

## DISPONIBILIDAD LEXICA. MUESTREO Y ESTADISTICOS<sup>1</sup>

Alba Valencia

Universidad de Chile

### Resumen

Considerando la importancia que tiene el léxico para el crecimiento cognitivo del ser humano, se realizó una investigación para conocer cuánto y cuál es el vocabulario de los estudiantes chilenos del último curso de Enseñanza Media, con la finalidad de colaborar a la planificación escolar en este aspecto. En este artículo se presentan los pasos metodológicos para la recolección del corpus: la muestra de carácter nacional con parámetros pioneros en este tipo de estudios, la encuesta, los estadísticos resultantes y las conclusiones preliminares.

### Abstract

(Considering the importance that the lexicon bears in the cognitive development of the human being, a piece of research was carried out to find out how much and which is the vocabulary that Chilean students of the last grade in Secondary School have so as to help in the school planning in this respect.

This article presents the methodological steps followed in order to gather the corpus: the sample which has a national character and contains pioneering parameters in studies of this type, the questionnaire, the resulting statistics and the preliminary conclusions).

## ANTECEDENTES

Existe consenso respecto a la importancia del léxico en el desarrollo cognitivo del ser humano. Es al ampliar su vocabulario que el niño extiende también su dominio sobre el entorno y, paulatinamente, va ensanchando su horizonte vivencial. El camino se conoce a grandes rasgos: en primer término nomina la realidad concreta circundante y, luego, avanza hacia lo abstracto hasta llegar a apropiarse de todo el patrimonio que la humanidad ha construido. La cantidad y la calidad del léxico hacen la diferencia al momento de evaluar la competencia lingüística y cultural del individuo. Wittgenstein dijo “los límites de

---

<sup>1</sup> Este artículo es derivación del Proyecto FONDECYT N° 1940533.

mi lenguaje son los límites de mi mundo". Y el mundo se codifica, principalmente, en unidades léxicas.

Las investigaciones sobre léxico disponible siguieron a las de léxico básico o fundamental y tuvieron su comienzo en 1953, con el trabajo de René Michea, "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage". Los estudios sobre la disponibilidad léxica en Hispanoamérica se inician con los trabajos de López Morales en Puerto Rico, en 1973, y se continúan con gran productividad: Juan López Chávez, en México; Pedro Benítez, José Antonio Samper y Clara Hernández, en España; Orlando Alba, en República Dominicana; Max Echeverría y Alba Valencia, en Chile. Además, hay numerosas tesis de posgrado realizadas en Puerto Rico, México y Chile.

El índice de disponibilidad léxica (IDL) es un indicador de cuáles son las palabras que un grupo de sujetos tiene disponibles para la intercomunicación lingüística en un tema determinado. Para calcular este índice hay que considerar la frecuencia de cada término en el grupo en estudio, el total de sujetos encuestados, y la posición en que aparece la palabra en la lista dada por cada uno de ellos. 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.

El proyecto "Disponibilidad léxica en estudiantes de cuarto año de Educación Media" aplica este procedimiento estadístico para a) establecer un corpus nacional medido y evaluado cuantitativa y cualitativamente, como base concreta de referencia para la construcción de instrumentos de medición que se apoyen en el lenguaje verbal; b) elaborar un catastro léxico de acuerdo al índice de disponibilidad léxica para ser utilizado en planificación educacional; y c) comparar la disponibilidad léxica chilena con la de otros países de habla hispana en el mismo nivel de escolaridad.

A continuación, entrego datos pormenorizados de los criterios de muestreo y del análisis cuantitativo del léxico de los estudiantes del curso terminal 1994 de los liceos chilenos.

## 1. LA MUESTRA

Se obtuvo mediante una encuesta respondida por 2.052 sujetos, alumnos de establecimientos de Educación Media de tipo científico-humanista-diurno, estudiantes de los liceos que se anotan a continuación indicando la Región Administrativa<sup>2</sup> y la ciudad a la que pertenecen.

<sup>2</sup> Desde 1976, son 13 Regiones: I Región de Tarapacá, II Región de Antofagasta, III Región de Atacama, IV Región de Coquimbo, V Región de Valparaíso, Región Metropolitana, VI Región del Libertador General Bernardo O'Higgins, VII Región del Maule, VIII Región del Bío-Bío, IX Región de la Araucanía, X Región de Los Lagos, XI Región Aysén del General Carlos Ibáñez del Campo, XII Región de Magallanes y de la Antártica Chilena.

Tabla N° 1  
Establecimientos por Regiones

I	<b>Iquique:</b> Liceo A 11 Elena Duvauchelle; Liceo María Auxiliadora; Colegio Don Bosco <b>Pozo Almonte:</b> Liceo Alcalde Sergio González
II	<b>Antofagasta:</b> Liceo José Santos Ossa; Colegio Inglés San José <b>Taltal:</b> Liceo Juan Cortés Monroy
III	<b>Copiapó:</b> Liceo Mercedes Fritis Mackenney; Liceo Católico de Atacama <b>Freirina:</b> Liceo Gral. Ramón Freire
IV	<b>La Serena:</b> Liceo Gabriela Mistral; Liceo Gregorio Cordovez; Liceo Gerónimo Rendic <b>Monte Patria:</b> Presidente Eduardo Frei Montalva
V	<b>Valparaíso:</b> Liceo Eduardo de la Barra; Colegio Sagrados Corazones <b>Viña del Mar:</b> Liceo Guillermo Rivera; Seminario San Rafael <b>Catemu:</b> Liceo Chagres
REG	<b>Santiago:</b> Liceo Victorino Lastarria; Instituto Nacional; Liceo Javiera Carrera; Liceo José Toribio Medina; Liceo Carmela Carvajal de Prat; Liceo Guillermo Labarca; Villa María Academy; Colegio del Verbo Divino; Colegio Santiago Bueras; Colegio Filipense; Colegio Metropolitano; Liceo San Ignacio; Liceo Ignacio Carrera Pinto
ME	<b>Lampa:</b> Complejo Educativo Manuel Plaza
TRO	<b>Colina:</b> Complejo Educativo Polivalente Rigoberto Fontt
PO	<b>Lo Barnechea:</b> Colegio Parroquial San Paulo VI
LI	<b>San José de Maipo:</b> Liceo San José de Maipo
TA	<b>Buin:</b> Liceo A 131
NA	<b>Paine:</b> Centro Educacional Municipal Enrique Bernstein <b>Curacaví:</b> Liceo Presidente Balmaceda <b>Melipilla:</b> Liceo Hermanos Sotomayor Baeza
VI	<b>Rancagua:</b> Liceo Oscar Castro; Instituto Inglés <b>Peumo:</b> Liceo Jean Buchanan
VII	<b>Talca:</b> Liceo Abate Molina; Liceo Integrado San Pío X <b>Pencahue:</b> Complejo Educativo
VIII	<b>Concepción:</b> Liceo Enrique Molina Garmendia; Liceo de Niñas; Colegio Concepción; Colegio Sagrados Corazones <b>Hualqui:</b> Liceo San Juan Bautista
IX	<b>Temuco:</b> Liceo Pablo Neruda; Colegio de la Salle <b>Pitrufquén:</b> Liceo de Ciencias y Humanidades
X	<b>Puerto Montt:</b> Liceo Andrés Bello; Colegio Santo Tomás de Aquino <b>Castro:</b> Liceo Galvarino Riveros
XI	<b>Coyhaique:</b> Liceo Josefina A. Montenegro; Liceo San Felipe Benicio <b>Chile Chico:</b> Liceo Rabanal Palma
XII	<b>Punta Arenas:</b> Liceo Juan Bautista Contardi; The British School <b>Puerto Porvenir:</b> Liceo Hernando de Magallanes

### 1.1. Factores de estratificación de la muestra

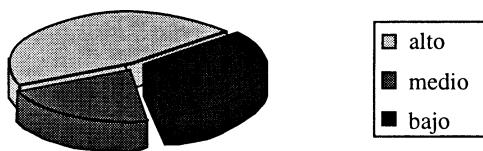
La selección de la muestra se hizo considerando los factores extralingüísticos usuales en estos estudios: sexo y nivel sociocultural; además, se incluyeron el régimen educacional, la regionalidad y la procedencia poblacional, para averiguar la gravitación que ellos tienen sobre el bagaje léxico de los informantes.

