asociativo, puesto

⁷ Para la aplicación de la prueba, se diseñó un cuadernillo conformado por dos partes, según el modelo que ofrecen Samper, Bellón y Samper (2003): una ficha inicial y las listas para escribir las unidades léxicas por cada centro de interés. La ficha inicial ocupaba la primera página del cuadernillo y estaba integrada por nueve ítems que tenían como finalidad obtener la siguiente información sociológica de cada informante: 1. sexo, 2. edad, 3. lugar de nacimiento, 4. si es originario de Bogotá o no, 5. barrio donde vive, 6. lugar de origen de los padres, 7. instituciones educativas en las que ha estudiado, 8. ocupación de los padres y 9. nivel de escolaridad de los padres. Todos estos datos se previeron a fin de realizar luego la estratificación posterior de la muestra. Las siguientes páginas del cuadernillo estaban destinadas para la aplicación de la prueba en sí. En cada página se encontraban cuatro columnas con 30 líneas cada una. Las columnas estaban numeradas en forma ascendente del 1 hasta el 16, de tal forma que cada número correspondía a cada uno de los centros de interés. Nótese que los centros de interés no aparecían escritos en el cuadernillo para hacer, en la medida de lo posible, más natural y espontánea la recolección de la información léxica.

que ésta se constituye en el instrumento “que mejor posibilita, en condiciones experimentales, la actualización (o evocación) de unidades léxicas con poca estabilidad estadística” (Gómez y Gómez, 2004: 47). Además, este tipo de pruebas se basa en los llamados *centros de interés*, los cuales hacen referencia a un conjunto de “estímulos temáticos relativamente amplios que persiguen provocar las respuestas de los sujetos entrevistados” (Samper, Bellón y Samper, 2003: 46); de esta forma, los centros de interés se tornan en estímulos cuya finalidad es hacer que los individuos actualicen su conocimiento léxico. Así, a cada informante se le entregó un cuadernillo en el que consignaron sus respuestas, discriminadas por cada centro de interés; de tal forma que se le presentaba el estímulo temático o centro de interés (por ejemplo, partes del cuerpo) y se le daban dos minutos para que escribiera todas las unidades léxicas relacionadas con éste y que pudiera evocar durante ese tiempo (véase nota 8).

Como el proyecto se acogió a los principios metodológicos generales del Proyecto Panhispánico, los centros de interés están determinados previamente para conservar la posibilidad de comparar los resultados de este proyecto con los de otras latitudes. En este orden de ideas, se contemplaron los 16 centros de interés que se consideran ya tradicionales en los estudios panhispánicos sobre léxico disponible,

por cuanto muchas de las investigaciones que ya se han realizado los han empleado. Dichos centros de interés son:

1. Partes del cuerpo.
2. Prendas de vestir.
3. Partes de la casa (sin incluir muebles).
4. Muebles de la casa (sin incluir electrodomésticos).
5. Alimentos y bebidas.
6. Objetos puestos en la mesa para la comida.
7. La cocina.
8. La escuela/el colegio y útiles escolares.
9. Instrumentos para iluminar, airear o calentar un recinto.
10. La ciudad.
11. El campo.
12. Medios de transporte.
13. Trabajos del campo y el jardín.
14. Animales.
15. Juegos y distracciones.
16. Profesiones y oficios.

Para la aplicación de las pruebas, en primer lugar, se contactaron las instituciones educativas, según los criterios establecidos en la definición de la muestra; una vez obtenida la autorización por parte de éstas, se realizaron las distintas pruebas. Para ello, en cada institución se solicitó un lugar adecuado que contara con la infraestructura mínima para que los estudiantes pudieran diligenciar las pruebas de forma cómoda y sin interrupciones⁸.

⁸ La prueba, en la medida de lo posible, se realizó en cada grado por separado. En primera instancia, se informaba a los estudiantes sobre la naturaleza y los objetivos del trabajo; luego se completaba la ficha sociológica y se procedía a explicar la prueba. Para esto se les decía que debían escribir bajo el número de la lista correspondiente, en columna, todas las palabras que viniesen a su memoria relacionadas con el tema que se les iba indicado; al respecto se les aclaraba que tenían dos minutos para escribir las palabras, sin reparar en ello cuando tuviesen dudas ortográficas. Además, se les explicaba que cada columna estaba conformada por 30 espacios para escribir en ellos las palabras, con la posibilidad de continuar al final del cuadernillo en los casos en los que se sobrepasara este número. Se previno a los estudiantes de abstenerse de preguntar en voz alta, para no sugerir palabras a sus compañeros, y de no volver a escribir en los centros ya trabajados.

Léxico frecuente y léxico disponible

El estudio del componente léxico de un sistema verbal puede abordarse desde distintas ópticas y perspectivas, según los intereses y objetivos de los investigadores. Así, este estudio puede centrar el interés en la naturaleza, los constituyentes y los procesos de formación de las unidades léxicas que integran un sistema léxico, tal es el caso de la morfología; o en el significado que dichas unidades portan, en consecuencia, analizar aspectos como la denotación, la connotación, la referencia, la sinonimia, la polisemia, los campos léxicos, de los que se ocupa la semántica léxica; o en la manera en que las palabras se almacenan en el léxico mental del individuo o cómo accede éste a ellas, preocupaciones asociadas al procesamiento léxico, propias de la psicolingüística; o en la relación palabra –ideología– cultura, que constituye una de las áreas de interés de la etnolingüística; o en las variedades lingüísticas reflejadas en la diversidad léxica asociadas a variables como la edad, el género, el nivel sociocultural, la ocupación o profesión, temas de la sociolingüística; o en la relación del léxico con la organización del conocimiento, aspecto por el que se interesa la lingüística cognitiva.

Otra forma de afrontar el estudio del componente léxico, que no riñe con las anteriores y que, por el contrario, se nutre de éstas y a la vez les ofrece elementos para fortalecer sus argumentaciones, es la que tiene que ver con la determinación de la norma léxica, ya sea de una lengua o de un grupo particular de hablantes, es decir, los estudios que centran su interés en determinar y cuantificar el inventario léxico de una comunidad de habla. Estos estudios, fundamentados en los desarrollos de la léxico-estadística, han recurrido a las listas de palabras como recurso metodológico y técnica de estudio. Dentro de esta línea de trabajo pueden establecerse, en principio, dos enfoques, que corresponden a dos momentos en el desarrollo de la léxico-estadística: los estudios basados en el criterio de *frecuencia léxica* y los basados en la noción de *disponibilidad léxica*.

Los estudios centrados en la frecuencia léxica se fundamentan en los siguientes principios: a) las

emisiones verbales son producto de la integración de unidades léxicas; b) algunas de estas unidades léxicas son más frecuentes que otras, esto es, que aparecen un número mayor de veces en un texto que otras; c) esta frecuencia se puede cuantificar y, a partir de ello, establecer cuál es la norma léxica de determinada lengua. Así pues, tienen como objetivo cuantificar la frecuencia de ocurrencia de una palabra y para ello se sustentan en el hecho de que “algunas palabras se presentan de forma más frecuente que otras” (Rodríguez, 1980: 21). En estos estudios prima entonces el número de veces que la palabra aparece o se actualiza en el discurso –oral o escrito–. Con base en este criterio se puede, en consecuencia, establecer el índice de frecuencia de las palabras; de tal forma que los estudios del léxico basados en el concepto de frecuencia dan cuenta de las palabras con *mayor estabilidad estadística*⁹. Para esto recurren a la elaboración de las listas de frecuencias de palabras, que están ordenadas de acuerdo con el índice de frecuencia de ocurrencia de las palabras en un corpus; a partir de estos listados se crean los *diccionarios de vocabulario básico*. De esta forma, los trabajos sustentados en la frecuencia de uso como indicador tienen como finalidad el conteo del empleo de unidades léxicas dentro de un corpus.

Según lo anterior, los antecedentes inmediatos de los estudios de disponibilidad léxica, en los que se ubican los trabajos desarrollados por los investigadores franceses entre los años cincuenta y sesenta del siglo pasado, se sustentan en el criterio de frecuencia léxica. Estos trabajos tenían como objetivo “fijar el tipo de vocabulario, evocado de forma más o menos inmediata mediante una asociación de ideas y graduado en diferentes niveles de dificultad para la enseñanza del francés como lengua extranjera” (Gómez y Gómez, 2004: 11). De esto puede plantearse, de un lado, que esos trabajos tenían una intención de didáctica orientada hacia la enseñanza de la lengua francesa y, del otro, que se basaban en la frecuencia

⁹ Con este término se hace referencia a las formas léxicas que suelen presentar un índice de aparición en un discurso, ya sea oral o escrito, relativamente constante y alto.

de aparición de los términos como el indicador más importante para definir “el vocabulario que manejan los hablantes de una lengua determinada” (Gómez y Gómez, 11).

Así, en los trabajos franceses se realizaba una selección del léxico usual de la lengua francesa con base en criterios de frecuencia. No obstante, la utilización del método basado en la frecuencia de uso puso en evidencia que muchos términos conocidos, e incluso comunes y de empleo cotidiano, no aparecían en las listas de vocabulario frecuente; asimismo, permitió establecer que las palabras de clase cerrada (gramaticales) presentaban los mayores índices de frecuencia. Esto llevó a establecer unas diferencias básicas para el desarrollo de los estudios de disponibilidad léxica. La primera de ellas planteaba la distinción entre palabras *temáticas* y *atemáticas* (Michéa, 1950)¹⁰.

