



Artículo de Investigación

Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación

Frequent vocabulary, lexical richness and stereotypes of reading in training teachers

Recibido: Enero 2020 Aceptado: Abril 2020 Publicado: Junio 2020

María Natalia Castillo Fadić

Pontificia Universidad Católica de Chile, Chile
mcastilf@uc.cl

Enrique Sologuren Insúa

Universidad de los Andes, Chile
esologuren@miuandes.cl

Resumen: El presente artículo analiza lexicoestadísticamente un corpus textual de relatos de vida lectora escritos por profesores en formación, con énfasis en los índices de frecuencia, riqueza léxica, representatividad y curva de cobertura; y compara los resultados obtenidos con los de otros corpus de español. Además, analiza cualitativamente las unidades léxicas de mayor frecuencia desde la semántica del estereotipo. Desde el punto de vista cuantitativo, releva los cien vocablos más frecuentes del corpus, determina que la curva de cobertura es más acelerada que las de otros corpus escritos y se acerca a la de corpus orales, precisa el índice de riqueza léxica a partir de dos fórmulas complementarias y presenta los listados de vocablos con contenido nocional y de nombres propios con $Fi \geq 2$, con indicación de su rango. Esto se acompaña de un análisis por estereotipos que da cuenta de una marcada asociación entre lectura y géneros literarios, así como de asociaciones contradictorias, entre las que destaca, por un lado, la de lectura con placer, tiempo libre y capacidad adquisitiva y, por otro, la de lectura con aburrimiento, obligatoriedad, instituciones educativas y bibliotecas.

Palabras clave: lexicoestadística - riqueza léxica - lexicosemántica - semántica del estereotipo

Citación: Castillo Fadić, M. N. & Sologuren Insúa, E. (2020). Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación. *Logos: Revista de Lingüística, Filosofía y Literatura*, 30(1), 69-85. DOI: 10.15443/RL3006

Dirección Postal: Av. Vicuña Mackenna 4860, Macul, Santiago, Chile

DOI: doi.org/10.15443/RL3006



Abstract: The present article analyzes a textual corpus formed by stories of reading life written by teachers in training, with emphasis on frequency index, lexical richness, representativeness and coverage curve with the use of statistics, and compares the results obtained through other corpora of Spanish language. In addition, it analyzes the most frequent lexical units from the Stereotype Semantics field qualitatively. From a qualitative point of view, it gathers the one hundred most frequent words in the corpus, determines a more accelerated coverage curve than the rest of the written corpora and it shows a similar curve to the one of the oral corpora. It also specifies the lexical richness index from complementary formulas and presents the list of words with notional content and proper names with $F_i \geq 2$, with range display. This information is accompanied by a stereotype analysis that shows a strong association between reading and literary genres, as well as contradictory associations between, on the one hand, reading with pleasure, free time and purchasing power and, on the other hand, reading with boredom, obligation, educational institution and libraries.

Keywords: lexical statistics - lexical richness - lexical semantics - stereotype semantics

1. Introducción

Los trabajos de estadística léxica comienzan en el mundo hispano a mediados del siglo XX, cuando lingüistas como Rodríguez Bou (1952a y 1952b) y Juilland y Chang-Rodríguez (1964) realizan estudios de frecuencia léxica a partir de corpus de referencia. Estos últimos desarrollan además fórmulas de dispersión compleja y uso. Más tarde, López Morales (1973) extiende esta línea de investigación cuantitativa a los estudios de disponibilidad léxica, a partir de la aplicación de tests asociativos organizados por centros de interés o áreas temáticas, ámbito en el que realiza numerosos aportes (cf., por ejemplo, Lorán & López Morales, 1983 y López Morales, 1999). Tanto los estudios de frecuencia como los de disponibilidad, complementarios, dan cuenta de distintos índices y tienen amplias posibilidades de aplicación en la enseñanza-adquisición de la lengua materna y de segundas lenguas (cf. Germain, 1993; Santos Díaz, 2017a y 2017b; Trigo, Romero y Santos Díaz, 2019). En ambos, importa determinar cuál es el núcleo estadístico del español, ya sea en lo relativo al léxico atemático de mayor uso, como al léxico temático, de mayor disponibilidad, organizado por centros de interés. Uno de los índices de mayor aplicación al ámbito pedagógico ha sido el de *riqueza léxica*, cuyo precursor en español es Humberto López Morales (1984). Respecto de este índice se han desarrollado distintas fórmulas, de las que daremos cuenta más adelante.

Este artículo analiza estadísticamente léxico en uso de profesores en formación en un corpus textual compuesto por textos pertenecientes a un género discursivo específico: *los relatos de vida lectora* (cf. Munita, 2017; Parrado, Romero & Trigo, 2018). El trabajo revisa los índices de frecuencia (F_i), representatividad acumulada, cobertura y riqueza léxica. Luego, analiza los estereotipos presentes en las unidades léxicas más frecuentes, con énfasis en aquellas con contenido nocional.

El estudio exploratorio, lexicoestadístico y lexicosemántico tiene finalidad pura e implicancias aplicadas; de corte sincrónico, sintópico, sinstrático y sinfásico, opera sobre un corpus empírico de primera mano y aplica herramientas de lingüística de corpus, lingüística computacional y estadística descriptiva.