**1.1.1.** Se procuró que hubiese un equilibrio proporcional en la variable sexo: 1032 estudiantes de sexo masculino (50,3%) y 1020, de sexo femenino (49,7%).

**1.1.2.** Para la determinación del nivel sociocultural (NSC), se aplicó la encuesta elaborada por Himmel y Maltes que considera: a) *nivel educacional de los padres*, en una escala de 0 punto (sin estudios) a 7 puntos (estudios universitarios completos); b) *nivel ocupacional*: de 0 punto (obreros no especializados) a 7 puntos (grandes empresarios, grandes directivos), y c) *enumeración de elementos del hogar*: artefactos tales como cálifon, plancha eléctrica, máquina de coser o tejer, aspiradora, radio, refrigerador, televisor, automóvil, etc. En a) y b) se codifica sólo el nivel más alto alcanzado por cualquiera de los padres; en c) se asigna a cada elemento un puntaje de 0,5 o 1. La suma final ubica al sujeto en una de estas categorías: si alcanza 16,5 puntos o menos, en el nivel bajo; si tiene entre 17 y 23,5 puntos, en el nivel medio; si suma 24 puntos o más, en el nivel alto.

De acuerdo con la tabulación de esta encuesta sociocultural, 954 estudiantes (el 46,49% de los sujetos) pertenecen al nivel medio; 670 (el 32,65%), al nivel bajo, y 428 (el 20,86%), al nivel alto, situación que se aprecia en el gráfico.

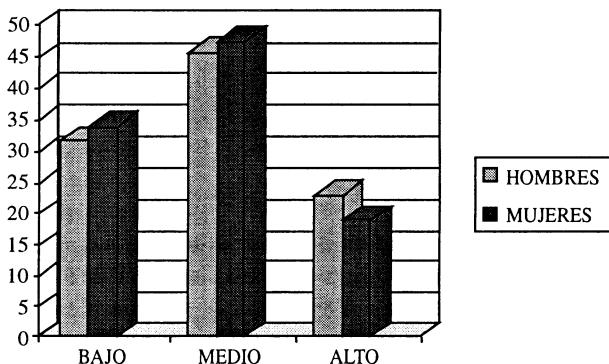
Gráfico N° 1  
Distribución por nivel sociocultural



**1.1.3.** Si se correlacionan las variables sexo y NSC, tenemos que en el nivel bajo se encuentran 325 varones y 345 niñas; en el nivel

medio, 472 varones y 482 niñas, y en el nivel alto, 235 varones y 193 niñas. Gráficamente, el universo estudiado se distribuye así, según estas dos variables:

Gráfico N° 2  
Distribución por sexo/NSC

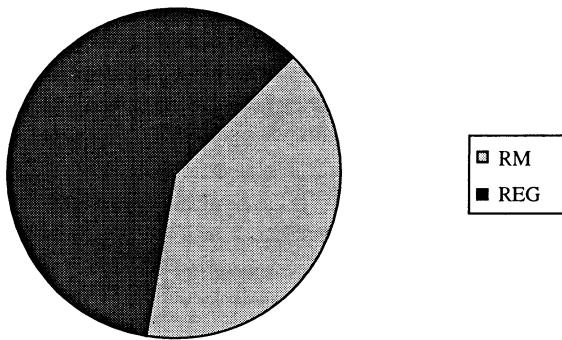


**1.1.4.** La variable régimen educacional considera establecimientos públicos y privados.<sup>3</sup> En la muestra están representados 63 establecimientos educacionales, de los cuales 39 son públicos y 23, privados. La población escolar de los liceos públicos asciende a 1198 y la de los particulares, a 854.

**1.1.5.** Hemos llamado convencionalmente “regionalidad” al factor geográfico, asociándolo a las Regiones Administrativas en que se divide el país. Se consideran en la muestra establecimientos de las trece Regiones, como se puede comprobar en la Tabla N° 1. De acuerdo con las estadísticas actuales, el 40% de la población de IV año medio estudia en la Región Metropolitana, por lo que se tuvo en cuenta esos datos al momento de elaborar la muestra. En consecuencia, se encuestaron 824 estudiantes en la Región Metropolitana (40,2% de la muestra), y 1228 estudiantes (59,8%), en el resto del país.

<sup>3</sup> No se distingue aquí entre particular pagado y subvencionado, puesto que en algunas Regiones Administrativas del país no existe la primera de estas categorías.

Gráfico N° 3  
Distribución Región Metropolitana vs Regiones



**1.1.6.** La procedencia poblacional de los sujetos fue determinada por la distinción entre urbano y rural. Se consideró **urbano** un establecimiento educacional inserto en plena ciudad, asumiendo que sus estudiantes también proceden de la urbe. Por oposición, se consideró **rural** un establecimiento cuya característica principal es su aislamiento relativo con respecto a los centros urbanos, en otras palabras, su condición de periférico. Subyace el supuesto de que los muchachos del curso terminal de los liceos rurales no están expuestos –o lo están en menor grado– a los múltiples estímulos de las grandes ciudades, y eso influye de algún modo en su léxico.

En esta investigación no se operó con la distinción “urbano/rural” del Instituto Nacional de Estadísticas, porque existen muchas localidades que, siendo “urbanas” según la definición censal de 1982 de dicho organismo, presentan características que las sitúan dentro de alguno de los criterios válidos para nuestra clasificación y, por lo tanto, pudieron haber sido elegidas.

Para la determinación de las localidades seleccionables según este factor, se recurrió a la asesoría de un experto en demografía, quien elaboró una nómina utilizando los siguientes criterios:

- a) el tamaño demográfico de la localidad con respecto a las restantes de su grupo<sup>4</sup>;
- b) la distancia en Kms. del centro urbano principal más cercano, lo que permite identificar el aislamiento relativo de la localidad;
- c) el número mínimo de alumnos cursando IV medio requerido por el proyecto.

---

<sup>4</sup> Se trabajó con la información del Censo 1982, porque no estaba disponible todavía la del Censo 1992

En consecuencia, en cada una de las 13 Regiones se procedió a encuestar estudiantes de liceos de la capital regional y de una localidad estimada de “ruralidad relativa”. De acuerdo a estos parámetros, el grupo de estudiantes de procedencia poblacional urbana está constituido por 1622 sujetos y el de procedencia rural, por 430.

## 2. LA ENCUESTA

Al contingente de 2052 estudiantes se le aplicó una encuesta para recopilar el léxico de que disponían en relación con 18 centros de interés, en cuya elección se trató de abarcar un amplio espectro de los núcleos de conocimiento y actividad de los adolescentes.

### 2.1. Los centros encuestados fueron:

Las partes del cuerpo	(CUE)	Juegos y entretenimientos	(JUE)
Ropa y calzado	(RYC)	Profesiones y oficios	(PRO)
La casa	(CAS)	La política	(POL)
Muebles	(MUE)	La actividad económica	(ECO)
Alimentos	(ALI)	Las artes	(ART)
La cocina	(COC)	Ciencia y tecnología	(CYT)
Medios de transporte	(TRA)	Mundo espiritual	(MUN)
Trabajo de campo y jardín	(CAM)	Procesos mentales	(PRO)
Plantas y animales	(PLA)	Problemas del ambiente	(AMB)

Después de exemplificar el procedimiento para responder la encuesta, se leyeron las instrucciones –un centro por vez–, dando 2 minutos de tiempo para que los muchachos anotaran las palabras, en un cuadernillo debidamente diagramado y de fácil manipulación.

