



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE HUMANIDADES Y ARTE

**COMPETENCIA LÉXICA EN EL DISCURSO ESPECIALIZADO EN
ESTUDIANTES DE PERIODISMO**

POR

KARINA ANDREA FUENTES RIFFO

Tesis presentada en la Facultad de Humanidades y Arte de la Universidad de Concepción para optar al grado académico de Doctor en Lingüística.

Profesor Guía

Dr. Pedro Antonio Salcedo Lagos.

Noviembre de 2019

Concepción, Chile.

*Para Sergio Andrés,
Sergio Tomás y
Diego Andrés.
¡Qué expectantes esperábamos este momento!*



Agradecimientos

Agradecer es un acto de necesario reconocimiento hacia quienes nos acompañan y apoyan cuando decidimos iniciar este recorrido; recorrido que no es fácil, por cierto. No he conocido a nadie que alguna vez haya dicho que esta empresa lo sea.

En primer lugar, quiero agradecer a quienes está dedicado este trabajo: Sergio, Sergito y Dieguito. Al primero por su constante apoyo. Fue, sin duda, quien tuvo que soportar mis tribulaciones y encauzar mis constantes arrebatos y devaneos con mi trabajo y con el doctorado en sí. A los segundos, mis pequeños niños, mis disculpas por robarles egoístamente tiempo que les era propio: todas esas tardes de juego, días de entretención y descubrimiento fueron sacrificados por mí, por la "mamita estudiosa".

También debo agradecer a mi familia: mis padres, mis hermanos, mi tía, mis cuñadas y mis suegros; quienes sin dudarlo tomaron mi lugar en todos esos momentos en que tenía que estar presente y que, sin enjuiciar, perdonaron cada ausencia.

A mis amigos, que desinteresadamente me acompañaron en cada rato de *libertad tesística*. Sin importar el momento, la hora o la fecha, siempre acudieron a mi llamado de auxilio. Especialmente a Jeannette Valenzuela y Pamela Cea; a los *Tunas* y los *Llanos*, amigos maravillosos que el destino puso en mi camino.

Por cierto, a los integrantes del querido Departamento de Metodología de la Investigación e Informática Educativa, de la Facultad de Educación, quienes me recibieron con los brazos abiertos.

Una inicia este camino doctoral con muchas ilusiones. A veces éstas se mantienen intactas hasta el final, otras -como en mi caso-, peligran y tienden a apagarse.

Este trabajo no hubiera sido posible si el Dr. Gastón Salamanca no se hubiera cruzado en mi camino, ayudándome a rearmar lo que estaba trizado. Tampoco hubiera sido posible sin la valentía del Dr. Pedro Salcedo, que tomó el desafío de guiar lo que estaba torcido, defender mi trabajo y apoyar el inicio de mi carrera académica. Y, por cierto, tampoco hubiera sido posible sin la fuerza que emana, probablemente sin intención y sin saberlo, la Doctora Paola Alarcón. A todos ustedes mis más sinceras e infinitas gracias.

Mi competencia léxica no alcanza para expresar todo lo que significan para mí.

Karina.

TABLA DE CONTENIDOS

RESUMEN	1
CAPÍTULO 1: INTRODUCCIÓN Y OBJETO DE ESTUDIO.....	3
1. INTRODUCCIÓN	3
1.1. ¿ES EL PERIODISMO UN DISCURSO ESPECIALIZADO?	8
1.2. OBJETO DE ESTUDIO Y/O PROBLEMÁTICA A ABORDAR	11
1.2.3. OBJETIVO GENERAL	14
CAPÍTULO 2: MARCO TEÓRICO.....	16
2. CONCEPTOS FUNDAMENTALES.....	16
2.1. COMPETENCIA COMUNICATIVA	17
2.1.1. COMPETENCIA LÉXICA	17
2.1.2. VOCABLO Y VOCABULARIO	22
2.1.3. PERFIL LÉXICO.....	24
2.1.4. LEXICÓN MENTAL	26
2.2. DISCURSO ESPECIALIZADO	27
2.2.1. LA NOTICIA COMO DISCURSO ESPECIALIZADO.....	32
2.2.2. LA ESTRUCTURA DE PIRÁMIDE INVERTIDA, EL LEAD o ENTRADILLA.....	33
2.3. RIQUEZA LÉXICA.....	35
2.3.1. DIVERSIDAD Y DENSIDAD LÉXICA.....	40
2.3.1.1. ÍNDICE DE RIQUEZA LÉXICA.....	42
2.4. DISPONIBILIDAD LÉXICA.....	46
2.4.1. CENTROS DE INTERÉS	51
2.4.2. ÍNDICE DE DISPONIBILIDAD LÉXICA.....	52
2.4.3. ESTUDIOS DE DISPONIBILIDAD LÉXICA EN CHILE.....	54
2.5. REDES DE ASOCIACIÓN LÉXICA.....	56
CAPÍTULO 3: MARCO METODOLÓGICO	62
3.1. DISEÑO DE INVESTIGACIÓN	62
3.1.1. ESTRUCTURA	63