El objetivo general es determinar el léxico en uso de profesores en formación. Los objetivos específicos son:

- Describir estadísticamente el léxico del grupo mencionado.
- Comparar los índices estadísticos con los de otros corpus de español.
- Revisar cualitativamente el léxico desde la semántica del estereotipo.

Esperamos que nuestros resultados contribuyan al conocimiento de la relación de futuros profesores con la lectura y, de manera mediata, puedan ser de utilidad para investigaciones ligadas al ámbito educativo.

2. Lexicoestadística y lexicosemántica

La estadística léxica se centra en el nivel léxico de la lengua desde un punto de vista cuantitativo. Por ello, opera sobre la unidad mínima de léxico, el *vocablo*, y no sobre la unidad mínima textual, la *palabra*. En este sentido, supone la necesidad de *segmentar* y *lematizar* las unidades léxicas actualizadas en los textos, de modo de abstraer la variable bajo la cual se agrupan los tipos léxicos o variantes que, a su vez, aparecen representados en dichos textos por medio de palabras. En lo relativo a la segmentación, importa distinguir no sólo unidades univocales de plurivocales, sino también resolver casos de amalgamas, construcciones con pronombres enclíticos, etc. Las variables, vocablos, lemas o unidades de cita operan como etiqueta de todo un paradigma de variantes y su determinación es convencional: así, en español, en el caso de los verbos, el lema es el infinitivo; en el de las formas nominales, usualmente el masculino singular; en el de otras clases léxicas, pueden existir distintas alternativas; por ejemplo, es posible lematizar los pronombres personales por persona y número, por persona con independencia del número e incluso separadamente según si se trata de un caso nominativo, dativo o ablativo, como hacemos en esta ocasión. Estas decisiones inciden, como es lógico, sobre el conteo estadístico y, normalmente, deben estar justificadas de acuerdo con el marco teórico y metodológico y alineadas con los objetivos de la investigación.

En esta investigación, la segmentación y la lematización siguen *grosso modo* los criterios de Castillo Fadić (2020). No obstante, se consideran también como unidades plurivocales algunas construcciones con carácter de colocación (cf. Bosque, 2001) relevantes para este estudio; es el caso de, por ejemplo, *séptimo_básico*, que permite ubicar temporalmente momentos relevantes en la vida lectora de los informantes.

La frecuencia aquí es absoluta (F_i), pues no está ponderada por el tamaño de la muestra. Para el cálculo de la representatividad acumulada (R), ordenamos los vocablos por frecuencia descendente y aplicamos la fórmula presentada en Castillo Fadić (2012b), a saber: $R = \frac{\sum_{i=1}^n F_i}{N}$, que corresponde al cociente entre la sumatoria de las frecuencias de los vocablos dentro de un rango y el total de vocablos en el mismo rango. La riqueza léxica (r), por su parte, se mide a partir de dos índices complementarios propuestos por López Morales (1984): el primero da cuenta del cociente entre el total de vocablos y el total de palabras de un corpus; el segundo, en tanto, observa el intervalo de palabras nocionales –sustantivos, adjetivos, verbos y adverbios– dentro de un corpus. Estos índices son ampliamente desarrollados en Haché de Yunén (1991) y López Morales (2011).

En lo relativo al análisis semántico, nos basamos en el concepto de *estereotipo* como lo entiende Lara (2001 y 2004), con quien concordamos en que las unidades léxicas presentan elementos semánticos teñidos socialmente, que dan cuenta de la cosmovisión de la comunidad. A diferencia del prototipo, el estereotipo no es universal, se asocia a las características de la sociedad y es, por tanto, dependiente de la experiencia y de las ideologías de los hablantes.

Puesto que los análisis realizados requieren operar sobre un corpus empírico homogéneo, para efectos de esta investigación entendemos los relatos de vida lectora como género discursivo (Parodi, 2008; Bazerman, 2012) y estimamos que este género puede ser definido como una interacción comunicativa cuyo modo de organización discursivo predominante es narrativo y se caracteriza por una fuerte orientación reflexiva. Así, este género reflexivo (Jarpa y Becerra, 2019) se construye a través de las historias que un estudiante cuenta de sí mismo (Romero y

Trigo, 2018) y a partir de las experiencias personales del sujeto lector (Munita, 2017). De allí se desprende su valor formativo e investigativo como herramienta heurística y también como género discursivo de formación. De este modo, nuestro corpus textual simple está constituido por material lingüístico correspondiente únicamente al género discursivo relato de vida lectora.

3. El relato de vida lectora como género discursivo

El concepto de género discursivo resulta fundamental para indagar en lo que las comunidades hacen a través de sus textos. La noción moderna de género se debe a Mijaíl Bajtín, quien en los años 50 escribió una serie de textos inéditos sobre los géneros discursivos, que fueron conocidos en occidente a partir de los años 80. Los géneros discursivos son, para Bajtín (2005, p. 248), tipos de enunciados relativamente estables que pertenecen a esferas de la comunicación determinadas. Se trata de unidades de comunicación discursiva que se caracterizan por el contenido temático, el estilo verbal –“o sea por la selección de los recursos léxicos, fraseológicos y gramaticales de la lengua” (Bajtín, 2005, p. 248)– y la composición o estructuración, tres aspectos inseparables que caracterizan a todo enunciado.