Las palabras *temáticas* son las que se presentan en situaciones específicas y su aparición está determinada por el tema o la necesidad comunicativa que tenga el hablante; estas palabras “designan los seres y los objetos, y son, en su mayor parte, palabras concretas” (Samper, Bellón y Samper 2003: 30); a esta categoría pertenecen los adjetivos, los verbos, y en especial, los sustantivos. Por su parte, las palabras *atemáticas* se presentan con más frecuencia en cualquier tipo de discurso y su aparición no está determinada por el tema ni el tipo de texto; éstas corresponden a las llamadas palabras gramaticales, integradas por artículos, preposiciones, conjunciones, verbos auxiliares; también forman parte de las palabras *atemáticas* un buen número de adjetivos (bonito, bueno, feo), verbos comunes (hacer, trabajar...) y algunos nombres muy generales. Esta distinción llevó a que Michéa (1953) estableciera,

¹⁰ En términos de la clasificación de las palabras según la categoría gramatical, las palabras *temáticas* corresponden a las denominadas *palabras de clase abierta*, es decir, las que tienen significado léxico y se expresan en forma de nombres, adjetivos, verbos y adverbios. Así mismo, las palabras *atemáticas* se relacionan con las *palabras de clase cerrada*, es decir, con las “palabras *funcionales* que, tradicionalmente conforman la arquitectura de las oraciones y no poseen ningún contenido” (Reeves, Hirsh-Pasek y Golinkoff, 1999: 183). El conjunto de palabras cerradas es limitado en cuanto al número de unidades, con respecto al de las palabras de clase abierta, que tiende –en principio– a ser ilimitado.

en segundo lugar, la diferencia entre *léxico frecuente* y *léxico disponible*.

De acuerdo con este autor, el individuo cuenta con dos tipos de vocabulario: el *frecuente* (o *básico*) y el *disponible*. El vocabulario frecuente o básico está relacionado directamente con las palabras atemáticas, dado que éstas le permiten al individuo estructurar formalmente su construcción discursiva; este vocabulario puede entenderse como el conjunto de palabras más frecuentes de una lengua, actualizado con independencia del tema tratado o de la situación comunicativa.

Además, el vocabulario disponible está asociado a las palabras temáticas y se actualiza en función de la situación comunicativa; de tal forma que este vocabulario se constituye en el “caudal léxico utilizable en una situación comunicativa dada” (Butrón, 1991: 80). Así, el vocabulario disponible se entiende como el conjunto de palabras que los hablantes tienen almacenadas en el lexicón mental, potencial, no actualizado, y cuyo uso está condicionado por el tema concreto de la interacción verbal en la que participe. Esto se sustenta en la hipótesis de que existe en el lexicón mental una serie de términos que no se actualiza a menos que sea necesario para comunicar información muy específica; de esta forma, este léxico disponible o potencial se concreta según la situación comunicativa, pero ello no quiere decir que no forme parte de la competencia léxica del individuo.

Lo anterior permitió establecer que los listados de palabras elaborados con base en criterios de frecuencia privilegiaban unos vocablos pero dejaban por fuera otros, es decir, que centraban su atención sólo en el léxico frecuente y dejaban de lado el léxico disponible; razón por la cual se estableció que el índice de frecuencia no era suficiente criterio para dar cuenta de la norma léxica de una lengua o de una comunidad lingüística.

Con el fin de superar este escollo e indagar sobre ese léxico potencial, en la metodología de trabajo se introdujeron los estímulos verbales temáticos con el fin de que el hablante actualizara dicho léxico; “de ahí que empezara a trabajarse con centros de interés que actuaban como estímulos para actualizar el cau-

dal léxico disponible” (Gómez y Gómez, 2004: 12); esto atendiendo al hecho de que este léxico sólo se hace evidente de acuerdo con el tema o la situación comunicativa. De esta forma, se plantean las pruebas escritas basadas en los centros de interés que se asumen como áreas asociativas –partes del cuerpo, prendas de vestir, partes de la casa, muebles de la casa, animales, profesiones y oficios, etc.–, las cuales se convierten en estímulos verbales para que se dé la producción artificial del léxico disponible.

Índice de disponibilidad léxica

De lo planteado hasta ahora, puede concluirse que los trabajos centrados en frecuencia se constituyen en los antecedentes inmediatos para los estudios del léxico basados en disponibilidad. Como se ha visto, en el primer caso se establece el orden de las unidades léxicas y su importancia a partir del número de veces que éstas aparezcan en las muestras, lo cual lleva a establecer el índice de frecuencia. A este respecto, los estudios que indagan sobre la disponibilidad léxica asumen como criterio no sólo el número de apariciones de la unidad léxica, frecuencia, sino que, además, tienen en cuenta el orden de aparición de éstas en las listas¹¹; la correlación de estos dos elementos permite establecer el *índice de disponibilidad léxica*.

Según lo expuesto, los trabajos en léxico-estadística han ido refinándose de tal forma que se ha pasado del cálculo del índice de frecuencia al cálculo del índice de disponibilidad léxica. Para ello se ha recurrido al diseño y desarrollo de fórmulas y algoritmos que ofrezcan mejores resultados con mayor confiabilidad. Estas herramientas matemáticas tienen la posibilidad de relacionar y ponderar la frecuencia de una unidad léxica con el lugar de aparición de ésta en las listas; esto por cuanto, como se anotó, al considerar sólo la frecuencia como parámetro para el cálculo se pierde la información

¹¹ El orden de aparición de la palabra en las listas, tanto individuales como grupales, recibe el nombre de *índice de espontaneidad*; éste señala el hecho de que unas palabras son evocadas por el individuo en forma más rápida que otras y, en consecuencia, ocupan las primeras posiciones en las listas, lo cual lleva a plantear que poseen un índice de disponibilidad más alto.

relacionada con el léxico disponible; pero al incluir la posición se obtienen resultados más completos. Lo anterior, de acuerdo con el planteamiento que sostiene que “cuando una palabra está más disponible para un sujeto, éste la incluirá antes en su lista-respuesta del correspondiente centro de interés porque acudirá más pronto a su memoria” (Samper, Bellón y Samper, 2003: 103). De esta forma, para el cálculo del índice de disponibilidad adquiere gran importancia la posición que ocupa la palabra dentro del listado; por lo tanto, la palabra que aparece más veces en los listados no es necesariamente la más disponible, porque ello está supeditado al orden de aparición de la forma léxica en los listados. En consecuencia, el índice de disponibilidad léxica se constituye en una forma de organizar unidades léxicas por medio de fórmulas matemáticas que “a partir de la frecuencia y del orden de mención de las palabras entregan el número llamado *índice de disponibilidad* a base del cual se realiza la ordenación de las palabras correspondientes, de manera que las palabras de mayor índice irán primero” (Butrón, 1991: 80).

Una de las fórmulas desarrolladas para el cálculo del índice de disponibilidad léxica es la propuesta por los investigadores mexicanos López Chávez y Strassburger Frías (1991: 92-93), la cual tiene en cuenta: a) la posición absoluta de cada palabra en cada posición; b) la frecuencia absoluta de cada palabra; c) el número total de informantes; d) el número de posiciones que se alcanzan en la encuesta en cada centro de interés, y e) la posición en la cual aparece la palabra. Esta fórmula es la que, precisamente, ha utilizado el proyecto panhispánico, y la que sirvió de base para el análisis de la disponibilidad léxica de los estudiantes bogotanos. Este cálculo permitió obtener los siguientes valores: a) índice de disponibilidad léxica por centros de interés y según las variables (edad, género, tipo de colegio y estrato); b) frecuencia de aparición de las unidades léxicas por centro de interés; c) orden de aparición de las unidades léxicas por centro de interés; d) total de vocablos; e) total de vocablos diferentes, y f) promedio de unidades léxicas por centro de interés e informantes.

Índice de cohesión léxica

Además del índice de disponibilidad léxica, también se ha introducido en los trabajos el *índice de cohesión léxica*, noción desarrollada por Echeverría (1991), la cual permite clasificar los centros de interés en abiertos (difusos) y cerrados (compactos). Este índice es resultado de dividir el promedio de respuestas de los informantes en un centro de interés por el número de palabras diferentes en dicho centro.

De acuerdo con Echeverría, si se divide el promedio de respuestas por el número de palabras diferentes se obtiene un valor igual a 1, “siempre y cuando todos los sujetos hayan respondido las mismas palabras en su encuesta léxica, cosa por lo demás quimérica. Por el contrario, cuanto más variadas sean las respuestas, menor será el índice de cohesión; es decir, la dispersión es mayor” (1991: 62). Así, si el valor que se obtiene se acerca a 1, el índice de cohesión léxica es mayor; en consecuencia, se considera que el centro de interés es más compacto; pero si este valor se aleja de 1, el índice de cohesión es menor; por lo tanto, el centro de interés tiende a ser abierto o difuso.