3.2. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN	64
3.2.1. OBJETIVO GENERAL	64
3.2.2. OBJETIVOS ESPECÍFICOS	64
3.2.3. PREGUNTAS DE INVESTIGACIÓN	65
3.2.4. HIPÓTESIS DE TRABAJO	65
3.3. PARTICIPANTES/MUESTRA.....	66
3.3.1. DETALLES DE LA MUESTRA EN CADA SUB-ESTUDIO	68
3.3.1.1. ESTUDIO DE DISPONIBILIDAD LÉXICA.....	68
3.3.1.2. ESTUDIOS DE RIQUEZA LÉXICA.....	69
3.3.1.3. ESTUDIO SOBRE RELACIONES LÉXICAS.....	70
3.4. PRUEBAS Y MEDICIONES	70
3.4.1. VARIABLES INDEPENDIENTES EN RIQUEZA LÉXICA	70
3.4.1.1. PROCEDIMIENTO PARA MEDIR LAS VARIABLES DIVERSIDAD LÉXICA Y DENSIDAD LÉXICA.....	71
3.4.1.2. CÁLCULO AUTOMÁTICO DE LA DIVERSIDAD LÉXICA Y LA DENSIDAD LÉXICA	75
3.4.1.3. CARACTERÍSTICAS DEL CORPUS PARA EL ANÁLISIS DE LA RIQUEZA LÉXICA	82
3.4.2. DISPONIBILIDAD LÉXICA.....	88
3.4.2.1. PROCEDIMIENTO PARA MEDIR LA DISPONIBILIDAD LÉXICA ..	90
3.4.2.2. INDICADORES EN ÍNDICE DE DISPONIBILIDAD LÉXICA.....	94
3.4.3. REDES DE ASOCIACIÓN	95
3.4.3.1. INDICADORES.....	98
CAPÍTULO 4: RESULTADOS Y ANÁLISIS DE DATOS	100
4.1. ANÁLISIS DESCRIPTIVOS DE DATOS	100
4.1.1. RIQUEZA LÉXICA	100
4.1.1.1. RESULTADOS GLOBALES DE RIQUEZA LÉXICA.....	106
4.1.2. DISPONIBILIDAD LÉXICA.....	127
4.1.3.1. RESULTADOS GENERALES.....	127
4.1.3.2. ANÁLISIS SEGÚN VARIABLES ESTUDIADAS	132
i) VARIABLE SEXO	132
4.1.3. REDES DE ASOCIACIÓN LÉXICA	156

4.1.3.1. CIENCIA Y TECNOLOGÍA.....	156
4.1.3.2. POLÍTICA.....	171
4.1.3.3. DEPORTE SEGÚN GRUPO	182
4.2. ANÁLISIS GENERAL SEGÚN OBJETIVOS ESPECÍFICOS	195
CAPÍTULO 5: CONCLUSIONES	200
5.1. LIMITACIONES DE LA INVESTIGACIÓN	208
5.2. PROYECCIONES DE LA INVESTIGACIÓN	210
REFERENCIAS BIBLIOGRÁFICAS.....	212
ANEXOS	220
ANEXO 01: LISTADO DE <i>STOP WORDS</i> UTILIZADAS PARA EL ANÁLISIS DE DENSIDAD LÉXICA.	220
ANEXO 02: PRUEBAS DE DISPONIBILIDAD LÉXICA.	223
<i>Prueba A: Periodistas</i>	223
<i>Prueba B: Estudiantes.</i>	225
ANEXO 03: LISTADOS DE RESULTADOS DE RIQUEZA Y DENSIDAD LÉXICA.	227
ANEXO 04: ÍNDICES DE DISPONIBILIDAD LÉXICA	236
1. <i>Ciencia y Tecnología</i>	236
2. <i>Política</i>	250
3. <i>Deporte</i>	274