Estos tipos de enunciados pueden ser orales o escritos, primarios o secundarios. Los primarios o simples (Bajtín, 1986) se adquieren por medio de las interacciones de crianza y se corresponden con diálogos cotidianos, conversaciones diarias, anécdotas y narraciones breves, principalmente orales; por su parte, los secundarios o complejos requieren mayor elaboración y generalmente son de modalidad escrita. Su adquisición y desarrollo está mediatizado por las instituciones escolares y de instrucción formal. Asimismo, los géneros primarios pueden integrarse a los géneros secundarios ya que “mientras los géneros discursivos se van formando, estos absorben y alteran los géneros primarios o simples” (Beke, 2011, p. 43). Visto lo anterior y desde esta perspectiva, en este artículo entendemos el relato de vida lectora como género discursivo escrito secundario.

Desde los estudios de la composición norteamericana, destaca la corriente de la nueva retórica –los Estudios Retóricos del Género (RGE, por sus siglas en inglés)– que postula que “Los géneros se corresponden con formas estandarizadas de expresiones que son reconocidas como las que llevan a cabo ciertas acciones en ciertas circunstancias y hacia entendimientos estandarizados de las situaciones” (Bazerman, 2012, p. 131). De fuerte raigambre bajtiniana y socioconstructivista, esta concepción de los géneros releva con mayor énfasis el carácter accional de estas formas, al catalogarlas como acciones tipificadas “by using these typified texts we are able to advance our own interests and shape our meanings in relation to complex social systems” (Bazerman, 1994, p. 79). En este marco, es muy importante distinguir que los géneros no corresponden a categorías fijas o estáticas atribuidas a un texto, sino a categorías de reconocimiento psicosocial, por lo cual “el género discursivo es lo que las personas, en tanto grupos y en tanto individuos, identifican como tal” (Bazerman, Little, Bethel, Chavquin, Fquette & Garufis, 2016, p. 170).

Se trata, por tanto, de una cuestión de percepción (Bazerman, 2015, p. 81) que va más allá de las apariencias convencionales superficiales. Esta concepción de género es relevante para nuestra investigación, puesto que no sólo permite tipificar el relato de vida lectora como género, sino que, además, al trascender los aspectos visibles de forma y contenido y plantear que “Los géneros discursivos son maneras de hacer cosas –y como tales encarnan aquello que se quiere hacer y muestran huellas del tiempo y del espacio en que se realizan, así como de los motivos y las acciones realizadas” (Bazerman, 2013, p. 24), se relaciona con la noción de estereotipo empleada en el análisis lexicosemántico, que también da cuenta del reflejo cultural y social, esta vez a través del significado léxico.

4. Metodología

El corpus consta de 7423 palabras en contexto, distribuidas en 88 párrafos extraídos de once relatos de vida lectora escritos por estudiantes universitarios chilenos del área de la pedagogía

inicial o primaria. Los relatos corresponden a fuentes primarias y fueron obtenidos en el marco del proyecto Fondecyt Regular n° 1170779. Este género discursivo, consistente en una redacción solicitada *ad hoc* a los informantes, se ha considerado una herramienta útil en los estudios educativos para acceder a la motivación por la lectura de los sujetos (Munita, 2017; Parrado, Romero & Trigo, 2018).

La consigna para obtener los relatos fue la siguiente:

Escriba su autobiografía lectora desde sus primeros recuerdos de lectura, intentando responder a la pregunta: a) ¿Qué tipo de lector estima usted ser? b) ¿Cómo cree usted que llegó a ser este tipo de lector? Al llegar a la actualidad, incluya toda la información que usted crea pertinente para describirse a sí mismo como lector (rutinas de lectura, tipos de lectura, últimos libros leídos, frecuentación de otros materiales escritos, etc.). (Fondecyt Regular n° 1170779).

El corpus fue procesado mediante el programa LexBas 1.0¹. Tras la segmentación y lematización automática encontramos 3408 unidades univerbales y pluriverbales, correspondientes a 2004 tipos y 1404 vocablos. Luego de la posterior y complementaria segmentación y lematización manual, las cifras quedaron en 3383 palabras, 1990 tipos y 1393 vocablos (cf. Figura 1).

Figura 1. Recorte de pantalla LexBas 1.0

	Total	Forms	Lemma
Report	3383	1990	1393

El análisis se centró primero en aspectos cuantitativos. Así, una vez obtenida la F_i por vocablo mediante Lexbas 1.0, los datos se exportaron a una planilla de cálculo, donde se computaron separadamente las unidades con contenido nocional, los nombres propios y las demás categorías detectadas. A continuación, se aplicó la fórmula de representatividad acumulada presentada en Castillo Fadić (2012b) para medir la velocidad de cobertura del corpus y establecer gráficamente su curvatura. Luego, se midió la riqueza léxica usando los dos parámetros complementarios, propuestos por López Morales (1984). A continuación, se realizó un análisis cualitativo de los vocablos con contenido nocional con $F_i \geq 2$, desde la semántica del estereotipo (Lara, 2001).

5. Resultados y discusión

5.1 Vocablos de mayor frecuencia: los primeros cien

Los cien vocablos de mayor frecuencia (cf. Tabla 1), con una $R=65,57995\%$, están encabezados por palabras gramaticales, como es habitual en este tipo de listado. No obstante, y a diferencia de corpus de concordancia complejos que abarcan diferentes géneros y temáticas, la falta de índices de dispersión que modulen los resultados se observa en la presencia de deícticos, entre los que destacan pronombres personales y adjetivos posesivos. También aparecen en rangos iniciales (8 en adelante) unidades con contenido nocional de alta especificidad semántica y fuertemente temáticos (ej. *leer*, *libro lectura*), claramente motivados por la instrucción de escritura y por el género en cuestión.