Unidad de análisis

Si bien hasta ahora se han utilizado de forma indistinta las nociones de palabra y vocablo, es necesario aclarar que la unidad que se ha constituido en objeto de conteo para efectos de los distintos análisis estadísticos no corresponde necesariamente a la palabra, pese a que ésta sea la unidad básica de la léxico-estadística, por el hecho de que se puede identificar, aislar y contar fácilmente; sin embargo, los datos recolectados presentan elementos conformados por una palabra (*ojo, boca, cuaderno, carro*) y elementos conformados por más de una palabra, esto es, estructuras sintagmáticas complejas (*mesa de noche, cortar flores, cuidar animales, sala de profesores*). Según esto, se ha optado por la noción de *unidad léxica* como unidad de análisis porque “se puede considerar como unidad de significado en el lexicón mental [...] y puede estar formada por una o más palabras” (Gómez y Gómez, 2004: 28).

Por tanto, en el análisis de los datos se utilizará, preferiblemente, el término unidad léxica para hacer referencia a las respuestas de los informantes y sus respectivos análisis estadísticos; de ahí que en la presentación de los resultados se hable del total de unidades léxicas (UL) y del total de unidades léxicas distintas (ULD) como los elementos sobre los cuales se realiza el análisis.

Datos y resultados

En este apartado se presenta el análisis de los datos obtenidos con base en la aplicación de las pruebas a los 911 informantes. Este análisis se hizo a partir de las cifras que arrojó la herramienta informática G-Lex¹². Primero, en la presentación de los datos y resultados se ha tenido en cuenta el número de unidades léxicas (UL) y de unidades léxicas diferentes (ULD) para analizar la productividad y riqueza

léxicas de los estudiantes bogotanos (6.1); luego, se han analizado los índices de cohesión léxica (ICL), de forma general y comparados entre las dos poblaciones (6.2); y, finalmente, se han abordado los índices de disponibilidad léxica (IDL) de los estudiantes bogotanos (6.3). En cada apartado se han realizado consideraciones sobre los anteriores aspectos a la luz del comportamiento de la variable edad, asociada, como ya se explicó, con los estudiantes de los grados quinto y once de la población escolar de Bogotá.

Resultados generales

Productividad léxica. Ante todo, es posible analizar el número de unidades léxicas expresadas en general y discriminadas por cada centro de interés para determinar la productividad léxica, es decir, el número de vocablos expresados ante cada estímulo temático o centro de interés por los 911 informantes.

Tabla 2

Centros de interés		Total UL	%
1	Animales	22.402	9,8
2	Alimentos y bebidas	20.872	9,2
3	Partes del cuerpo	18.824	8,3
4	La escuela/colegio y útiles escolares	18.304	8,0
5	La ciudad	17.622	7,7
6	El campo	15.316	6,7
7	Prendas de vestir	15.314	6,7
8	Profesiones y oficios	14.639	6,4
9	Juegos y distracciones	13.596	6,0
10	La cocina	13.190	5,8
11	Medios de transporte	12.841	5,6
12	Partes de la casa	11.179	4,9
13	Objetos puestos en la mesa para la comida	10.083	4,4
14	Muebles de la casa	9.288	4,1
15	Trabajos del campo y el jardín	7.561	3,3
16	Instrumentos para iluminar, airear o calentar un recinto	6.981	3,1

¹² Herramienta desarrollada por el ingeniero Luis Guillermo Ortiz, profesor adscrito al Departamento de Tecnología de la UPN; dicha herramienta informática se sustenta en la fórmula propuesta por Juan López Chávez y Carlos Strassburger (1991) para el cálculo del índice de disponibilidad léxica. Esta herramienta permite entonces calcular el índice de disponibilidad léxica y la frecuencia léxica; de igual forma, permite obtener distintos reportes de acuerdo con cada una de las variables que se definieron. La herramienta ofrece tres tipos de archivos: diccionarios por centros de interés, índice de disponibilidad léxica y frecuencia, y posición de aparición de las unidades léxicas.

Así, el número de unidades léxicas (UL) obtenidas ascendió a un total de 228.012. Al analizar esta cifra puede establecerse que el promedio de UL por centro de interés es de 14.251; mientras que el promedio de UL por informante corresponde a 250. En la tabla 2 se presenta en orden descendente el número total de UL asociadas a cada centro de interés, CI.

De acuerdo con estos resultados, puede apreciarse que el CI más productivo corresponde a

Animales, con un total de 22.402 unidades léxicas, que equivale al 9,8% del total de unidades obtenidas; enseguida se encuentra *Alimentos y bebidas*, que presenta 20.872 unidades léxicas (9,2%). A estos dos centros de interés les siguen, en orden descendente, *Partes del cuerpo*, *La escuela/colegio y útiles escolares*, *La ciudad*, *El campo*, *Prendas de vestir y Profesiones y oficios*, que se encuentran por encima del promedio ya mencionado (14.521 UL).

Por su parte, *Juegos y distracciones*, *La cocina*, *Medios de transporte*, *Partes de la casa*, *Objetos puestos en la mesa para la comida*, *Muebles de la casa*, *Trabajos del campo y el jardín* e *Instrumentos para iluminar, airear o calentar un recinto* son los CI menos productivos, pues el número de UL se encuentra por debajo de la media; de éstos, *Instrumentos para iluminar, airear o calentar un recinto* se constituye en el centro menos productivo, en tanto presenta el menor número de UL (6.981). Como se muestra en la tabla, ocho de los CI superan el promedio, mientras que los restantes ocho, lógicamente, se encuentran por debajo del mismo. De esta manera, se evidencian diferencias en cuanto a la productividad de UL en los CI, que arrojan la existencia de centros de interés especialmente más productivos que otros, hecho que indica que las características de los centros inciden directamente en la posibilidad de asociar UL a los mismos. Las razones de estas diferencias de productividad entre los CI residirían, inicialmente, en las posibilidades de asociación semán-

tica de los centros según estos designen, por ejemplo, conjuntos específicos de objetos (*i.e. Animales, Partes del cuerpo*) o actividades variadas (*i.e. Trabajos del campo y el jardín, Juegos y distracciones*).

Riqueza léxica. Otro aspecto que revela los datos generales es el que tiene que ver con la riqueza léxica de los CI. Ésta se refiere al número de vocablos diferentes expresados en cada CI. A diferencia del anterior aspecto, en éste sólo se tienen en cuenta las unidades léxicas diferentes (ULD). De las 228.012 UL, solamente el 7,3%, 16.662, corresponde a ULD. Ante esta selección, podría pensarse en formular una relación entre productividad y riqueza léxica, por cuanto podría esperarse que los CI con mayor número de unidades léxicas tuvieran mayor número de unidades léxicas diferentes, es decir, que cuanto más productivo sea un CI, más rico debería ser en términos léxicos.

Sin embargo, a la luz de los datos, se refuta esta afirmación, y se comprueba una tendencia contraria a la formulación propuesta. La tabla 3 demuestra que no hay una relación directamente proporcional entre productividad y riqueza léxica. En su lugar, se evidencia que éstos son aspectos léxicos diferentes, derivados de las asociaciones establecidas respecto a contextos/estímulos de determinado CI con una tendencia inversa, pues algunos de los CI con menor número de UL están en los primeros lugares del listado de ULD:

Tabla 3

	Centros de interés	Total ULD	%
1	Juegos y distracciones	1.853	11,1
2	Profesiones y oficios	1.651	9,9
3	Trabajos del campo y el jardín	1.608	9,7
4	La ciudad	1.550	9,3
5	El campo	1.376	8,3
6	La escuela/colegio y útiles escolares	1.263	7,6
7	Alimentos y bebidas	1.037	6,2
8	La cocina	940	5,6
9	Instrumentos para iluminar, airear o calentar un recinto	771	4,6
10	Muebles de la casa	701	4,2
11	Animales	684	4,1
12	Objetos puestos en la mesa para la comida	681	4,1
13	Prendas de vestir	677	4,1
14	Partes de la casa	663	4,0
15	Partes del cuerpo	620	3,7
16	Medios de transporte	587	3,5

En contraste con la productividad léxica, en este caso solamente seis de los 16 CI rebasan el promedio de ULD (1.041). El centro que presenta el mayor número de ULD es *Juegos y distracciones*; en su orden le siguen *Profesiones y oficios*, *Trabajos del campo y el jardín*, *La ciudad*, *El campo* y *La escuela/colegio y útiles escolares*, que también están por encima del promedio. Los restantes se encuentran por debajo del promedio y los que presentan el menor número de unidades léxicas diferentes son *Partes del cuerpo* y *Medios de transporte*.

En consecuencia, existe cierta tendencia a que los CI con menor número de UL presenten el mayor número de ULD. Aquí se destaca el caso del centro *Trabajos del campo y el jardín*, que sólo cuenta con 7.561 UL (posición 15 en la tabla de productividad), pero presenta 1.608 ULD, ubicándose en la tercera posición de la tabla de riqueza léxica.