- 2.2. El volumen de palabras recogidas hubo que someterlo a estandarización, a fin de homogeneizar los datos teniendo cuidado de respetar, en lo posible, la versión original.
- a) Se eliminaron las palabras que aparecían repetidas o no correspondían al centro. Considerando que se pide al sujeto una relación asociativa de palabras, hubo que estudiar cuidadosamente en algunos casos –muy pocos, en definitiva–, la eliminación. En general, la aparición de un término que no pertenecía al campo fue bastante evidente.
  - b) Se revisó la ortografía para, en algunos casos, regularizar variantes gráficas y, en muchos casos, corregir errores. Como ejemplos “para la historia”, anotaré –sin comentarios– algunas ortografías relativamente recurrentes como *helefante* por “ele-

fante”, *huelles* o *guelles* por “bueyes”, *llena* por “hiena”, *hawar* por “jaguar”, *gayina* por “gallina”, *narango* por “naranjo”, *holla* por “olla”, *descrasia* por “desgracia”, *ignotismo* por “hipnotismo”, *asinamiento* por “hacinamiento”, *absila* por “axila”, *disahue* por “desagüe”. Frente a estos casos, resultan hasta disculpables otros como *liben* por “living”, *jaguallanas* o *jaguallaras* por “hawaianas”, *Bancoc* por “Van Gogh”, *Mosar* por “Mozart”... y otros.

- c) Para los casos de variantes morfológicas, entradas léxicas compuestas y otros casos específicos, se tomaron las siguientes decisiones: anotar en singular sustantivos y adjetivos, privilegiando la forma masculina cuando la variación de género no implicaba cambio de valor léxico; adscribir los diminutivos a la forma primitiva; adscribir todas las formas verbales al infinitivo verbal; gerundio y participio como entradas independientes, este último, en su forma masculina; las lexías unidas por guiones: “rayito-de-sol”, “ingeniero-en-computación”, “zapato-de-fútbol”, “mesa-de-centro”; escribir los extranjerismos con grafía españolizada, como aparecen en el diccionario de la RAE; los que no están españolizados, como en la lengua de origen. Otra opción fue la de digitar todo en minúsculas, incluso los nombres propios.

Huelga decir que ésta es una de las etapas más laboriosas de la investigación y la que consume la mayor cantidad tiempo, pero de su acuciosidad depende, en gran parte, la confiabilidad de los resultados.

### 3. EL PROCESAMIENTO DE LOS DATOS

El material, preparado como se ha descrito, se sometió a proceso computacional para el cálculo de índices y frecuencias. El índice de disponibilidad léxica (IDL), que considera la frecuencia de menciones de cada palabra en un grupo de sujetos en relación con el número de éstos, pues la probabilidad de ocurrencia de una palabra es función del número de ellos. Además, es importante determinar el lugar en que aparece la palabra en cuestión, porque es evidente que un término que acude a la mente de inmediato está disponible en un grado altamente mayor que otro que se actualiza con dificultad. Para el cálculo del IDL se utilizó la fórmula de Echeverría (1987: 67), artículo al que remitimos al lector. Otros datos son el promedio de respuestas, las palabras diferentes y el índice de cohesión. Este último fue acuñado por Echeverría y establece el grado de homogeneidad del léxico grupal.

#### 4. EL ANÁLISIS ESTADÍSTICO

A continuación se entrega el análisis, desde el punto de vista cuantitativo, de los datos obtenidos: en primer lugar, los datos generales y, luego, los obtenidos al considerarlos desde la perspectiva de las variables independientes.

##### 4.1. Datos generales

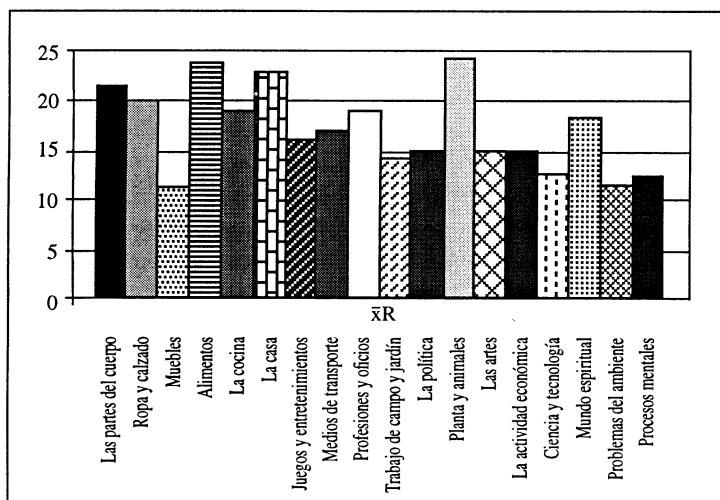
Como se sabe, los índices que más aportan a la determinación de la riqueza léxica de los sujetos son el *promedio de respuestas* a cada centro de interés ( $\bar{x}R$ ), el total de *palabras diferentes* por centro (PD), y el *índice de cohesión* (IC). En la tabla siguiente se anotan los resultados obtenidos por los 2052 sujetos en estos tres índices.

Tabla N° 2  
Datos generales de indicadores por Centro de Interés

Centro de interés	$\bar{x}R$	PD	IC
Las partes del cuerpo	21,39	483	0,0440
Ropa y calzado	20,30	481	0,0420
Muebles	11,06	314	0,0350
Alimentos	23,84	819	0,0290
La cocina	19,13	871	0,0210
La casa	22,86	927	0,0246
Juegos y entretenimientos	16,02	1433	0,0111
Medios de transporte	16,84	470	0,0358
Profesiones y oficios	18,97	1353	0,0140
Trabajo de campo y jardín	14,14	1176	0,0120
La política	15,02	1640	0,0091
Plantas y animales	24,20	895	0,0270
Las artes	14,80	1696	0,0087
La actividad económica	14,92	1749	0,0085
Ciencia y tecnología	12,42	1578	0,0078
Mundo espiritual	18,15	2096	0,0086
Problemas del ambiente	11,14	1251	0,0089
Procesos mentales	12,29	1894	0,0064

**4.1.1.** El  $\bar{x}R$  indica cuántos son los vocablos que, en promedio, poseen los sujetos para su comunicación en el ámbito de conocimiento que explora el centro de interés respectivo. Los resultados obtenidos con esta muestra se pueden apreciar en el gráfico.

Gráfico N° 4  
Promedio de respuestas por Centro de Interés



De acuerdo a este índice, se advierte que la mayor riqueza léxica está en el centro de interés “Plantas y animales”; en segundo lugar está el centro “Alimentos”, y en tercer lugar, “La casa”. Esto significa que es en estos tres ámbitos donde se concentra el mayor número de vocablos disponibles para los sujetos de la muestra. Los  $\bar{x}R$  más deficitarios son “Muebles”, “Problemas del ambiente”, “Procesos mentales” y “Ciencia y Tecnología”.

El total de palabras que, en promedio, manejan los estudiantes de 4º medios en relación con los 18 campos nacionales encuestados, es de 307. El cálculo del promedio ponderado indica que éste es de 17 palabras por campo, cifra ligeramente superior al promedio ponderado de 15 palabras obtenido por Valencia (1994) en la encuesta aplicada en 1990 a una muestra estratificada de 298 estudiantes de 4º medio.

**4.1.2.** El índice PD da cuenta del total de vocablos conocidos por el grupo muestral. En la Tabla N° 2, se puede apreciar cómo se distribuye ese conocimiento lingüístico en los distintos núcleos temáticos encuestados. El mayor número de PD se encuentra en el centro de interés “Mundo espiritual”, con 2096 vocablos. Por lo tanto, es el ámbito que más conocen, en conjunto, los sujetos de la muestra. Sin embargo, el índice no da garantía de que todos los sujetos utilicen efectivamente este variado vocabulario. La riqueza léxica es menor

en los campos que le siguen: “Procesos mentales” (1894), y “La actividad económica” (1.749). Los centros con las menores cantidades de vocablos son “Muebles”, donde se recogen 314 PD, y “Medios de transporte”, cuyo listado es de 470 palabras. Comparados con los centros que presentan mayor número de PD, éstos apuntan a campos nacionales de carácter más concreto.