Tabla 1. Cien vocablos de mayor frecuencia

Rango	Vocablo	Pos (EAGLES)	F _i	R
1	el	DA	641	8,635%
2	de	SP	360	13,485%
3	que	PR	225	16,516%
4	en	SP	216	19,426%
5	y	CC	210	22,255%
6	a	SP	199	24,936%
7	me	PP	188	27,469%
8	leer	VM	170	29,759%
9	mi	DP	149	31,766%
10	ser	VS	131	33,531%
11	libro	NC	128	35,255%
12	por	SP	117	36,831%
13	uno	DI	115	38,381%
14	que	CS	111	39,876%
15	no	RN	78	40,927%
16	con	SP	76	41,951%
17	para	SP	74	42,948%
18	lectura	NC	66	43,837%
19	tener	VM	61	44,658%
20	como	CS	57	45,426%
21	o	CC	55	46,167%
22	este	DD	47	46,800%
23	más	RG	46	47,420%
24	estar	VA	44	48,013%
25	lo	PP	41	48,565%
26	texto	NC	39	49,091%
27	se	P0	33	49,535%
28	poder	VM	32	49,966%
29	lector	NC	29	50,357%
30	tiempo	NC	28	50,734%
31	gustar	VM	27	51,098%
32	cuando	CS	25	51,435%
33	hacer	VM	25	51,772%
34	vez	NC	25	52,108%
35	ir	VM	24	52,432%
36	pero	CC	24	52,755%
37	cuento	NC	23	53,065%
38	ese	DD	23	53,375%
39	siempre	RG	23	53,684%
40	año	NC	22	53,981%
41	otro	DI	22	54,277%
42	su	DP	22	54,574%

Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación

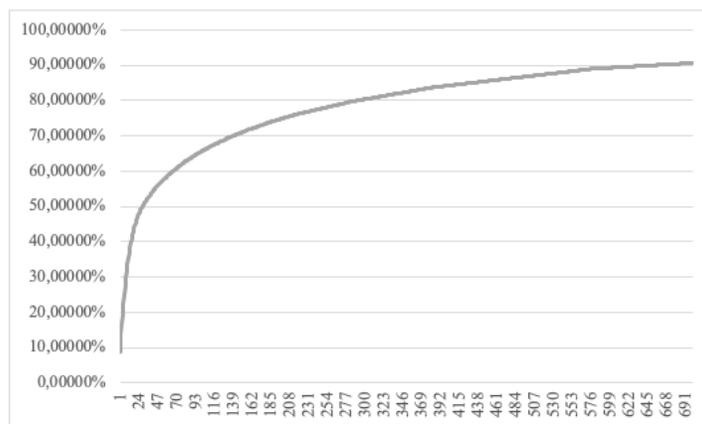
43	comenzar	VM	21	54,857%
44	los	PP	21	55,139%
45	sobre	SP	21	55,422%
46	todo	DI	21	55,705%
47	porque	CS	20	55,975%
48	día	NC	18	56,217%
49	momento	NC	18	56,460%
50	primero	AO	17	56,689%
51	comprar	VM	17	56,918%
52	desde	SP	17	57,147%
53	gusto	NC	17	57,376%
54	universidad	NC	17	57,605%
55	uno	PI	17	57,834%
56	conocer	VM	16	58,049%
57	cosa	NC	16	58,265%
58	deber	VM	16	58,480%
59	también	RG	16	58,696%
60	buscar	VM	15	58,898%
61	cual	PR	15	59,100%
62	historia	NC	15	59,302%
63	mismo	AQ	15	59,504%
64	recordar	VM	15	59,706%
65	si	CS	15	59,908%
66	alguno	DI	14	60,097%
67	encontrar	VM	14	60,286%
68	este	PD	14	60,474%
69	llegar	VM	14	60,663%
70	mucho	DI	14	60,851%
71	muy	RG	14	61,040%
72	terminar	VM	14	61,229%
73	último	AO	14	61,417%
74	volver	VM	14	61,606%
75	aprender	VM	13	61,781%
76	menos	RG	13	61,956%
77	mundo	NC	13	62,131%
78	puesto_que	SP	13	62,306%
79	ver	VM	13	62,481%
80	colegio	NC	12	62,643%
81	considerar	VM	12	62,805%
82	crear	VM	12	62,966%
83	tipo	NC	12	63,128%
84	así	RG	11	63,276%
85	ellos	PP	11	63,424%
86	familia	NC	11	63,573%
87	luego	RG	11	63,721%

88	novela	NC	11	63,869%
89	nunca	RG	11	64,017%
90	pedir	VM	11	64,165%
91	persona	NC	11	64,314%
92	querer	VM	11	64,462%
93	sentar	VM	11	64,610%
94	tomar	VM	11	64,758%
95	vida	NC	11	64,906%
96	además	RG	10	65,041%
97	casa	NC	10	65,176%
98	durante	SP	10	65,311%
99	haber	VM	10	65,445%
100	mí	PP	10	65,580%

5.2 Índices de cobertura y representatividad acumulada

De los 1393 vocablos, 578 tienen $F_i \geq 2$, con una representatividad acumulada $R = 89,02\%$. Como es habitual en estudios de esta naturaleza, la curva de cobertura es acelerada en los primeros rangos y se va ralentizando hasta volverse asintótica, alrededor del rango 600 y coincidentemente con el descenso de las $F_i < 2$ (cf. Gráfico 1).