Para la productividad léxica, esto es, el número de unidades léxicas, los CI también presentan diferencias en lo que atañe a la riqueza léxica, es decir, el número de unidades léxicas diferentes. Esta vez, la delimitación de referentes-objetos en CI como *Animales* y *Partes del cuerpo*, por ejemplo, cierra la posibilidad de evocar numerosas ULD; por su parte, los CI referidos a actividades, *Juegos y distracciones*, *Profesiones y oficios* y *Trabajos del campo y el jardín*, parecen facilitar la riqueza de las unidades léxicas. Lo anterior lleva a establecer que centros de interés como *Animales* y *Partes del cuerpo* tienden a ser compactos o cerrados, mientras que centros como *Juegos y distracciones* y *Profesiones y oficios* tienden a ser más dispersos o abiertos, de ahí que los primeros presenten menores unidades léxicas diferentes en relación con éstos.

Variable edad. Ahora bien, vale la pena examinar las diferencias de productividad y riqueza léxicas según las dos poblaciones investigadas (estudiantes de 5° y 11°). Así, los informantes del grado once presentan mayor productividad que los de quinto grado. La diferencia es de 36.714 UL respecto a los estudiantes de grado quinto, es decir, que se produjo un 16,1% más de vocablos en 11° que en 5°.

En lo atinente a la riqueza léxica, se mantiene la diferencia cuantitativa entre los dos grupos; de tal forma que los estudiantes del grado 11° aportaron 5.535 ULD más que los de grado 5°. Los datos comparados de los dos grados se resumen en la tabla 4:

Tabla 4

Grado	Productividad léxica		Riqueza léxica	
	Total UL	%	Total ULD	%
11°	132.363	58,1	12.143	55,5
5°	95.649	41,9	9.754	44,5
	228.012	100	21.897	100

Un hecho para resaltar es que específicamente entre quinto y once las variaciones en la productividad léxica en los CI son de orden cuantitativo, mas no cualitativas. De los 16 CI, 14 conservan su lugar en el orden de productividad, y sólo *Partes del cuerpo* y *La escuela* intercambian sus posiciones. Tanto los CI más productivos (*Animales*, *Alimentos y bebidas*) como los de menor productividad (*Trabajos del campo y el jardín*, *Instrumentos para iluminar, airear,...*) mantienen sus características. En la tabla 5 se compendia esta semejanza en los datos entre las dos poblaciones.

No obstante estas semejanzas aumenta en uno, de 7 a 8, las cifras de CI que están por encima del promedio entre 5° y 11°. Además, es notable que en todos los centros de interés el número de UL (productividad léxica) es mayor en grado 11° que en 5°; constante que evidencia una mayor capacidad para evocar vocablos a medida que aumenta la edad de la población.

La riqueza léxica, además, presenta un comportamiento distinto entre los dos grupos de edades. A diferencia de la productividad léxica, hay mayores diferencias en los CI en el paso de quinto a once. Si bien son los mismos siete CI, aquellos que están por encima del promedio de ULD en los dos grados, cuatro de ellos cambian de lugar. No obstante, el centro de interés *Juegos y distracciones* es el más rico en términos léxicos en todos los estudiantes. El orden y el número de ULD en cada CI, según su riqueza léxica, se presentan en la tabla 6.

Tabla 5

Grado quinto			Grado once		
	Centro de interés	Total UL		Centro de interés	Total UL
1	Animales	9.947		Animales	12.455
2	Alimentos y bebidas	9.091		Alimentos y bebidas	11.781
3	La escuela / el colegio y útiles escolares	8.041	1▲	Partes del cuerpo	10.854
4	Partes del cuerpo	7.970	1▼	La escuela / el colegio y útiles escolares	10.263
5	La ciudad	7.428		La ciudad	10.194
6	Prendas de vestir	6.548		Profesiones y oficios	9.169
7	El campo	6.244		El campo	9.072
8	Juegos y distracciones	5.484		Prendas de vestir	8.766
9	Profesiones y oficios	5.470		Juegos y distracciones	8.112
10	La cocina	5.390		La cocina	7.800
11	Medios de transporte	5.213		Medios de transporte	7.628
12	Partes de la casa	4.760		Partes de la casa	6.419
13	Objetos puestos en la mesa para la comida	4.273		Objetos puestos en la mesa para la comida	5.810
14	Muebles de la casa	3.887		Muebles de la casa	5.401
15	Trabajos del campo y el jardín	3.024		Trabajos del campo y el jardín	4.537
16	Instrumentos para iluminar, airear o calentar un recinto	2.879		Instrumentos para iluminar, airear o calentar un recinto	4.102

Tabla 6

Grado quinto			Grado once		
	Centro de interés	Total ULD		Centro de interés	Total ULD
1	Juegos y distracciones	1.085		Juegos y distracciones	1.271
2	Profesiones y oficios	1.078	1▲	La ciudad	1.130
3	La ciudad	917	2▲	El campo	1.079
4	Trabajos del campo y el jardín	911	2▼	Profesiones y oficios	1.058
5	El campo	762	1▼	Trabajos del campo y el jardín	1.046
6	La escuela/colegio y útiles escolares	719		La escuela/colegio y útiles escolares	957
7	Alimentos y bebidas	664		Alimentos y bebidas	808
8	La cocina	536		La cocina	716
9	Animales	460	2▲	Instrumentos para iluminar, airear o calentar un recinto	557
10	Prendas de vestir	403	1▼	Animales	530
11	Objetos puestos en la mesa para la comida	393	1▼	Prendas de vestir	521
12	Instrumentos para iluminar, airear o calentar un recinto	389	1▲	Muebles de la casa	521
13	Muebles de la casa	385	1▲	Partes de la casa	513
14	Partes de la casa	380	1▲	Partes del cuerpo	506
15	Partes del cuerpo	348	4▼	Objetos puestos en la mesa para la comida	489
16	Medios de transporte	324		Medios de transporte	441

Del mismo modo, los nueve CI de menor riqueza léxica también presentan algunas variaciones de orden. Sin embargo, los estudiantes de los dos grados evocaron el menor número de ULD en el mismo CI: *Medios de transporte*.

Tal como sucedió en la producción léxica, en grado once hay indicadores numéricos mayores que en quinto. Los estudiantes de aquel grado expresaron más vocablos diferentes en todos los CI comparados con los del grado quinto de educación básica primaria.

Así las cosas, es posible evidenciar que tanto la capacidad para evocar UL –productividad léxica– como ULD –riqueza léxica– mejora en términos cuantitativos a favor de los estudiantes de grado once. Sin embargo, estas diferencias no conducen a grandes variaciones cualitativas entre los centros de interés en las dos poblaciones en cuestión.

Índice de cohesión léxica (ICL)

Otra manera de explicar los CI consiste en analizar el *índice de cohesión léxica* (Echeverría, 1991). Este índice es el resultado de dividir el promedio de respuestas de cada sujeto por el número de ULD en cada centro, con lo cual se da cuenta de la dispersión o coincidencia en las respuestas de los sujetos por CI. Al respecto, recuérdese que cuanto más cerca se encuentre el ICL de 1, se considera el centro de interés más compacto; mientras que si se aleja de ese punto de referencia, el CI se considerará difuso o abierto. La tabla 7 presenta los ICL de los CI ordenados de forma decreciente:

Tabla 7

	Centros de interés	ICL
1	Animales	0,03595
2	Partes del cuerpo	0,03333
3	Prendas de vestir	0,02483
4	Medios de transporte	0,02401
5	Alimentos y bebidas	0,02209
6	Partes de la casa	0,01851
7	Objetos puestos en la mesa para la comida	0,01625
8	La escuela/colegio y útiles escolares	0,01591
9	La cocina	0,01540
10	Muebles de la casa	0,01454
11	La ciudad	0,01248

12	El campo	0,01222
13	Instrumentos para iluminar, airear o calentar un recinto	0,00994
14	Profesiones y oficios	0,00973
15	Juegos y distracciones	0,00805
16	Trabajos del campo y el jardín	0,00516

Según estos datos, puede plantearse que centros como *Animales, Alimentos y bebidas y Partes del cuerpo humano* son más compactos, es decir, tienen mayor índice de cohesión, razón por la cual su ICL se acerca a 1; entre tanto, los centros *Juegos y distracciones, Profesiones y oficios y Trabajos del campo y el jardín* tienden a ser más difusos o abiertos, por cuanto presentan un menor número de unidades léxicas y mayor cantidad de vocablos diferentes, de ahí que el índice de cohesión se aleje del punto de referencia 1.

De acuerdo con Gómez y Gómez (2004: 83), los centros de interés, según el ICL, pueden clasificarse en tres grupos: *compactos o cerrados, de coincidencia media y difusos o abiertos*. Con base en esta clasificación y los ICL obtenidos, los CI pueden organizarse como sigue:

- *Centros compactos o cerrados:* Animales, Partes del cuerpo, Ropa, Medios de transporte y Alimentos y bebidas.
- *Centros con un grado de compactación medio:* Partes de la casa, Objetos puestos en la mesa para la comida, La escuela/colegio y útiles escolares, La cocina, Muebles de la casa, La ciudad y El campo.
- *Centros difusos o abiertos:* Instrumentos para iluminar, airear o calentar un recinto, Profesiones y oficios, Juegos y distracciones y Trabajos del campo y el jardín¹³.