**4.1.3.** El IC es un indicador del grado de coincidencia en las respuestas. Esto permite interpretar más adecuadamente los datos del índice anterior, pues el grado de homogeneidad en el uso del léxico está en proporción directa con del grado de disponibilidad que ese léxico tiene para todos los sujetos como conjunto (v. Echeverría 1987).

El centro de interés que muestra el índice más alto es “Las partes del cuerpo”: 0,044, esto es, un área de conocimiento para la que todos los informantes tienen disponible aproximadamente el mismo léxico: un total de 483 palabras diferentes.

Tabla Nº 3  
Valor de IC por Centros de Interés

IC	CENTROS DE INTERÉS
0,04	CUE, RYC, MUE, TRA
0,03	ALI, PLA, CAS
0,02	COC
0,01	PYO, CAM, JUE, POL, AMB, ART, MUN, ECO, CYT, PRO

Como se puede apreciar, los centros en los que el léxico disponible es más homogéneo son sólo cuatro, y justamente se trata de áreas temáticas con referentes bien concretos y muy cercanos a la experiencia vital de los estudiantes. Como estos cuatro centros fueron los que acusaron el menor número de palabras diferentes, es lícito pensar que son campos muy acotados, pues los sujetos tienen respecto de ellos una experiencia común y, en consecuencia, hay mayor coincidencia en la mención de sus referentes.

En la base del cuadro se ubican centros que presentan un vocabulario muy disperso. Correlativamente, aquí se encuentran los centros con gran número de palabras diferentes. Si se observa, se constatará que estos centros, en la mayoría de los casos, implican conocimientos específicos dados por la escolaridad y, en los menos, una exposición a experiencias muy particulares como en “Juegos y entretenimientos” y “Trabajos de campo y jardín”.

#### 4.2. La variable sexo de los informantes

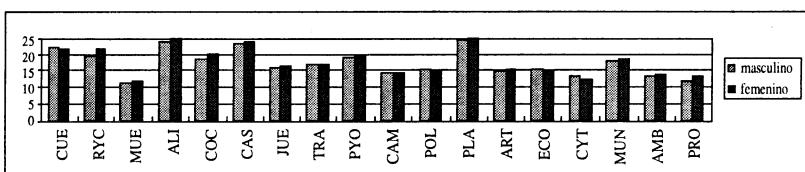
Como se dijo, se encuestaron 1032 hombres y 1020 mujeres. Los resultados obtenidos en los tres índices considerados, se presentan a continuación:

Tabla N° 4  
Cuantificación de índices según variable sexo

CENTRO	SEXO	$\bar{x}R$	PD	IC
CUE	MASC	21.89	454	0.04
	FEM	20.96	400	0.05
RYC	MASC	19.41	441	0.04
	FEM	21.34	422	0.05
MUE	MASC	10.84	313	0.03
	FEM	11.46	273	0.04
ALI	MASC	23.20	758	0.03
	FEM	24.54	690	0.03
CO <sub>C</sub>	MASC	18.47	797	0.02
	FEM	20.00	784	0.02
CAS	MASC	22.69	851	0.02
	FEM	23.18	798	0.02
JUE	MASC	15.80	1230	0.01
	FEM	16.25	1214	0.01
TRA	MASC	17.00	518	0.03
	FEM	16.83	353	0.04
PYO	MASC	18.68	1209	0.01
	FEM	19.26	1151	0.01
CAM	MASC	14.26	1104	0.01
	FEM	14.15	961	0.01
POL	MASC	15.22	1504	0.01
	FEM	14.81	1323	0.01
PLA	MASC	24.15	827	0.02
	FEM	24.39	771	0.03
ART	MASC	14.51	1513	0.009
	FEM	15.10	1395	0.01
ECO	MASC	15.09	1509	0.01
	FEM	14.38	1414	0.01
CYT	MASC	12.92	1484	0.008
	FEM	11.93	1237	0.009
MUN	MASC	17.86	1784	0.01
	FEM	18.44	1671	0.01
AMB	MASC	13.38	1150	0.01
	FEM	13.44	1133	0.01
PRO	MASC	11.31	1587	0.007
	FEM	13.28	1604	0.008

**4.2.1.** El más alto  $\bar{x}R$  lo alcanzan las mujeres en el centro de interés “Alimentos” (24.54), y los otros altos promedios se dan en el centro de interés “Plantas y animales”, en el que las mujeres presentan 24.39 y los hombres 24.15. El menor promedio lo presentan los hombres en el centro “Muebles”: 10.84, aunque las mujeres sólo alcanzan un promedio ligeramente superior en este campo: 11.46, como muestra el gráfico.

Gráfico N° 5  
Promedio de respuestas por sexo



En general, se advierte que las niñas tienen mejores promedios; aventajan a los hombres en 12 de los 18 centros encuestados. En algunos casos, y de acuerdo a la perspectiva que se adopte, ese mayor conocimiento podría estar motivado por una inclinación natural, o por la aceptación pasiva de un rol social, como, por ejemplo, en los centros ALI, COC, CAS, RYC, MUE.

Las niñas manejan, en promedio, 314 palabras para el total de los centros encuestados; los hombres, 307. La diferencia, pues, no es significativa, ya que el promedio por centro, en ambos grupos, es de 17 palabras. En el estudio de Valencia 1994 tampoco se encontraron diferencias significativas en esta variable, pero en esa muestra los hombres presentaban un total de PD tres puntos mayor que el de las niñas.

**4.2.2.** Si se considera el número de PD que manejan los sujetos de la muestra, la situación es la siguiente: el mayor número de vocablos lo presentan tanto hombres como mujeres en el centro MUN: 1784 y 1671, respectivamente. El menor número de PD está en el centro MUE: mujeres 273 y hombres 313. Los varones presentan una mayor riqueza léxica en todos los campos. Las diferencias van de 13 palabras más en COC, a 181 en POL. Las niñas aventajan solamente en el centro PRO, con una diferencia de 17 vocablos.

**4.2.3.** Los siguientes son los resultados de la aplicación del IC

Tabla N° 5  
Indice de cohesión y Centros de interés

IC	CENTROS DE INTERÉS	
	MASCULINO	FEMENINO
0,05		CUE, RYC
0,04	CUE, RYC	MUE, TRA
0,03	MUE, ALI, TRA	ALI, PLA
0,02	COC, CAS, PLA	COC, CAS
0,01	JUE, PYO, CAM, POL, ECO, MUN, AMB	JUE, PYO, CAM POL, ART, ECO MUN, AMB
0,009	ART	CYT
0,008	CYT	PRO
0,007	PRO	

Como se puede apreciar, las mujeres presentan un vocabulario más homogéneo que el de los hombres. Tal vez eso explique por qué la lista de PD que conocen las niñas en cada centro de interés es más reducida que la de los varones. Estos, como grupo, tienen más léxico, pero es un léxico más disperso que el de las niñas. El campo nocional más cohesionado es “Las partes del cuerpo” en ambos grupos, pero alcanza un mejor IC en las niñas. Los centros en los que no se comprueban vivencias comunes son, aunque con distinto IC, pero igualmente deficitarios en ambos grupos, CYT y PRO.

#### **4.3. La variable nivel sociocultural**

Al analizar los datos léxicos considerando los niveles socioeconómicos A(lto), M(edio) y B(ajo), se obtuvieron, en cuanto al promedio de respuestas, las palabras diferentes y el índice de cohesión de ellas, los resultados que consigna la Tabla N° 6. El grupo de NSC-B está constituido por 670 sujetos; el de NSC-M, por 954 y el de NSC-A, por 428.