Gráfico 1. Curva de cobertura



Si se compara la velocidad de cobertura con la del *Corpus Básico del Español de Chile* © (Castillo Fadić, 2012a), se observa una cobertura más rápida (cf. Tabla 2), similar a la que se aprecia en corpus orales (Ávila, 1998, p. 257) y distinta de la que suele darse en los listados de frecuencia obtenidos a partir de fuentes escritas, donde “las primeras 15 palabras de un listado de frecuencias viene a representar un 25% del total de ocurrencias (tokens) aparecidas en un texto, las primeras 50 un 50%, las primeras 100 un 60%, las primeras 1.000 un 85%” (Izquierdo Gil, 2005, p. 71)”. A modo de hipótesis, esto podría dar cuenta de un dominio aún en desarrollo de géneros discursivos secundarios, reflejado en el recurso a mecanismos propios de géneros primarios, característicos del discurso oral.

Tabla 2. Comparación entre curvas de cobertura de relatos de vida lectora y los distintos mundos del *Corpus Básico del Español de Chile* ©.

Vocablos	Relatos de vida lectora	Drama	Narrativa	Ensayo	Técnico-Científico	Prensa
Primeros 50	56,68867%	52,45317%	50,36652%	50,15457%	50,38526%	49,86673%
Primeros 100	65,57995%	60,64938%	56,28222%	55,33256%	55,39573%	54,38839%
Primeros 500	86,91904%	77,15970%	71,24919%	71,37668%	71,69272%	68,61179%
Primeros 1000	94,70564%	83,27389%	78,02884%	79,32885%	79,94999%	76,20969%

5.3 Riqueza léxica

Si aplicamos la medida de número de vocablos (1393) sobre el total de palabras (7423), tenemos una riqueza léxica $r = 18,765\%$, superior a la de todos los mundos del *Corpus Básico del Español de Chile* ©; no obstante, estos resultados que parecen auspiciosos deben ser relativizados. Los resultados de esta fórmula dependen del n y a idéntica riqueza arrojan cifras mayores cuanto menor sea el tamaño del corpus; por ello, su valor debe observarse siempre en relación con el obtenido del análisis de corpus de tamaño similar. Y puesto que los distintos mundos del *Corpus Básico del Español de Chile* © presentan riquezas entre 9,7 y 12,7 para subcorpus de 100.000 palabras, la riqueza proporcional de este pequeño corpus de relatos de sólo 7423 palabras se ve considerablemente disminuida.

Por ello, en este trabajo en particular nos parece de mayor interés observar el intervalo de palabras con contenido nocional, que corresponde al cociente entre el tamaño del corpus (7423) y la F_i total de las palabras con contenido nocional (3670); puesto que se encuentran 1094 vocablos con contenido nocional, con una frecuencia total de 3670, según este parámetro, la riqueza léxica $r = 2,022$. Este valor es similar al obtenido del análisis estadístico del mundo Técnico-Científico del *Corpus Básico del Español de Chile* © y da cuenta tanto de la riqueza conceptual de los textos como del menor uso de conectores en relación con otros géneros discursivos. Es importante recordar, en relación con esta fórmula, que la riqueza es inversamente proporcional al tamaño del intervalo, de modo que, a menor intervalo, mayor riqueza.

5.4 Unidades con contenido nocional y nombres propios: estadísticos y estereotipos

Los estereotipos son preferentemente observables en las unidades con contenido nocional. Nos centramos por ello en los sustantivos, adjetivos, verbos y adverbios. Aunque no corresponden a clases de denotación, sino de designación, revisamos también los nombres propios más frecuentes.

Dado que la consigna entregada a los informantes contiene explícitamente los vocablos *lector*, *lectura*, *libro* y *escrito*, su presencia en el corpus, así como su elevada frecuencia, se considera motivada por la instrucción. Nos enfocamos, pues, de manera preferente, en unidades léxicas no contenidas en la consigna.

5.4.1 Sustantivos

Hay 499 vocablos en esta categoría, con un total de 1470 ocurrencias. De estos, 213 tienen $F_i \geq 2$, con 1183 ocurrencias. A diferencia de lo que sucede con listados de F_i obtenidos a partir del análisis de corpus estratificados, el hecho de que todos los textos analizados correspondan a un mismo género discursivo, con un mismo tipo de autor y sobre un mismo tema, amén, de la no consideración de índices de dispersión, incide en que los vocablos con contenido nocional que ocupan los primeros rangos presenten alta especificidad semántica y estén fuertemente vinculados a las circunstancias de enunciación e, incluso, a las instrucciones de realización de

la tarea (*libro, lectura y lector*). Texto es el primer sustantivo de alta especificidad que no figura en la consigna, por lo que podríamos considerarlo en rigor el sustantivo que con más frecuencia los futuros profesores asocian con la lectura. Se trata de un hiperónimo de unidades léxicas presentes en la instrucción, como “autobiografía lectora”, “tipo de lectura”, “libro” o “material escrito”.

A continuación, encontramos una serie de vocablos referidos al tiempo, desde el vocablo homónimo *tiempo* hasta *vez, año, día, momento, periodo, etapa, instancia, noche, semana, verano*, etc., que dan cuenta de que los futuros profesores reconocen instancias específicas para la lectura. Como contraparte, este reconocimiento puede funcionar como excusa para no leer, en la medida en que no todas las instancias serían propicias para esta actividad determinada fuertemente por la temporalidad.