¹³ Estos resultados guardan algunas semejanzas con los presentados por García y García (2004) en la investigación *Disponibilidad léxica en estudiantes preuniversitarios valencianos*. Allí, los CI compactos son *cuerpo humano, ropa, colores, medios de transporte y objetos colocados en la mesa para la comida*; los de coincidencia media de respuestas, *partes de la casa, la cocina, comidas y bebidas, iluminación, calefacción y ventilación, muebles de la casa y la escuela*; finalmente, los más difusos son *juegos y distracciones, y trabajos del campo y el jardín, y el campo*. Las faltas de coincidencias pueden explicarse por las diferencias culturales que generan distintas asociaciones semánticas en las dos sintopías (bogotana y valenciana).

Lo anterior se relaciona con la naturaleza misma del centro de interés, esto es, con la realidad misma a la que está asociado; de tal forma que los individuos comparten y tienen más conocimientos acerca de una realidad como los animales, en la que coinciden en más respuestas, mientras que su conocimiento de un ambiente como las actividades propias del campo y/o de la jardinería son menos comunes, de ahí que se tornen más difusas y el grado de coincidencia en las respuestas disminuya. Esto lleva a plantear que la naturaleza del centro de interés, entendida como la realidad a la que se asocia, determina en buena medida el índice de disponibilidad léxica, al haber centros que abarcan realidades más amplias o complejas, si se quiere (*Juegos y distracciones, Profesiones y oficios*), y otras más delimitadas (*El cuerpo*

humano, Medios de transporte), lo cual se refleja en el número de unidades léxicas que los conforman y que se encuentran disponibles para los hablantes en las mismas.

Variable edad. En lo que respecta a la comparación entre los grados quinto y once, el ICL en los centros de interés de los estudiantes de educación media (11°) superan en general los de los niños de educación básica primaria. Únicamente en el CI *Animales*, el ICL es superior en éstos que en aquellos.

A estas disparidades cuantitativas se suma que el orden de los CI en cada categoría de cohesión (*compacto, medio y difuso*) presenta variaciones entre las dos poblaciones estudiantiles; no obstante, hay algunas constantes:

Tabla 8

Grado quinto			Grado once		
	Centro de interés	ICL		Centro de interés	ICL
1	Partes del cuerpo	0,05022	1▲	Animales	0,05165
2	Animales	0,04742	1▼	Partes del cuerpo	0,04714
3	Prendas de vestir	0,03563	1▲	Medios de transporte	0,03802
4	Medios de transporte	0,03528	1▼	Prendas de vestir	0,03698
5	Alimentos y bebidas	0,03002		Alimentos y bebidas	0,03204
6	Partes de la casa	0,02747		Partes de la casa	0,02750
7	La escuela / el colegio y útiles escolares	0,02453	1▲	Objetos puestos en la mesa para la comida	0,02611
8	Objetos puestos en la mesa para la comida	0,02384	2▲	La cocina	0,02394
9	Muebles de la casa	0,02214	2▼	La escuela / el colegio y útiles escolares	0,02357
10	La cocina	0,02205	1▼	Muebles de la casa	0,02278
11	El campo	0,01797	1▲	La ciudad	0,01983
12	La ciudad	0,01776	2▲	Profesiones y oficios	0,01905
13	Instrumentos para iluminar, airear o calentar un recinto	0,01623	2▼	El campo	0,01848
14	Profesiones y oficios	0,01113	1▼	Instrumentos para iluminar, airear o calentar un recinto	0,01619
15	Juegos y distracciones	0,01108		Juegos y distracciones	0,01403
16	Trabajos del campo y el jardín	0,00728		Trabajos del campo y el jardín	0,00953

De acuerdo con los datos de la tabla 8, los cinco CI *compactos* en quinto grado son los mismos en el grado once, salvo algunas alteraciones en su orden. Los CI de *cohesión media* sí presentan más cambios, pues un CI pasa de ser difuso a tener el carácter de compactación media (*Profesiones y oficios*), mientras

que los otros cinco alteran su orden. Otro es el comportamiento de los CI *difusos*, puesto que tienden a mantenerse los mismos CI en los dos grados con ligeros cambios en sus posiciones.

Estos datos conducen a afirmar que las dos poblaciones son homogéneas en cuanto conservan la

coincidencia de respuestas asociadas a los distintos centros de interés, pues tan sólo dos de los 16 CI cambian de categoría entre los dos grados. Como complemento, también puede aseverarse que esta homogeneidad cualitativa en los índices de cohesión léxica en los dos grados sufre alteraciones que se evidencian por el incremento de los índices cuantitativos en grado once; variaciones explicables por el mayor desarrollo lingüístico y cognitivo de éstos últimos.

Índices de disponibilidad léxica

La disponibilidad léxica da cuenta de un conjunto de unidades léxicas ordenadas según la frecuencia y lugar de aparición en la lista en cada centro de interés, con el fin de analizar el potencial orden de

acceso a dichas unidades. En otras palabras, con base en estos datos se calcula el índice de disponibilidad léxica (ID) que, como se expuso, permite ordenar unidades léxicas de mayor a menor disponibilidad; ello al correlacionar la frecuencia y el lugar de aparición de cada unidad léxica.

A continuación se presentan los listados de unidades léxicas ordenados por índice de disponibilidad en cada centro de interés y luego se analizan algunas coincidencias y diferencias entre las unidades léxicas en los CI de los dos grados. Se han elegido de estas listas, a manera de ilustración, solamente los 20 primeros vocablos de cada centro; sin embargo, los cálculos del índice de disponibilidad se realizaron sobre la totalidad de unidades léxicas expresadas en cada centro de interés:

1. Partes del cuerpo				
	Grado 5°		Grado 11°	
	ID	Vocablo	Vocablo	ID
1	0,572553	ojo	ojo	0,657193
2	0,49861	nariz	cabeza	0,655335
3	0,467786	cabeza	nariz	0,554466
4	0,464609	boca	mano	0,525949
5	0,455525	pie	pie	0,515714
6	0,445664	mano	boca	0,476652
7	0,327198	pierna	dedo	0,470825
8	0,32576	brazo	brazo	0,463631
9	0,324042	dedo	pierna	0,429066
10	0,300423	corazón	corazón	0,335547
11	0,288807	estómago	uña	0,321653
12	0,265738	oreja	oreja	0,317924
13	0,197394	hígado	rodilla	0,262153
14	0,184934	uña	estómago	0,259268
15	0,178139	rodilla	cabello	0,230609
16	0,165655	lengua	hígado	0,22821
17	0,16265	cerebro	lengua	0,208441
18	0,15846	pulmón	pulmón	0,207065
19	0,153135	hueso	hombro	0,192355
20	0,145558	diente	cuello	0,192276

2. Prendas de vestir				
	Grado 5°		Grado 11°	
	ID	Vocablo	Vocablo	ID
1	0,691338	camisa	camisa	0,738661
2	0,679101	pantalón	pantalón	0,680046
3	0,57591	media	media	0,575191
4	0,465314	zapato	camiseta	0,522429
5	0,41937	saco	chaqueta	0,452955
6	0,395693	falda	zapato	0,42031
7	0,389159	camiseta	saco	0,412251
8	0,387692	chaqueta	pantaloneta	0,360502
9	0,310303	pantaloneta	falda	0,3304
10	0,243886	buzo	brasier	0,286508
11	0,201675	calzoncillos	buzo	0,254916
12	0,192383	blusa	blusa	0,250475
13	0,190597	brasier	boxer	0,248834
14	0,168265	cucos	corbata	0,208008
15	0,162536	vestido	bufanda	0,189387
16	0,149011	corbata	tenis	0,185421
17	0,138294	sudadera	jeans	0,184603
18	0,127973	tenis	sudadera	0,158524
19	0,115565	bufanda	calzoncillos	0,145277
20	0,106615	guante	tanga	0,140939

11. El campo

Grado 5°		Grado 11°	
ID	Vocablo	Vocablo	ID
1	0,494645	vaca	vaca 0,591009
2	0,483765	árbol	árbol 0,503082
3	0,429226	pasto	pasto 0,385491
4	0,317739	casa	caballo 0,375808
5	0,292718	animal	río 0,320023
6	0,280171	caballo	casa 0,292532
7	0,229379	río	perro 0,276393
8	0,215896	perro	finca 0,266938
9	0,194842	finca	gallina 0,262278
10	0,181396	gallina	campesino 0,227755
11	0,165423	flor	cultivo 0,219859
12	0,16203	montaña	animal 0,21322
13	0,136323	cultivo	montaña 0,165126
14	0,121156	campesino	flor 0,164062
15	0,116897	burro	burro 0,14316
16	0,110101	cabaña	cerdo 0,129547
17	0,108819	cerdo	pollo 0,108559
18	0,103846	planta	fruta 0,105067
19	0,077707	granja	quebrada 0,104657
20	0,077283	persona	oveja 0,104459