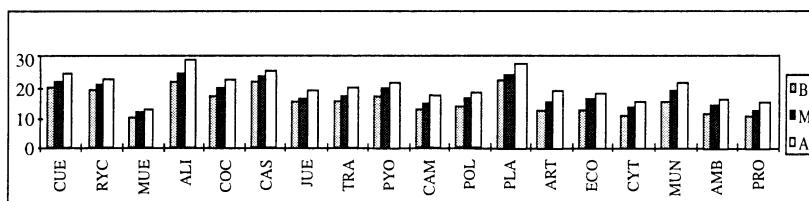
**Tabla N° 6**  
**Indicadores  $\bar{x}R$ , PD e IC y nivel sociocultural**

CENTRO	NSC	$\bar{x}R$	PD	IC
CUE	B	19.37	332	0.05
	M	21.55	432	0.04
	A	24.39	356	0.06
RYC	B	18.72	366	0.05
	M	20.45	414	0.04
	A	22.77	347	0.06
MUE	B	10.11	195	0.05
	M	11.40	296	0.03
	A	12.22	238	0.05
ALI	B	21.31	535	0.03
	M	23.87	706	0.03
	A	28.10	628	0.04
CO C	B	16.79	584	0.02
	M	19.42	728	0.02
	A	22.55	662	0.03
CAS	B	21.30	662	0.03
	M	22.96	784	0.02
	A	25.37	684	0.03
JUE	B	14.65	946	0.01
	M	16.27	1176	0.01
	A	18.43	972	0.01
TRA	B	15.27	307	0.04
	M	16.83	407	0.04
	A	19.59	401	0.04
PYO	B	17.09	930	0.01
	M	19.33	1202	0.01
	A	21.87	896	0.02
CAM	B	12.44	780	0.01
	M	14.51	979	0.01
	A	16.61	801	0.02
POL	B	12.74	973	0.01
	M	15.54	1444	0.01
	A	18.22	1115	0.01
PLA	B	22.49	680	0.03
	M	24.26	459	0.03
	A	27.07	687	0.03
ART	B	12.28	990	0.01
	M	15.38	1415	0.01
	A	18.29	1157	0.01

CENTRO	NSC	$\bar{x}R$	PD	IC
ECO	B	12.28	1086	0.01
	M	15.85	1432	0.01
	A	17.68	1034	0.01
CYT	B	10.53	958	0.01
	M	12.80	1309	0.009
	A	15.38	1113	0.01
MUN	B	15.26	1270	0.01
	M	18.89	1701	0.01
	A	21.67	1289	0.01
AMB	B	11.56	805	0.01
	M	13.64	1116	0.01
	A	15.81	903	0.01
PRO	B	10.63	1126	0.009
	M	12.64	1522	0.008
	A	15.06	1260	0.01

**4.3.1.** El grupo A alcanza los más altos promedios en todos los centros de interés propuestos, sigue el grupo M y, por último, el grupo B. Esto lleva a pensar que la riqueza léxica está fuertemente asociada a la pertenencia del sujeto a un estrato sociocultural. El  $\bar{x}R$  más alto alcanzado está en el centro “Alimentos” (28.10), el más bajo, en “Muebles” (10.11), como se aprecia en el gráfico.

Gráfico N° 6  
Promedio de respuestas por nivel sociocultural



Los mayores promedios del grupo A están en ALI y PLA; los del grupo M, en PLA y CAS; los del grupo B, en PLA y ALI. El centro menos rico en léxico para todos los niveles socioculturales es MUE.

El grupo A presenta en promedio, para los 18 centros, 361 palabras; el grupo M, 316 y el grupo B, 275. El total de palabras que, en promedio, maneja cada uno de los grupos estudiados confirma que el factor sociocultural influye en forma decisiva en la cantidad de vocabulario que el sujeto puede actualizar rápidamente como res-

puesta a un estímulo. Estos datos estadísticos señalan que, para un discurso en una de las áreas temáticas que exploramos, los sujetos del nivel B disponen de un promedio de 15 palabras, donde los del M disponen de 18 y los del A, de 20.

**4.3.2.** La mayor cantidad de PD se encuentra en el estrato M en el área temática “Mundo espiritual”, con 1701 vocablos. Este es también el centro en el que los otros dos estratos presentan la cumbre más alta: B con 1270 y A con 1289 vocablos, respectivamente. El centro de interés más deficitario para los tres estratos es “Muebles”, en el que no sobrepasan los 300 vocablos.

**4.3.3.** En la Tabla Nº 7 se puede visualizar cuáles son los centros más aglutinados para cada estrato sociocultural, de acuerdo con el valor de su índice de cohesión.

Tabla Nº 7  
Índice de cohesión y nivel sociocultural

IC	CENTROS DE INTERÉS		
	BAJO	MEDIO	ALTO
0,06			CUE, RYC
0,05	CUE, RYC, MUE		MUE
0,04	TRA	CUE, RYC, CAS, TRA	ALI, TRA
0,03	ALI, CAS, PLA	MUE, ALI, PLA	COC, CAS, PLA
0,02	CO	CO	PYO, CAM
0,01	JUE, PYO, CAM, POL, ART, ECO, CYT, MUN	JUE, PYO, CAM, POL, ART, ECOMUN, AMB	JUE, POL, ART, ECO, CYT, MUN, PRO
0,009	PRO	CYT	
0,008		PRO	

El grupo A es el que presenta más coincidencia en las respuestas. El grupo M se aprecia como el de vocabulario más disperso. Los centros con léxico más homogéneo son CUE, RYC y MUE, tanto para el estrato B como para el A, como se puede apreciar en la tabla.

#### **4.4. La variable régimen de estudios**

En la muestra hay 1198 sujetos que estudian en la enseñanza municipalizada, es decir, pertenecen al régimen de estudio público. Los

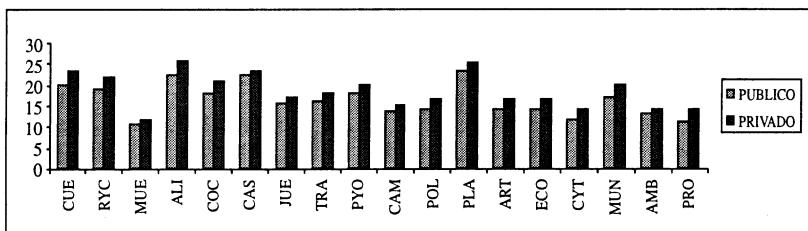
otros 854 sujetos estudian en establecimientos privados. Según esta variable, los índices fueron los que se consignan en la siguiente tabla.

Tabla N° 8  
Indicadores  $\bar{x}R$ , PD e IC y régimen de estudios

CENTRO INTERÉS	RÉGIMEN ESTUDIOS	$\bar{x}R$	PD	IC
CUE	Público	20.01	442	0.04
	Privado	23.41	422	0.05
RYC	Público	19.24	438	0.04
	Privado	21.95	427	0.05
MUE	Público	10.80	285	0.03
	Privado	11.63	304	0.03
ALI	Público	22.64	717	0.03
	Privado	25.71	727	0.03
COC	Público	17.97	766	0.02
	Privado	20.96	788	0.02
CAS	Público	22.41	801	0.02
	Privado	23.63	840	0.02
JUE	Público	15.59	1298	0.01
	Privado	17.05	1201	0.01
TRA	Público	15.83	419	0.03
	Privado	18.40	461	0.04
PYO	Público	18.36	1240	0.01
	Privado	20.20	1143	0.01
CAM	Público	13.58	1040	0.01
	Privado	15.24	1003	0.01
POL	Público	14.14	1404	0.01
	Privado	16.65	1429	0.01
PLA	Público	23.43	825	0.02
	Privado	25.44	788	0.03
ART	Público	13.89	1439	0.009
	Privado	16.51	1461	0.01
ECO	Público	14.02	1460	0.009
	Privado	16.55	1413	0.01
CYT	Público	11.69	1334	0.008
	Privado	13.88	1400	0.009
MUN	Público	16.95	1764	0.009
	Privado	20.16	1685	0.01
AMB	Público	12.93	1108	0.01
	Privado	14.07	1177	0.01
PRO	Público	11.49	1552	0.007
	Privado	13.89	1642	0.008

**4.4.1.** El gráfico muestra los datos correspondientes al  $\bar{x}R$  que los sujetos dieron a cada centro:

Gráfico N° 7  
Promedio de respuestas por régimen de estudios



Los estudiantes del segmento privado obtienen, en todos los centros de interés, promedios de respuestas más altos que los de los estudiantes del segmento público, es decir, tienen, en promedio, comparativamente, más palabras disponibles en cada área temática que los de establecimientos municipalizados. El más alto promedio alcanzado se encuentra en los establecimientos privados, en el centro de interés ALI (25,71) y el más bajo, en el de estudiantes de establecimientos públicos, en el centro MUE: 10,80.