Otros vocablos dan cuenta de los tipos de lectura evocados por los informantes: desde el más amplio, *texto*, hasta *cuento, historia, novela, saga, artículo*, entre otros. La asociación de la lectura con géneros literarios es predominante.

Los espacios que se asocian preferentemente con la lectura son instituciones formales como *universidad, colegio y biblioteca* (F_i total=36), que suman mayor frecuencia que *casa*, lo que parece contradecir la asociación entre lectura y tiempo libre que los informantes esbozan discursivamente en los mismos textos, por cuanto si la lectura se asocia preferentemente con espacios de educación formal o que cuentan con reglas y horarios fijos, no puede asociarse de manera especial con el tiempo libre.

Como agentes motivadores de la lectura, encontramos *familia* ($F_i = 11$), *padre / papá* ($F_i = 12$), *madre / mamá* ($F_i = 9$), *profesor / docente* ($F_i = 10$), *amigo* ($F_i = 4$), *hermano* ($F_i = 3$), *tío* ($F_i = 3$) y *compañero* ($F_i = 2$), lo que da cuenta de la relevancia del núcleo familiar en este ámbito ($F_i = 38$), por sobre la figura del académico ($F_i = 10$) o del par ($F_i = 6$).

Resulta también de interés la alta frecuencia de unidades léxicas que dan cuenta de etapas de la vida, como *niño, infancia y adolescencia*, con énfasis en las dos primeras, donde los informantes anclan su vinculación con la lectura.

En la Tabla 3 se presentan los sustantivos de mayor frecuencia, con $F_i \geq 6$.

Tabla 3. Sustantivos con $F_i \geq 6$

Rango	vocablo	F_i
11	libro	128
18	lectura	66
26	texto	39
29	lector	29
30	tiempo	28
34	vez	25
37	cuento	23
40	año	22
48	día	18
49	momento	18
53	gusto	17

54	universidad	17
57	cosa	16
62	historia	15
77	mundo	13
80	colegio	12
83	tipo	12
86	familia	11
88	novela	11
91	persona	11
95	vida	11
97	casa	10
101	niño	10
109	infancia	9
110	internet	9
112	padre	9
118	tema	9
122	contenido	8
125	educación	8
127	forma	8
129	interés	8
131	profesor	8
133	recuerdo	8
139	biblioteca	7
152	página	7
154	parte	7
156	saga	7
161	artículo	6
162	atención	6
168	hábito	6
171	madre	6
175	obligación	6
176	palabra	6
178	prueba	6
180	relación	6
183	sensación	6
188	temática	6

5.4.2 Adjetivos

Hay 160 vocablos, con 315 ocurrencias totales. Sólo 58 presentan $F_i \geq 2$, con 213 ocurrencias. De ellos, nos centramos especialmente en los que tienen carácter de subjetivema (Kerbrat-Orecchioni, 1986, p. 93), donde destacan aquellos que aluden a la infancia, como pequeño e infantil ($F_i=17$); a la educación formal, como académico y escolar ($F_i=12$); a la importancia de la lectura, como bueno, importante, principal, necesario ($F_i=15$); al gusto de leer, como favorito, interesante, maravilloso y placentero ($F_i=15$); al disgusto de leer, como complejo, interminable, tedioso ($F_i=6$); en esta línea, llama la atención que los antónimos libre y obligatorio tengan idéntica $F_i=5$.

La Tabla 4 contiene los adjetivos más frecuentes, con $F_i \geq 5$.

Tabla 4. Adjetivos con $F_i \geq 5$

Rango	Vocablo	F_i
63	mismo	15
107	gran	9
114	pequeño	9
128	infantil	8
136	académico	7
149	lector	7
159	anterior	6
174	nuevo	6
200	escolar	5
203	favorito	5
206	interesante	5
207	libre	5
212	obligatorio	5

5.4.3 Verbos

En esta categoría hay 330 vocablos, con 1438 ocurrencias. De ellos, 150 presentan $F_i \geq 2$, con 1258 ocurrencias. Si descartamos los verbos auxiliares y semiauxiliares y dejamos sólo los modales, nos quedamos con sólo 147 verbos con 1077 ocurrencias.

Resulta llamativo que el primer verbo de alta especificidad semántica del listado, descontado *leer*, cuya aparición en el rango 1 se explica por la presencia de palabras de la misma familia léxica en la consigna, sea *comprar*; la asociación de compra y lectura ($f_i=17$) parece dar cuenta de una relación de dependencia, donde la posibilidad de leer se percibe como supeditada a la posibilidad de comprar y, por ende, a la capacidad económica, lo que se ve refrendado por la aparición de *económico* en el listado de adjetivos (cf. 5.4.2) y por la de *tener* (que no destacamos inicialmente por su menor especificidad semántica), *poseer*, *regalar* y *gastar* en el de verbos. Esta percepción, no se condice, sin embargo, con el hecho de que los espacios más asociados con la lectura sean las bibliotecas e instituciones educativas (cf. 5.4.1), donde la lectura no depende del poder adquisitivo; de hecho, *prestar* sólo tiene $F_i=2$.

En la Tabla 5 es posible revisar los verbos más frecuentes, con $F_i \geq 12$.