12. Medios de transporte

Grado 5°		Grado 11°	
ID	Vocablo	Vocablo	ID
1	0,69727	carro	bus 0,713837
2	0,674107	bus	avión 0,651084
3	0,563057	avión	carro 0,650249
4	0,436677	taxi	taxi 0,449676
5	0,350884	barco	barco 0,436195
6	0,275487	bicicleta	bicicleta 0,427596
7	0,269918	moto(cicleta)	Transmilenio 0,326431
8	0,264635	Transmilenio	moto(cicleta) 0,323245
9	0,23001	caballo	tren 0,320067
10	0,224952	tren	buseta 0,302574
11	0,185382	buseta	caballo 0,285862
12	0,180313	helicóptero	camión 0,215999
13	0,161466	camión	helicóptero 0,210786
14	0,122276	colectivo	lancha 0,195815
15	0,119297	camioneta	avioneta 0,191453
16	0,115912	burro	patín 0,190377
17	0,104856	lancha	colectivo 0,183495
18	0,097852	patín	patineta 0,178094
19	0,095698	avioneta	burro 0,1518
20	0,092613	"cicla"	camioneta 0,146016

13. Trabajos del campo y el jardín

Grado 5°		Grado 11°	
ID	Vocablo	Vocablo	ID
1	0,347913	sembrar	sembrar 0,4882
2	0,322736	cultivar	cultivar 0,390385
3	0,216346	ordeñar	ordeñar 0,240353
4	0,123498	cosechar	cortar 0,236431
5	0,106722	cortar	podar 0,212367
6	0,073356	agricultura	cosechar 0,20693
7	0,071781	plantar	regar 0,184746
8	0,070877	regar	arar 0,182443
9	0,070561	cuidar	recoger 0,141863
10	0,058868	podar	abonar 0,110416
11	0,057753	jardinero	cuidar 0,106894
12	0,049348	correr	limpiar 0,08533
13	0,048639	arreglar	jardinero 0,084844
14	0,048246	limpiar	plantar 0,073995
15	0,046344	regar	jardinería 0,070695
16	0,045407	cultivo	agricultor 0,060383
17	0,044859	comer	agricultura 0,059402
18	0,044382	ganadería	talar 0,05497
19	0,04383	trabajar	recolectar 0,054898
20	0,042493	alimentar	cocinar 0,052393

14. Animales

Grado 5°		Grado 11°	
ID	Vocablo	Vocablo	ID
1	0,774322	perro	perro 0,840012
2	0,623825	gato	gato 0,712327
3	0,53303	vaca	vaca 0,543583
4	0,483478	caballo	caballo 0,51585
5	0,461892	león	león 0,424497
6	0,359978	tigre	tigre 0,336285
7	0,276573	burro	gallina 0,325331
8	0,276127	gallina	burro 0,310922
9	0,25437	elefante	elefante 0,276127
10	0,24386	conejo	ratón 0,265051
11	0,232548	tiburón	ballena 0,244116
12	0,215294	ratón	pez 0,216141
13	0,201588	pez	pájaro 0,215239
14	0,186403	pato	jirafa 0,210681
15	0,181864	mico	tiburón 0,208477
16	0,175856	pollo	conejo 0,208204
17	0,169903	jirafa	pollo 0,197032
18	0,166326	culebra	pato 0,187999
19	0,165485	loro	mico 0,183861
20	0,165389	cerdo	oso 0,17404

15. Juegos y distracciones

	Grado 5°		Grado 11°	
	ID	Vocablo	Vocablo	ID
1	0,570714	fútbol	fútbol	0,637252
2	0,262082	escondidas	voleibol	0,323646
3	0,249334	cogidas	baloncesto	0,318
4	0,224548	voleibol	parqués	0,30237
5	0,173332	congelados	ajedrez	0,252767
6	0,156021	basquetbol	tenis	0,250401
7	0,152093	tenis	cartas	0,191324
8	0,141958	parqués	billar	0,180944
9	0,138084	baloncesto	dominó	0,177194
10	0,130002	montaña rusa	escondidas	0,168441
11	0,1149	ajedrez	Play Station	0,145183
12	0,113019	béisbol	Atari	0,131024
13	0,109705	Play Station	béisbol	0,129084
14	0,10742	correr	natación	0,127517
15	0,083528	dominó	golosa	0,121002
16	0,080646	rodadero	ping-pong	0,119282
17	0,080544	Atari	cogidas	0,112565
18	0,070935	Stop	basketball	0,104974
19	0,066483	ponchados	correr	0,10015
20	0,065886	columpio	microfútbol	0,093443

De estos datos se puede plantear:

- **Partes del cuerpo:** en este centro de interés, el vocablo más disponible en los dos grados es el mismo: *ojo*. Coinciden, además, 17 de los 20 vocablos iniciales, que corresponden al 85%. Así, tres UL, *cerebro*, *hueso* y *diente*, que aparecen en 5° grado en los 20 primeros lugares, son remplazadas en 11° por *cabello*, *hombro* y *cuello*.
- **Prendas de vestir:** de manera semejante al anterior CI, la coincidencia en las 20 primeras UL es del 85% entre los dos grados. Resulta interesante que los tres primeros vocablos, *camisa*, *pantalón*, *media*, concuerdan en ambos grados, incluso en el orden de disponibilidad. En grado once aparecen dos vocablos de origen extranjero, *boxer* y *jeans*, junto a *tanga*, que sustituyen a *cucos*, *vestido* y *guante* (quinto grado).
- **Partes de la casa (sin incluir muebles):** este CI presenta el mismo porcentaje de coincidencia en las 20 primeras unidades léxicas que en los anteriores centros de interés (85%). Del mismo modo que en *Prendas de vestir*, los primeros vocablos (*baño*, *cocina*, *sala* y *patio*) son idénticos

16. Profesiones y oficios

	Grado 5°		Grado 11°	
	ID	Vocablo	Vocablo	ID
1	0,430257	profesor	profesor	0,500259
2	0,316221	doctor	ingeniero	0,327901
3	0,207684	policía	abogado	0,303983
4	0,152305	abogado	médico	0,277889
5	0,146918	ama de casa	doctor	0,246584
6	0,142957	enfermero	arquitecto	0,232023
7	0,133238	secretario	enfermero	0,221002
8	0,12005	bombero	policía	0,218007
9	0,111706	futbolista	secretario	0,214937
10	0,109583	vendedor	conductor	0,188091
11	0,106027	médico	administrador	0,172252
12	0,101987	conductor	vendedor	0,166527
13	0,085314	arquitecto	estudiante	0,144379
14	0,080846	panadero	ama de casa	0,138545
15	0,076179	cantante	contador	0,137158
16	0,074908	barredor	aseador	0,116961
17	0,070847	constructor	celador	0,113954
18	0,068887	actor	panadero	0,112986
19	0,06786	ingeniero	bombero	0,111111
20	0,067853	mecánico	comerciante	0,106603

en los grados quinto y once. Por su parte, *pieza*, *piso* y *cama*, del listado de 5° son desplazadas por *ventana*, *habitación* y *balcón* en el listado de 11°.

- **Animales:** Así como en los anteriores tres CI, el porcentaje de semejanza en los vocablos es del 85%. El parecido de las listas es tal que las seis primeras UL, *perro*, *gato*, *vaca*, *caballo*, *león* y *tigre*, y las que ocupan las posiciones 7, 8 y 9 son las mismas en quinto y once (*burro*, *gallina* y *elefante*).
- **Instrumentos para iluminar, airear o calentar un recinto:** cierra el grupo de CI con identidad de las listas en los 20 primeros vocablos del 85% este centro. También aquí, en los dos grados, las cuatro primeras UL son las mismas: *bombillo*, *ventilador*, *lámpara* y *ventana*.
- **Muebles de la casa (sin incluir electrodomésticos):** a diferencia de los anteriores CI, en éste el vocablo más disponible en 5°, *silla*, no es el mismo en grado 11°, *cama*. No obstante, el porcentaje de semejanza de los 20 primeras UL en los dos grados es del 90%. Para los estudiantes

de 5º, los vocablos *mueble* –que hace parte del enunciado del CI– y *cajón*, son relevados por *cuadro* y *alacena* en 11º.

- **Alimentos y bebidas:** el grado de identidad en las 20 primeras UL persiste en este CI; es el mayor de los analizados hasta ahora: 95%. Las tres primeras unidades léxicas en las dos poblaciones son: *arroz*, *gaseosa* y *jugo*. Un hecho interesante tiene lugar entre estas dos listas, pues el único vocablo que varía de 5º a 11º es una bebida, *chocolate*, la cual se reemplaza por otra bebida: *cerveza*.
- **Medios de transporte:** al igual que en *Alimentos y bebidas*, el porcentaje de semejanza en la selección realizada en las dos listas es el más alto entre los demás CI: 95%. No obstante, el vocablo más disponible en 5º grado es *carro*, en tanto que en 11º es *bus*. El vocablo que ocupó la posición 20 en quinto grado es de la voz de origen popular para bicicleta, *cicla*, la cual desaparece, y en su lugar asciende *patineta*, en grado once.

Hasta ahora se aprecia una alta coincidencia cualitativa en la disponibilidad léxica de ambos grados, la cual sería índice de una identidad sociocultural en las dos poblaciones en cuanto a los referentes y asociaciones en los anteriores ocho centros de interés, que implica, en este caso, operaciones psicolingüísticas estadísticamente cercanas en sus resultados: similitud en la evocación de entidades mentales y léxicas (vocablos).