Si se comparan los  $\bar{x}R$  dentro de cada estrato, en el grupo de educación privada el promedio más alto es el que ya se anotó: ALI 25,71; seguido de PLA 23,43. El más bajo es MUE 11,63. En el grupo de educación pública, el más alto es PLA 23,43; seguido de ALI 22,64. Los más bajos son MUE 10,80 y PRO 11,49.

**4.4.2.** En cuanto al número de PD que manejan, en grupo, los sujetos, se puede decir que el corpus mayor está en el centro “Mundo espiritual” del sector que recibe educación estatal, con 1764 palabras diferentes, y el menor, en el centro “Muebles” del mismo grupo. Si se interpretan estos datos según lo acotado sobre el grado de homogeneidad del vocabulario, se puede concluir que las PD del centro MUE son el reflejo verbalizado de la exposición de los sujetos a una experiencia común, en tanto que en el área de lo espiritual prima la experiencia individual.

En el centro CUE ambos grupos conocen el mismo número de vocablos, pero en los otros centros exhibe un mayor número el grupo de régimen privado. Solamente en 7 de los 17 centros restantes los estudiantes de régimen público aventajan a los de régimen particular.

**4.4.3.** La aplicación del IC da la verdadera dimensión a los datos sobre palabras diferentes.

Tabla N° 9  
Indice de cohesión y régimen de estudios

IC	CENTROS DE INTERÉS	
	PUBLICO	PRIVADO
0,05		CUE , RYC
0,04	CUE, RYC	TRA
0,03	MUE, ALI, TRA	MUE, ALI, PLA
0,02	COC, CAS, PLA	COC, CAS
0,01	JUE, PYO, CAM, POL, AMB	JUE, PYO, CAM, POL, ART, ECO, MUN, AMB
0,009	ART, ECO, MUN	CYT
0,008	CYT	PRO
0,007	PRO	

Los centros en los que se manifiesta más homogeneidad en ambos grupos se ubican en la parte alta de la tabla y –considerando también el centro TRA que para los estudiantes de establecimientos públicos tiene un IC de 0,3–, representan un 17% del total. Los centros menos cohesionados son PRO para ambos grupos y también CYT para el grupo que recibe educación pública. La experiencia codificada relativa a estos centros temáticos, no presenta grandes coincidencias en los grupos muestrales.

#### 4.5. La variable regionalidad

Considerando que poco más del 40% de los sujetos del curso terminal de Educación Media estudia en la Región Metropolitana, a continuación se presentan los datos léxicos comparando los índices de la Región Metropolitana con el resto de las Regiones como un bloque.

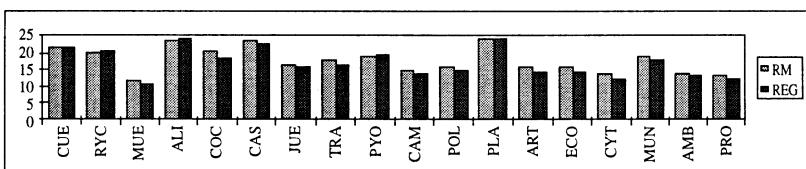
Tabla N° 10  
Indicadores  $\bar{x}R$ , PD e IC y regionalidad

CENTRO	REGIONALIDAD	$\bar{x}R$	PD	IC
CUE	RM	21.61	409	0.05
	REG	21.31	445	0.04
RYC	RM	20.27	419	0.04
	REG	20.44	443	0.04
MUE	RM	11.80	288	0.04
	REG	10.71	292	0.03

CENTRO	REGIONALIDAD	$\bar{x}R$	PD	IC
ALI	RM REG	23.61 24.12	684 742	0.03 0.03
COG	RM REG	20.59 18.29	769 798	0.02 0.02
CAS	RM REG	23.65 22.46	765 859	0.03 0.02
JUE	RM REG	16.50 15.98	1192 1260	0.01 0.01
TRA	RM REG	17.64 16.40	462 413	0.03 0.03
PYO	RM REG	19.11 19.13	1138 1230	0.01 0.01
CAM	RM REG	14.79 13.92	946 10.87	0.01 0.01
POL	RM REG	15.91 14.70	1410 1428	0.01 0.01
PLA	RM REG	24.36 24.21	729 846	0.03 0.02
ART	RM REG	15.98 14.31	1420 1473	0.01 0.009
ECO	RM REG	16.05 14.41	1297 1582	0.01 0.009
CYT	RM REG	13.48 12.01	1347 1395	0.01 0.008
MUN	RM REG	19.04 17.77	1592 1838	0.01 0.009
AMB	RM REG	13.88 13.10	1109 1157	0.01 0.01
PRO	RM REG	13.28 11.96	1529 1669	0.008 0.007

**4.5.1.** Si se observa el siguiente gráfico, se podrá apreciar más fácilmente la distribución resultante al aplicar a los datos el índice  $\bar{x}R$ .

Gráfico N° 8  
Promedio de respuestas por regionalidad



La conclusión general es que, en la mayoría de los centros, los estudiantes de la Región Metropolitana presentan mayor riqueza léxica que los de las otras Regiones aunque, en general, las diferencias son pequeñas. Estos últimos, aventajan a los de la Región Metropolitana en los centros RYC y ALI. Ambos grupos coinciden en el  $\bar{x}R$  obtenido en PYO, pero éste es un dato estadístico. En los listados por índice de disponibilidad léxica (IDL) se podrá apreciar cuál es efectivamente el léxico que manejan y sus características.

El  $\bar{x}R$  más alto está en el centro PLA: 24,36 la Región Metropolitana y 24,21 las Regiones. Los menores promedios se alcanzan en MUE: 11,80 la Región Metropolitana y 10,71 las Regiones.

Como se ve, la riqueza léxica de los estudiantes, según esta variable, está en las áreas temáticas que representan PLA, ALI, CAS, CUE y RYC, en orden decreciente. La mayor diferencia, a favor de la RM, se da, curiosamente, en COC.

**4.5.2.** El total de PD que aportan en cada centro de interés los dos grupos considerados por la variable se pueden apreciar, comparativamente, en la Tabla N° 11. Los sujetos que estudian en las diferentes Regiones del país presentan el corpus mayoritario en el centro MUN: 1838, en el que los de la RM sólo anotan 1592 palabras diferentes. Los resultados de este índice muestran que los estudiantes regionales conocen, en conjunto, más palabras que los metropolitanos en casi todos los núcleos nacionales abordados. El centro con menos vocablos es MUE (288) y corresponde al grupo Región Metropolitana.

Si interpretamos correctamente los datos de este índice, podemos decir que el vocabulario de los estudiantes de las Regiones acusa más dispersión que el de los de la RM. Esto podría atribuirse al hecho de que en el grupo que hemos llamado “Región” se recoge un léxico que apunta a realidades muy dispares, en tanto que en el grupo “Región Metropolitana”, la acotación geográfica podría favorecer la mayor acumulación de experiencias comunes. La respuesta la aportará un análisis más fino de los datos.