Tabla 5. Verbos con $F_i \geq 11$

Rango	Vocablo	F_i
8	leer	170
19	tener	61
28	poder	32
31	gustar	27
33	hacer	25
35	ir	24
43	comenzar	21
51	comprar	17
56	conocer	16
58	deber	16
60	buscar	15
64	recordar	15
67	encontrar	14
69	llegar	14
72	terminar	14
74	volver	14
75	aprender	13
79	ver	13
81	considerar	12
82	crear	12

5.4.4 Adverbios

En esta categoría hay 104 vocablos con un total de 447 ocurrencias. De estos, 50 vocablos tienen $F_i \geq 2$, con 393 ocurrencias. Si nos centramos en los subjetivemas, abundan los adverbios de tiempo, como *siempre, luego, nunca, ahora, antes, hoy, nuevamente, aún, después, finalmente, actualmente, a veces, constantemente, generalmente, posteriormente, anteriormente, a diario, jamás, tarde, todavía y usualmente*.

Los adverbios más frecuentes, con $F_i \geq 6$, se presentan en la Tabla 6.

Tabla 6. Adverbios con $F_i \geq 6$

Rango	vocablo	F_i
15	no	78
23	más	46
39	siempre	23
59	también	16
71	muy	14
76	menos	13

84	así	11
87	luego	11
89	nunca	11
96	además	10
103	ya	10
111	mucho	9
117	sólo	9
119	ahora	8
121	bien	8
160	antes	6
169	hoy	6
173	nuevamente	6
186	tan	6
187	tanto	6

5.4.5 Nombres propios

El corpus presenta 159 ocurrencias de nombres propios o unidades de designación. En esta categoría, sólo siete vocablos tienen $F_i \geq 2$, lo que da cuenta de la amplia dispersión de resultados; dos corresponden a nombres o apellidos de escritores (C.S. Lewis y Kafka) y cuatro a títulos de libros o sagas (*Crepúsculo*, *El mundo de Sofía*, *El principito* y *Harry Potter*). Todos presentan $F_i=2$, excepto *El Principito* con $F_i=4$.

6. Conclusiones

Puesto que el análisis lexicoestadístico y lexicosemántico de este corpus de relatos de vida lectora ha resultado provechoso para los propósitos de esta investigación, estimamos que este tipo de género discursivo puede ser, en efecto, una herramienta útil para acceder a los estereotipos sobre la lectura de los hablantes. Para poder realizar un análisis estadístico inferencial, sería conveniente tomar nuevos datos, para ampliar y estratificar sociolingüísticamente el corpus, de modo de permitir comparaciones que enriquezcan los análisis y que, además, arrojen resultados que puedan extrapolarse al universo en estudio.

Pese al carácter exploratorio de este trabajo, se han obtenido resultados que se perfilan con interesantes proyecciones. Entre ellos, destacan:

- Los cien vocablos de mayor frecuencia, con una $R=65,57995\%$.
- La curva de cobertura acelerada, más parecida a la de corpus orales que a la esperable en corpus escritos.
- La determinación de la riqueza léxica, semejante a la de textos técnicos y científicos en lo relativo al intervalo de vocablos con contenido nocional.
- La asociación preferente de la lectura con géneros literarios.
- Las asociaciones contradictorias de lectura con tiempo libre, con dinero y con placer, pero a su vez con obligatoriedad, tedio, bibliotecas e instituciones educativas.
- La consideración de la familia como promotora principal de la lectura, en detrimento de profesores y pares.

En investigaciones futuras, consideramos relevante comparar corpus de relatos de vida lectora estratificados por carrera, curso, nivel socioeconómico, sexo y edad, entre otras variables. También nos resulta de interés realizar un estudio contrastivo de los resultados obtenidos del

análisis lexicoestadístico y lexicosemántico con otro centrado específicamente en estrategias discursivas, para indagar en posibles contradicciones entre los niveles léxico-semántico y discursivo y profundizar en los mecanismos empleados por los hablantes para resolver estas posibles contradicciones en sus textos.

Agradecimientos y financiamiento

Este trabajo se desarrolló en el marco del Proyecto Fondecyt Regular n° 1170779.

Agradecemos la colaboración de Carla Muñoz, quien gestionó la recolección de los relatos de vida lectora y comentó el primer borrador de este artículo, y la de José Alejandro Martínez Lara, quien realizó valiosos comentarios a una de las últimas versiones de este trabajo.

Referencias bibliográficas

Ávila, A. M. (1998). *Elaboración, anotación y análisis del corpus oral del Proyecto V.U.M.* Málaga: Universidad de Málaga, Departamento de Filología Griega, Estudios Árabes y Traducción e Interpretación, Área de Lingüística General

Bajtín, M. (1986). *Problemas de la poética de Dostoievski*. México: F. C. E.

Bajtín, M. (2005). El problema de los géneros discursivos. En *Estética de la creación verbal* (pp. 248-293). Buenos Aires: Siglo XXI Editores.

Bazerman, C. (1994). Systems of genre and the enactment of social intentions. En A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-99). London: Taylor & Francis.

Bazerman, C. (2012). Actos de habla, géneros y sistemas de actividades: de qué manera los textos organizan las actividades y los grupos sociales. En C. Bazerman (Ed.), *Géneros textuales, tipificación y actividad* (pp. 122-161). Puebla: Benemérita Universidad Autónoma de Puebla.