En el siguiente conjunto de centros de interés descienden un poco los porcentajes de semejanza en contraposición con el conjunto recién analizado¹⁴.

- **Objetos puestos en la mesa para la comida:** el porcentaje de identidad en los vocablos de los dos grados en las 20 unidades seleccionadas es del 80%. Mientras en el quinto el vocablo más disponible es *plato*, en once es *cuchara*. De las cuatro UL que reemplazan en grado once a las de quinto se resalta la aparición de un vocablo correspondiente a una palabra compuesta: *por-*

tavazos, tipo de palabra bastante escaso en esta selección.

- **La cocina:** en este centro también el porcentaje de semejanza es del 80%. Si bien los tres primeros vocablos son iguales en los dos grados (*estufa*, *olla*, *nevera*), el más disponible en quinto es *estufa*, en tanto que en once es *olla*.
- **La escuela/colegio y útiles escolares:** este es el tercer CI con 80% de igualdad en las unidades de las dos poblaciones. Del mismo modo que en el anterior CI, las dos listas comparten las UL iniciales (*lápiz*, *cuaderno*, *esfero*, *libro* y *borrador*) pero en distinto orden. Además, el primer vocablo en quinto es *lápiz* mientras que en once es *cuaderno*. Es de resaltar que de las cinco UL que son reemplazadas por otras entre ambos grados sale el vocablo *computador*, que se esperaría como “inevitable” en esta selección en grado once.
- **El campo:** corresponde al cuarto CI que concluye el grupo de coincidencia del 80%. En este centro, los vocablos, *vaca*, *árbol* y *pasto* ocupan las tres primeras posiciones en ambos grados.
- **La ciudad:** el grado de coincidencia en los vocablos de este CI es del 70%. Es uno de los cuatro centros con menor porcentaje de identidad estadística en sus UL. En quinto el vocablo más disponible es *casa*, mientras que en once es *carro*. No obstante, los tres primeros vocablos en las dos listas son similares: *carro*, *casa* y *edificio*. Se destaca el descenso entre quinto y once de las UL *persona*, *carretera*, *restaurante*, *almacén*, *supermercado* y *poste* a lugares por debajo de la posición 20, frente a *avenida*, *gente*, *transmilenio*, *taxi*, *museo* y *universidad* en once.
- **Profesiones y oficios:** junto al anterior CI, éste también posee 70% de identidad entre las respuestas de las dos poblaciones. Sin embargo, a diferencia del anterior CI, sí coincide el vocablo más disponible: *profesor*.
- **Juegos y distracciones:** éste es uno de los centros de interés con menor igualdad entre las 20 primeras UL en quinto y once: 65%. A pesar de este bajo porcentaje, el vocablo *fútbol* es el de mayor IDL en ambos grados. En quinto se presentan UL referidas a juegos y artefactos de naturaleza

¹⁴ Para esto, los CI se agruparon según superaran el promedio del porcentaje de semejanza en los veinte primeros vocablos de los 16 centros de interés, el cual corresponde al 80,3%.

más infantil, como *cogidas*, *congelados*, *rodadero*, *stop* y *columpio*, que junto a *montaña rusa* fueron remplazados por actividades más adultas en once, como *cartas*, *billar* y *ping pong*. Además, se resalta que aparecen unidades léxicas que devienen de voces extranjeras en las dos poblaciones, como *basketball*. También sorprende que en esta selección los estudiantes de once tengan más disponible un vocablo como *golosa*.

- **Trabajos del campo y el jardín:** finalmente, éste es el centro de interés con menor porcentaje de identidad en las respuestas de los estudiantes de ambos grados: 60%. No obstante este porcentaje, las tres primeras UL son idénticas, *sembrar*, *cultivar* y *ordeñar*, en las dos selecciones por grado.

Conclusiones

Como quedó señalado en el acápite 2 de este documento, el objetivo fundamental del estudio sobre la disponibilidad léxica en estudiantes bogotanos apuntaba a realizar el diagnóstico del léxico disponible de esta sintopía; objetivo que se ha cumplido a cabalidad en la medida en que se han obtenido los listados de las unidades léxicas, discriminadas por centros de interés. Estos listados corresponden a los diccionarios del léxico disponible de los estudiantes bogotanos. En dichos listados se recogen las unidades léxicas que los estudiantes actualizaron y verbalizaron ante los 16 centros de interés definidos.

Con base en el análisis informático de dichos datos, pudo establecerse el número total de unidades léxicas, o productividad léxica, aportadas por los 911 informantes; número que, como se señaló, ascendió a 228.012. Al estudiar esta cifra, se estableció esta gradación en la distribución de los 16 centros de interés, de acuerdo con el número de UL que presentaron: *Animales*, *Alimentos y bebidas*, *Partes del cuerpo*, *La escuela/colegio y útiles escolares*, *La ciudad*, *El campo*, *Prendas de vestir*, *Profesiones y oficios*, *Juegos y distracciones*, *La cocina*, *Medios de transporte*, *Partes de la casa*, *Objetos puestos en la mesa para la comida*, *Muebles de la casa*, *Trabajos del campo y el jardín* e *Instrumentos para iluminar*,

airear o calentar un recinto. De éstos, los ocho primeros se encuentran por encima del promedio de unidades léxicas, mientras que los restantes ocho se encuentran por debajo de esta media.

De igual forma, el tratamiento informático de los datos permitió establecer el número de unidades léxicas diferentes, o riqueza léxica (14.251). Al revisar los centros de interés según este criterio, se obtuvo que el orden de éstos se altera de la siguiente forma: *Juegos y distracciones*, *Profesiones y oficios*, *Trabajos del campo y el jardín*, *La ciudad*, *El campo*, *La escuela/colegio y útiles escolares*, *Alimentos y bebidas*, *La cocina*, *Instrumentos para iluminar*, *airear o calentar un recinto*, *Muebles de la casa*, *Animales*, *Objetos puestos en la mesa para la comida*, *Prendas de vestir*, *Partes de la casa*, *Partes del cuerpo* y *Medios de transporte*. Asimismo, en lo que respecta al promedio de unidades léxicas diferentes por centro, el listado sufre cambios, de tal forma que sólo los seis primeros centros se encuentran por encima de éste.

Lo anterior permitió establecer que no existe una relación proporcional entre el número de unidades léxicas, esto es, la productividad léxica, y el número de unidades léxicas diferentes, riqueza léxica, de los centros de interés; así, el hecho de que un centro presente el mayor número de unidades léxicas y, en consecuencia, ocupe los primeros puestos del orden según este criterio, no implica necesariamente que ofrezca el mayor número de unidades léxicas diferentes. Tal es el caso, por ejemplo, del centro *Animales*, que en el listado de productividad léxica muestra el mayor número de unidades léxicas (22.402), pero que en el listado de riqueza léxica pasa a ocupar el puesto 11 (sólo 684 ULD).

Esto tiene que ver, con el tipo de centro de interés, es decir, si éste tiende a ser compacto o difuso. De esta forma, el índice de cohesión léxica muestra que cuanto más elevado sea este índice, el número de unidades léxicas diferentes se reduce y que mientras este índice disminuye, el número de estas unidades aumenta. Este fenómeno podría explicarse por el hecho de que las posibilidades de asociación semántica o de elección léxica, por ejemplo, son más limitadas que otras, según el estímulo temático. Centros como *Animales* o *Partes del cuerpo* presen-

tan mayor delimitación conceptual, de ahí que se dé mayor coincidencia en las respuestas de los informantes, mientras que *Juegos y distracciones* genera más asociaciones diversas: pasatiempos, deportes o espectáculos, razón por la cual se presentan menos coincidencias en las respuestas de los informantes y, por lo tanto, ofrece el mayor número de unidades léxicas diferentes, al punto que ocupa el primer puesto en la tabla de riqueza léxica (1.853 ULD). Según esto, "ciertos centros de interés alcanzan un alto grado de cohesión semántica, en el sentido que se les puede limitar a un número reducido de clases lógicas; y otros, en cambio, presentan un débil grado de coherencia semántica porque el número de clases lógicas que los componen es difícilmente delimitable" (Gómez y Gómez, 2004: 82).

Al hacer el análisis de los datos a partir de la variable edad, pudo establecerse que tanto a nivel de productividad léxica (UL) como de riqueza léxica (ULD) se dan diferencias de orden cuantitativo, esto es, los estudiantes de grado once ofrecen más unidades léxicas y unidades léxicas diferentes que los estudiantes de grado quinto. Hecho que puede atribuirse al mayor grado de desarrollo lingüístico y cognitivo de los primeros frente a los segundos.

No obstante estas diferencias, al cotejar los resultados de los dos grupos en cuanto al orden de los centros de interés, puede apreciarse que el grado de coincidencia es alto, a tal punto que sólo presentan una leve diferencia en el orden de aparición de dos centros, *Partes del cuerpo* y *La escuela/colegio y útiles escolares*. En lo que respecta a los centros de mayor y menor productividad, éstos se mantienen; así, *Animales* ocupa el primer puesto mientras que *Instrumentos para iluminar, airear o calentar un recinto* ocupa el último lugar en ambos grupos. Además, en cuanto a la riqueza léxica, debe resaltarse que las diferencias se hacen un poco más notorias, en la medida en que el orden de los centros cambia de 5° a 11°, aunque se mantienen los mismos, la gradación es distinta entre los dos grupos; sin embargo, tanto en 5° como en 11° *Juegos y distracciones* constituye el centro más rico en términos léxicos, mientras que *Medios de transporte* presenta el menor número de unidades léxicas.