**4.5.3.** De acuerdo con el valor del IC, los grupos alcanzaron los resultados que se muestran a continuación.

Tabla N° 11  
Índice de cohesión Región Metropolitana vs Regiones

IC	CENTROS DE INTERÉS	
	REGIÓN METROPOLITANA	REGIONES
0,05	CUE	
0,04	RYC, MUE	CUE, RYC
0,03	ALI, TRA, CAS PLA	MUE, ALI, TRA
0,02	COG	COG, CAS, PLA
0,01	JUE, PYO, CAM POL, ART, ECO CYT, MUN, AMB	JUE, PYO, CAM, POL, AMB
0,009		ART, ECO, MUN
0,008	PRO	CYT
0,007		PRO

Se puede comprobar que el léxico de que disponen los estudiantes presenta una mayor coincidencia en los de la RM y es menos aglutinado en los sujetos que estudian en las Regiones. El centro más homogéneo es el CUE en el grupo RM. Compárese con el resultado anterior de 409 vocablos para este centro, lo que significa que los informantes manejan, poco más o menos, las mismas palabras al momento de elaborar un texto relacionado con la temática del centro. Volviendo a la Tabla, vemos que los resultados de este grupo se ubican en la parte alta del mismo. El centro que presenta la mayor heterogeneidad es PRO del grupo Regiones. Conviene relacionar estos resultados con los analizados en el punto anterior sobre palabras diferentes conocidas por los informantes: este centro registra 1669 vocablos.

#### 4.6. La variable procedencia poblacional de los sujetos

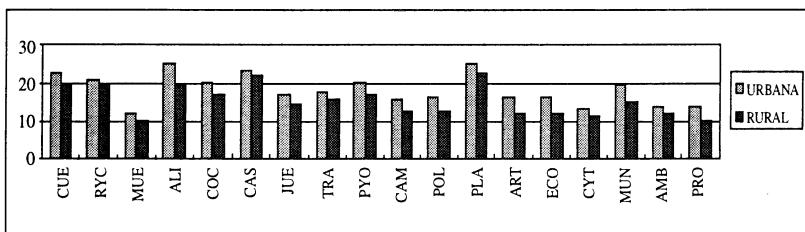
De acuerdo con la zona en que estudian los sujetos: urbana o rural, los resultados obtenidos en los índices  $\bar{x}R$ , PD e IC se presentan en la siguiente tabla. En la muestra, los jóvenes que estudian en establecimientos ubicados en zonas urbanas son 1622 y los que estudian en establecimientos de zonas rurales, son 430.

Tabla N° 12  
Indicadores  $\bar{x}R$ , PD e IC por procedencia poblacional

CENTRO	PROCED.	$\bar{x}R$	PD	IC
CUE	URBANA	21.94	490	0.04
	RURAL	19.50	310	0.06
RYC	URBANA	20.66	514	0.04
	RURAL	19.26	314	0.06
MUE	URBANA	11.49	366	0.03
	RURAL	9.81	179	0.05
ALI	URBANA	25.04	942	0.02
	RURAL	19.69	467	0.04
COG	URBANA	19.90	942	0.02
	RURAL	16.62	509	0.03
CAS	URBANA	23.26	982	0.02
	RURAL	21.64	560	0.03
JUE	URBANA	16.65	1556	0.01
	RURAL	14.47	766	0.01
TRA	URBANA	17.39	545	0.03
	RURAL	15.06	263	0.05
PYO	URBANA	19.74	1445	0.01
	RURAL	16.82	773	0.02
CAM	URBANA	15.50	1244	0.01
	RURAL	12.39	656	0.01
POL	URBANA	15.92	1819	0.008
	RURAL	12.43	777	0.01
PLA	URBANA	24.77	913	0.02
	RURAL	22.38	594	0.03
ART	URBANA	15.86	1854	0.008
	RURAL	11.66	765	0.01
ECO	URBANA	15.94	1839	0.008
	RURAL	11.79	818	0.01
CYT	URBANA	13.11	1763	0.007
	RURAL	10.68	761	0.01
MUN	URBANA	19.24	2173	0.008
	RURAL	14.67	963	0.01
AMB	URBANA	13.91	1465	0.009
	RURAL	11.53	625	0.01
PRO	URBANA	13.20	2087	0.006
	RURAL	9.79	729	0.01

**4.6.1.** El gráfico siguiente permite visualizar los  $\bar{x}R$  obtenidos en cada centro, por cada grupo muestral.

Gráfico N° 9  
Promedio de respuestas por procedencia poblacional



Se advierte que los promedios más altos son alcanzados por el grupo “urbano”, en todos los centros. Se puede decir, entonces, que este grupo muestral posee una mayor riqueza léxica que el grupo “rural”.

Los mayores promedios del grupo “urbano” se encuentran en ALI (25,04), PLA (24,77) y CAS (23.26), en tanto que los mayores del grupo “rural” son PLA (22,38) y CAS (21,64). El grupo “rural” presenta sus menores promedios de respuestas en los centros MUE (9,81) y PRO (9,79). Esto significa que, como grupo, hipotéticamente tendrían problemas para expresarse fluidamente en un discurso sobre la temática de dichos centros.

**4.6.2.** Los sujetos del grupo “urbano” presentan, en todos los centros, un mayor número de vocablos, especialmente en el centro PRO, donde alcanzan la suma de 2087 PD. Los sujetos del grupo “rural”, por su parte, alcanzan su máximo en el centro MUN, con 963 vocablos, lo que de cualquier manera resulta menoscabado frente a los 2173 vocablos que anotan como conocidos los del grupo “urbano”. El corpus menor de acuerdo con esta variable se encuentra en MUN “rural”, con 179 vocablos.

**4.6.3.** El índice de cohesión, como hemos dicho, permite acotar el valor de los índices anteriores. Veamos los resultados obtenidos en esta variable.

Tabla N° 13  
Índice de cohesión y procedencia poblacional

IC	CENTROS DE INTERÉS	
	URBANO	RURAL
0,06		CUE, RYC
0,05		MUE, TRA
0,04	CUE, RYC	ALI
0,03	MUE, TRA	COC, CAS, PLA
0,02	ALI, COC, CAS PLA	PYO
0,01	JUE, PYO, CAM	JUE, CAM, POL, ART, ECO, CYT MUN, AMB, PRO
0,009	AMB	
0,008	POL, ART, ECO, MUN	
0,007	CYT	
0,006		PRO

Como se puede apreciar, el grupo “rural” maneja un léxico más homogéneo, en todas las áreas temáticas exploradas. En otras palabras, su léxico corresponde a la codificación de una experiencia compartida por todos los individuos del grupo. Además, a juzgar por el índice PD alcanzado, su léxico es reducido. El grupo “urbano” demuestra poseer más vocabulario como grupo, pero la falta de cohesión de ese léxico indica que no está disponible para todos o no lo está en el mismo grado.

## 5. CUADRO COMPARATIVO

Si se considera –como en todos los estudios que se realizan actualmente– que el promedio de respuestas dadas por los sujetos refleja la riqueza de su léxico disponible, comparando los resultados obtenidos en este índice por los subgrupos de estratificación de la muestra, se podría comprobar cuál de los factores extralingüísticos gravita con mayor intensidad sobre el aspecto del lenguaje que analizamos aquí.

A continuación se presenta el promedio de respuestas por sexo (M/F), nivel sociocultural (B/M/A), régimen educacional (PUB/PRI), regionalización (RM/REG) y procedencia poblacional (U/R).

TABLA N° 14

Promedio de respuestas por centro de interés en las cinco variables de estratificación de la muestra.

C de INT	SEXO		NSC			REG.ED		REG.		PROCED.	
	M	F	B	M	A	PUB	PRI	RM	REG	URB	RUR
CUE	22	21	19	22	24	20	23	22	21	22	20
RYC	19	21	19	20	23	19	22	20	20	21	19
MUE	11	11	10	11	12	11	12	12	11	12	10
ALI	23	25	21	24	28	23	26	24	24	25	20
COG	19	20	17	19	23	18	21	21	18	20	17
CAS	23	23	21	23	25	22	24	24	23	23	22
JUE	16	16	15	16	18	16	17	17	16	17	15
TRA	17	17	15	17	20	16	18	18	16	17	15
PYO	19	19	17	19	22	18	20	19	19	20	17
CAM	14	14	12	15	17	14	15	15	14	16	12
POL	15	15	13	16	17	14	17	16	15	16	12
PLA	24	24	23	24	27	23	25	24	24	25	22
ART	15	15	12	15	18	14	17	16	14	16	12
ECO	15	14	12	16	18	14	17	16	14	16	12
CYT	13	12	11	13	15	12	14	14	12	13	11
MUN	18	18	15	19	22	17	20	19	18	19	15
AMB	13	13	12	14	16	13	14	14	13	14	12
PRO	11	13	11	13	15	12	14	13	12	13	10

La tabla permite comprobar que el NSC es, con mucho, el factor que establece la discriminación entre los grupos. Al igual que en los resultados obtenidos en otros países de habla hispana, se constata, una vez más, que existe una estrecha relación entre este factor y la cantidad de léxico que maneja el sujeto. Como el nivel sociocultural está asociado con el tipo de establecimiento al que asiste el estudiante, también la variable régimen de estudios presenta promedios de respuesta menores para el grupo público en relación con el grupo privado.