Bazerman, C. (2013). *A rhetoric of literate action: Literate action* Volume I. Fort Collins: The WAC Clearinghouse & Parlor Press.

Bazerman, C. (2015). *Teoria da ação letrada. Tradução e adaptação de Milton Camargo Mota et al.* São Paulo: Parábola.

Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouquette, D. & Garufis, J. (2016). *Escribir a través del Currículum. Una guía de referencia*. Córdoba: Universidad Nacional de Córdoba.

Beke, R. (2011). *Las voces de los otros en el discurso académico*. Caracas: Universidad Central de Venezuela.

Bosque, I. (2001). Bases para un diccionario de restricciones léxicas. *Moenia*, 7, 11-52.

Castillo Fadić. (2012a). *Corpus Básico del Español de Chile* ©.

Castillo Fadić. (2012b). *Léxico Básico del Español de Chile*. Tesis para optar al grado de doctora, Universidad de Valladolid, España.

Castillo Fadić. (2020, en prensa). *Léxico Básico del Español de Chile*. Liberalia Ediciones. Fondo del Libro y la Lectura.

Germain, C. (1993). *Évolution de l'enseignement des langues : 5000 ans d'histoire*. Paris: CLE International.

Haché de Yunén, A. M. (1991). Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna. En H. López Morales (Ed.), *La enseñanza del español como lengua materna* (pp. 49-60). Río Piedras: Universidad de Puerto Rico.

Izquierdo Gil, M. (2005). *La selección del léxico en la enseñanza del español como lengua extranjera*. Málaga: ASELE, Colección Monografías n° 8.

Juilland, A., & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words, The Romance Languages and their Structures, First Series SI*. La Haya: Mouton.

Jarpa Azagra, M., & Becerra Rojas, N. (2019). Escritura para la reflexión pedagógica: concepciones y géneros discursivos que escriben los estudiantes en dos carreras de pedagogía. *Logos: Revista de Lingüística, Filosofía y Literatura*, 29(2), 364-381.

Kerbrat-Orecchioni, C. (1986). *La enunciación. De la subjetividad en el lenguaje*. Buenos Aires: Hachette.

Lara, L. F. (2001). *Ensayos de teoría semántica: lengua natural y lenguajes científicos*. México: El Colegio de México.

Lara, L. F. (2004). *De la definición lexicográfica*. México: El Colegio de México.

López Morales, H. (1973). *Disponibilidad léxica de los escolares de San Juan*. MS inédito.

López Morales, H. (1984). *La enseñanza de la lengua materna*. Madrid: Playor.

López Morales, H. (1999). *Léxico disponible del español de Puerto Rico*. Madrid: Arco Libros.

López Morales, H. (2011). Los índices de 'riqueza léxica' y la enseñanza de lenguas. En J. de Santiago Guervós, H. Bongaerts, J. J. Sanchez Iglesias & M. Seseña Gomez (Eds.), *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 15-28). Salamanca: Asociación para la Enseñanza del Español como Lengua Extranjera.

Lorán, R. & López Morales, H. (1983). *Nouveau calcul de l'indice de disponibilité*. MS.

Munita, F. (2017). Yo, lector: los relatos de vida lectora en la construcción del sujeto didáctico. *Cronía*, 17(13), 1-9.

Parodi, G. (2008). Géneros del discurso escrito: Hacia una concepción integral desde una perspectiva sociocognitiva. En G. Parodi (Ed.), *Géneros académicos y géneros profesionales: Accesos discursivos para saber y hacer* (pp. 17-38). Pontificia Universidad Católica de Valparaíso.

Parrado, M.^a M., Romero, M. F. & Trigo, E. (2018). La experiencia literaria en la formación de futuros docentes: el viaje iniciático de nuestras biografías en 10 hashtag. En V. Amar (Ed.), *Miradas y voces de futuros maestros* (pp. 59-86), Barcelona: Octaedro.

Rodríguez Bou, I. (1952a). *Recuento de vocabulario español* (Vol. I). Río Piedras, Puerto Rico: Consejo Superior de Enseñanza, Universidad de Puerto Rico, Publicaciones Pedagógicas.

Rodríguez Bou, I. (1952b). *Recuento de vocabulario español* (Vols. II, Partes I y II). Río Piedras, Puerto Rico: Consejo Superior de Enseñanza, Universidad de Puerto Rico, Publicaciones Pedagógicas.

Romero, M. F. & Trigo, E. (2018). Entre las creencias y la formación inicial de los estudiantes del máster de profesorado de secundaria: una mirada hacia la enseñanza de la literatura en las aulas. Profesorado. *Revista de Curriculum y formación del profesorado* 22(1), 89-112.

Santos Díaz, I. C. (2017a). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum*, 27, 122-139.

Santos Díaz, I. C. (2017b). Organización de las palabras en la mente en lengua materna y lengua extranjera (inglés y francés). *Pragmalingüística*, 25, 603-617.

Trigo, E., Romero, M. F. & Santos Díaz I. C. (2019). Aproximación al léxico gastronómico dialectal andaluz desde los repertorios de disponibilidad léxica para una propuesta didáctica. *Verba Hispánica*, 27, 115-130.

Notas

1. El programa LexBas 1.0 fue desarrollado en España ad hoc para la obtención del *Léxico Básico del Español de Chile* (Castillo Fadić, 2012b y 2020) gracias a las gestiones y financiamiento de Humberto López Morales, a quien agradecemos por haber autorizado el empleo del software en esta investigación.