Según lo anterior, los resultados de los análisis de los datos dejan entrever que a medida que aumenta la edad y, en consecuencia, el grado de madurez cognitiva, lingüística y cultural de los individuos, la capacidad para evocar y tomar decisiones de orden léxico mejora en términos cuantitativos a favor de los estudiantes del grado once, pero estas diferencias no son tan grandes en términos cualitativos en lo que a la composición de los centros de interés se refiere, coincidencias que están asociadas al hecho de que los informantes comparten, en principio, la misma realidad urbana. Aquí, cabe recordar que en este artículo no se ha hecho referencia a la incidencia de las otras variables extralingüísticas que se han considerado en el desarrollo del proyecto: el tipo de colegio, el nivel sociocultural y el género, las cuales ofrecen elementos para matizar y explicar las diferencias y similitudes encontradas en los análisis de los datos.

Asimismo, en lo relacionado con el índice de cohesión léxica, pudo establecerse que hay una gradación de los centros de interés que, en términos generales, coincide con los resultados de otros trabajos, en especial el adelantado con estudiantes preuniversitarios en Valencia. Dicha gradación permite clasificarlos en tres grandes categorías: compactos o cerrados, de coincidencia media y difusos o abiertos. Según esto, los centros *Animales, Alimentos y bebidas, y Partes del cuerpo humano* se ubican en la categoría de compactos, por cuanto presentan los índices de cohesión léxica más altos; los centros de interés *Partes de la casa, Objetos puestos en la mesa para la comida, La escuela/colegio y útiles escolares, La cocina, Muebles de la casa, La ciudad y El campo* tienden a ser de coincidencia media; mientras que los centros *Instrumentos para iluminar, airear o calentar un recinto, Profesiones y oficios, Juegos y distracciones y Trabajos del campo y el jardín* se consideran como difusos puesto que presentan los menores índices de cohesión léxica, así como el mayor número de unidades léxicas diferentes.

Esto se debe a que la realidad con la cual está asociado el centro de interés determina en el proceso de acceso y decisión léxica; esto es, el individuo tendrá más o menos posibilidades de productividad

léxica, según el tipo de centro que se le presente, debido a que hay centros que hacen referencia a realidades más complejas y menos delimitadas conceptualmente que otras que son más delimitadas, lo cual tiene incidencia en el número de unidades léxicas que pueden conformarlas y que estén, por lo tanto, disponibles para los hablantes. Con base en esto, la investigación psicolingüística ha establecido que factores como la frecuencia, la imaginabilidad, la concreción y la abstracción, las asociaciones semánticas, la categoría gramatical de la palabra y los aspectos fonéticos, tienen incidencia en las tareas de decisión léxica (cfr. Reeves, Hirsh-Pasek y Golinkoff, 1999). Asimismo, hay que resaltar que el grado de conocimiento que el individuo tenga sobre el campo temático incidirá en su productividad y riqueza léxicas.

Al cruzar el índice de cohesión léxica en los dos grupos, nuevamente se presentan coincidencias, por cuanto ambos grados los centros con mayor índice concuerdan, sólo cambia el orden de éstos en los dos grupos. En cuanto a los centros con cohesión media, así como los difusos, también se presenta el mismo fenómeno, esto es, hay coincidencias en los dos grupos en cuanto al número de centros ubicados en cada una de estas dos categorías, pero con distribución distinta en cada grado; esto quiere decir que existen algunas diferencias en lo que se refiere al índice de cohesión léxica, lo que hace que en grado 5° el orden de los centros, según este parámetro, sea distinto al que presenta el grado 11°; sin embargo, prima el hecho de que los informantes presentan un nivel de coincidencia en sus respuestas en los distintos centros de interés. No obstante, la homogeneidad en los índices de cohesión léxica sufre alteraciones que se hacen evidentes en el ordenamiento distinto de los centros en el grado 11° frente al del grado 5°, debido al aumento del índice de cohesión léxica en los estudiantes de 11°, determinado por el mayor nivel de desarrollo de éstos.

En relación con el índice de disponibilidad léxica, el análisis de los datos permite determinar que en ocho centros de interés (*Partes del cuerpo, Prendas de vestir, Partes de la casa, Animales, Instrumentos para iluminar, airear o calentar un recinto, Muebles*

de la casa, Alimentos y bebidas, y Medios de transporte), se da un alto grado de concordancia (cercano al 90%) en las respuestas entre las dos edades, de tal forma que en las 20 primeras unidades léxicas que se han tomado como ilustración hay coincidencia en aproximadamente 17 vocablos, que se encuentran tanto en los resultados de grado 5° como en los de grado 11°, esto es, que en estos centros las unidades con mayor índice de disponibilidad son prácticamente las mismas. Asimismo, en los restantes ocho centros el grado de coincidencia se va reduciendo; así, en los centros *Objetos puestos en la mesa para la comida, La cocina, La escuela*, el porcentaje de coincidencia en las primeras 20 unidades léxicas se reduce al 80%; en los centros *La ciudad y Profesiones y oficios* se da un porcentaje de coincidencia del 70%; mientras que en *Juegos y distracciones* y en *Trabajos del campo y del jardín*, la coincidencia sólo se da en 65% y 60%, respectivamente.

Finalmente, la estabilidad conceptual que se ha identificado en las primeras unidades léxicas de cada centro de interés permite plantear que dichas unidades pueden asumirse como nucleares o, si se prefiere, como *prototípicas* de cada centro, en la medida en que los individuos los toman como los característicos o representativos de cada una de las categorías cognitivas en las que las han ubicado; por ello ocupan los primeros lugares de los listados con los índices de disponibilidad más elevados. Así, aunque con diferencias en el orden, debidas a que el índice de disponibilidad es distinto, los dos grupos presentan, en los primeros lugares de la lista de *Partes del cuerpo*, las mismas diez unidades léxicas: *ojo, nariz, cabeza, boca, pie, mano, pierna, brazo, dedo, corazón*, que pueden considerarse, *prototípicas* de esta categoría; esto quiere decir que la probabilidad de aparición de éstas, en una prueba de decisión léxica, será más elevada que *esternocleidomastoideo*. Desde luego, si el tema discursivo es este músculo, la probabilidad de actualización, o el índice de disponibilidad, de esta unidad léxica se tornará elevada, por cuanto el léxico disponible hace referencia al conjunto de unidades léxicas que el hablante posee potencialmente en su lexicón mental y que se actualiza sólo si la situación comunicativa así lo requiere. ■

Bibliografía

- AITCHISON, Jean (1987). *Words in the mind. An introduction to the mental lexicon*. Oxford: Basil Blackwell.
- BUTRÓN, Gloria (1991). Nuevos índices de disponibilidad léxica. En López M., H. (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico.
- CARCEDO G., Alberto (1998). Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística*, (Alfal), 10.
- CUENCA, Joseph y Hilferty, Joseph (1999). *Introducción a la lingüística cognitiva*. Barcelona: Ariel.
- ECHEVERRÍA, Max (1991). Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. En López M., H. (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico.
- GALOSO C., María Victoria (2002). *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Universidad de Salamanca.
- GÓMEZ M., José Ramón y Gómez D., María Begoña (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos. Estudio de estratificación sociolingüística*. Valencia: Universitat de Valencia.
- GOUGENHEIM, Georges, et al (1964). *L'élaboration du français fondamental (1er degree). Étude sur l'élaboration d'un vocabulaire et d'un grammaire de base*. Paris: Didier.
- LÓPEZ CH., Juan y Strassburger F., Carlos (1991). Un modelo para el cálculo del índice de disponibilidad léxica individual. En: López M., H. (ed.). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico.
- LÓPEZ M., Humberto (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
- _____ (ed.) (1991). *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico.
- _____ (1983). Lingüística estadística. En López M., Humberto (coord.). *Introducción a la lingüística actual*. Madrid: Playor.
- MICHÉA, René (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les langues modernes*, 47.
- MORENO F., Francisco (1998). *Principios de sociolingüística y sociología de lenguaje*. Barcelona: Ariel.
- MULLER, Charles (1973). *Estadística lingüística*. Madrid: Gredos.
- REEVES, Laretta, Hirsh-Pasek, Kathy y Golinkoff, Roberta (1999). Palabras y significado. De los elementos simples a la organización compleja. En Berko G., Jean y Berstein R., Nan. *Psicolingüística*. Madrid: McGraw-Hill.
- RODRÍGUEZ T., Nelson (1980). Listas de frecuencias de palabras: una revisión de la literatura en español y de sus posibles usos en investigación. *Lectura y vida*, 4.
- SAMPER, José Antonio, Bellón, Juan José y Samper, Martha (2003). *El proyecto de estudio de la disponibilidad léxica en español*. (En prensa).