El otro factor que, sin duda, influye en el léxico disponible es la procedencia poblacional: sin excepción, el grupo que estudia en un lugar relativamente aislado maneja un promedio de palabras notablemente menor, para cada área temática, que el que maneja el grupo

que estudia en un centro urbano. La diferencia entre los sujetos de la Región Metropolitana y el resto de las Regiones se aprecia ligeramente favorable a la primera.

## 6. COMENTARIOS

Este acercamiento cuantitativo al léxico disponible de los estudiantes de cuarto año medio permite las primeras consideraciones con respecto a su riqueza léxica. El examen de este material desde el punto de vista cualitativo queda fuera de los límites propuestos en este artículo.

Con los resultados presentados aquí, es posible adelantar algunas conclusiones:

- a) Corroborando investigaciones anteriores, no se advierten diferencias significativas entre hombres y mujeres en cuanto a la cantidad de léxico disponible, aun cuando las niñas presentan, en general, mejores promedios que los varones.
- b) La variable nivel sociocultural es la que establece una marcada diferencia entre los sujetos con respecto a la cuantificación de su léxico disponible. Así, se perfilan claramente tres sociolectos: el alto, con un mayor número de palabras disponibles para cada centro; el bajo, con una cantidad notablemente menor; y el medio, entre ambos, equidistante en la mayoría de las áreas temáticas encuestadas.
- c) El régimen educacional también permite comprobar diferencias, puesto que está en relación directa con el nivel sociocultural. Los estudiantes de establecimientos fiscales presentan, cuantitativamente, menos manejo léxico que los de establecimientos privados.
- d) El contingente estudiantil de la Región Metropolitana, aparentemente, tiene una mayor cantidad de léxico disponible que el que estudia en las otras doce Regiones, pues presenta, en casi todos los centros de interés, mejores promedios y corpus más homogéneos.
- e) La variable procedencia poblacional hace explícita una realidad: los estudiantes de los centros urbanos están estimulados lingüísticamente en un grado mucho mayor que el de sus pares de zonas rurales. Por lo tanto, estos últimos presentan un léxico bastante reducido.
- f) Los resultados cuantitativos obtenidos son similares a los obtenidos en Puerto Rico, República Dominicana y Las Palmas de Gran Canaria en cuanto a la fuerte influencia que ejerce sobre el

- léxico el nivel sociocultural del sujeto; y a los resultados de República Dominicana en cuanto al léxico no diferenciado por sexo y la influencia diferenciadora del régimen de estudios.
- g) Los centros de interés en los que los estudiantes muestran un léxico más homogéneo son “Las partes del cuerpo”, “Ropa y calzado”, “Muebles” y “Medios de transporte”. Los que aparecen más dispersos o abiertos son “Mundo espiritual”, “Ciencia y tecnología” y “Procesos mentales”, en los que parece primar la experiencia individual.
  - h) Los centros de interés que alcanzan mayores promedios de respuestas en los estudiantes de la muestra son “Plantas y animales”, “Alimentos” y “Casa”. Los que aparecen como más deficitarios son “Muebles”, “Ciencia y tecnología” y “Procesos mentales”.
  - i) La muestra recogida es la de mayor volumen con que se ha trabajado en este tipo de estudios. Además, por tratarse de un muestreo con una red de puntos que abarca todo el territorio nacional, el material recolectado puede dar cuenta, de un modo bastante fidedigno, de una interesante gama de fenómenos léxicos presentes en los sujetos del último nivel de escolaridad secundaria en Chile.

## BIBLIOGRAFIA

- ALBA, Orlando. 1995. “El léxico disponible de la República Dominicana”, Santiago de los Caballeros, Ediciones de la Pontificia Universidad Católica Madre y Maestra.
- \_\_\_\_\_. 1996. *Disponibilidad léxica en el español dominicano: Aspectos sociolingüísticos*, Actas del X Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina, México, UNAM: 742-749.
- BENÍTEZ, Pedro. 1992. *Disponibilidad léxica en la zona metropolitana de Madrid*, “Boletín de la Academia Puertorriqueña de la Lengua Española” 1,1: 71-102.
- ECHEVERRÍA, Max S. 1990. *Crecimiento del vocabulario pasivo en comunidades escolares*, “Actas del VIII Seminario de Investigación y Enseñanza de la Lingüística”, Santiago, Central de Publicaciones de la Facultad de Humanidades de la Universidad de Santiago de Chile: 272-282.
- \_\_\_\_\_. 1991. *Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio*, “Actas del II Seminario Internacional sobre aportes de la lingüística a la enseñanza del español como lengua materna”, Río Piedras, Edit. de la Universidad de Puerto Rico: 61-78.
- \_\_\_\_\_. et al. 1987. *Disponibilidad léxica en Educación Media*, “RLA Revista de Lingüística Teórica y Aplicada”, 25: 55-112.
- LÓPEZ CHÁVEZ, Juan. 1992. *Alcances panhispánicos del léxico disponible*, “Lingüística”, 4: 26-124.

- LÓPEZ-MORALES, Humberto. 1973. *Frecuencia léxica, disponibilidad y programación curricular*, “Boletín de la Academia Puertorriqueña de la Lengua Española”, VI/1:73-86.
- \_\_\_\_\_. 1979. *Disponibilidad léxica y estratificación socioeconómica*, “Dialectología y Sociolingüística. Temas puertorriqueños”, Madrid, Ed. Playor: 173-181.
- \_\_\_\_\_. 1983. *Lingüística estadística*, H. López-Morales (ed.), Madrid, Ed. Playor: 209-225.
- \_\_\_\_\_. 1996. *Los estudios de disponibilidad léxica: pasado y presente*, “Boletín de Filología de la Universidad de Chile”, XXXV: 245-259.
- MENA, Mónica. Tesina inédita. “Disponibilidad léxica infantil en tres niveles de Enseñanza Básica”, Universidad de Concepción.
- MICHEA, René. 1953. *Mots fréquents et mots disponibles*. Un aspect nouveau de la statistique du langage, “Les Langues Modernes”, 47: 338-344.
- SAMPER, José Antonio y Clara HERNÁNDEZ. Inédito. *El estudio de la disponibilidad léxica en Gran Canaria: Datos iniciales y variación sociolingüística*.
- VALENCIA, Alba; M. VÉLIZ, E. ÁVILA, G. MUÑOZ, N. NÚÑEZ y M. ECHEVERRÍA. 1991. *Evaluación de la riqueza léxica en estudiantes chilenos de cuarto medio*, “Estudios Filológicos”, 26: 71-81.
- VALENCIA, Alba. 1994. “El léxico de los estudiantes de 4º año de Educación Media. Centro de Interés “Procesos Mentales”, C.P.E.I.P., Serie Documentos de Estudio, 26.
- \_\_\_\_\_. 1994. “Disponibilidad léxica en Educación Media: III y IV”, X Encuentro en Torno a la Admisión Universitaria, Santiago, Talleres Gráficos DIVEST.
- \_\_\_\_\_. y Max ECHEVERRÍA. 1996. *El factor geográfico en la disponibilidad léxica chilena*, “Actas XI Congreso de la ALFAL”, Las Palmas de Gran Canaria, en prensa.