ÍNDICE DE TABLAS

TABLA 1. CRUCE SEGÚN SEXO Y GRUPO.	67
TABLA 2. VARIABLES INDEPENDIENTES Y SUS VALORES.	68
TABLA 3. RESUMEN DE CASOS SEGÚN TEMA Y GRUPO.....	102
TABLA 4. RESUMEN DE CASOS SEGÚN SEXO Y CORPUS.	103
TABLA 5. RESUMEN DE CASOS SEGÚN TIPO DE INSTITUCIÓN Y CORPUS.	104
TABLA 6. RECUENTO SEGÚN EXTENSIÓN DE TEXTOS.	105
TABLA 7. ESTADÍGRAFOS DESCRIPTIVOS SOBRE DENSIDAD LÉXICA	106
TABLA 8. ESTADÍGRAFOS DESCRIPTIVOS DE DIVERSIDAD LÉXICA TOTAL.	107
TABLA 9. ESTADÍGRAFOS DESCRIPTIVOS DE DIVERSIDAD LÉXICA ALEATORIA.	108
TABLA 10. ESTADÍGRAFOS DESCRIPTIVOS DE DIVERSIDAD LÉXICA DEL LEAD.....	108
TABLA 11. RESUMEN DEL CONTRASTE DE HIPÓTESIS.	110
TABLA 12. ESTADÍGRAFOS DESCRIPTIVOS DENSIDAD LÉXICA SEGÚN GRUPO.	111
TABLA 13. ESTADÍGRAFOS DESCRIPTIVOS DIVERSIDAD LÉXICA SEGÚN GRUPO.	112
TABLA 14. ESTADÍGRAFOS DESCRIPTIVOS DIVERSIDAD LÉXICA ALEATORIA SEGÚN GRUPO.	113
TABLA 15. ESTADÍGRAFOS DESCRIPTIVOS DIVERSIDAD LÉXICA LEAD SEGÚN GRUPO.	114
TABLA 16. RESUMEN DE CONTRASTE DE HIPÓTESIS.	116
TABLA 17. ESTADÍGRAFOS DESCRIPTIVOS DENSIDAD LÉXICA.....	120
TABLA 18. CONTRASTE DE HIPÓTESIS TIPO DE ESTABLECIMIENTO EDUCACIONAL.	121
TABLA 19. ESTADÍSTICOS DESCRIPTIVOS DIVERSIDAD LÉXICA SEGÚN EXTENSIÓN.	123
TABLA 20. ESTADÍGRAFOS DESCRIPTIVOS DENSIDAD LÉXICA SEGÚN EXTENSIÓN.	125
TABLA 21. ORDEN SEGÚN ÍNDICE DE COHESIÓN.....	130
TABLA 22. LISTADO DE LAS CINCO PALABRAS MÁS FRECUENTES.	131
TABLA 23. CONTRASTE DE HIPÓTESIS.	136
TABLA 24. ÍNDICE DE COHESIÓN.	139
TABLA 25. ÍNDICE DE COHESIÓN EN LOS CENTROS DE INTERÉS SEGÚN GRUPO.	146
TABLA 26. DISTRIBUCIÓN DE LOS SUJETOS SEGÚN TIPO DE ESTABLECIMEINTO EDUCACIONAL DE EGRESO.	148

ÍNDICE DE FIGURAS

FIGURA 1. PIRÁMIDE INVERTIDA.	34
FIGURA 2. COMPLEJIDAD LÉXICA EN DIFERENTES NIVELES.	36
FIGURA 3. PROCESO PARA CALCULAR EXTENSIÓN DEL LEAD (ELABORACIÓN PROPIA).	72
FIGURA 4. MEDIDAS UTILIZADAS PARA ANALIZAR RIQUEZA LÉXICA.	74
FIGURA 5. PREGUNTAS QUE EL SCRIPT REALIZA AL USUARIO.	78
FIGURA 6. RESULTADOS QUE ENTREGA EL SCRIPT EN LA CONSOLA.	81
FIGURA 7. RESULTADO DE UN ARCHIVO DE SALIDA DEL SCRIPT (LOS VALORES SOLO CONSIDERAN DOS DECIMALES PARA FACILITAR SU PRPRESENTACIÓN).	82
FIGURA 8. ESQUEMA DE ANÁLISIS DE DIVERSIDAD Y DENSIDAD LÉXICA.	87
FIGURA 9. LISTADO DE PALABRAS DISPONIBLES EN PLANILLA EXCEL.	91
FIGURA 10. DEPURACIÓN DEL CORPUS DE LÉXICO DISPONIBLE	92
FIGURA 11. IMAGEN DEL CENTRO DE INTERÉS CIENCIA Y TECNOLOGÍA, PROVENIENTE DE DISPOGRAFO.	96
FIGURA 12. RED DEL CENTRO DE INTERÉS CIENCIA Y TECNOLOGÍA, EN GEPHI.	97
FIGURA 13. ESTADÍSTICO KRUSKAL WALLIS EN DISTRIBUCIÓN DE VOCABLOS SEGÚN GRUPO..	142
FIGURA 14. PRUEBA DE VARIANZAS ENTRE CENTROS DE INTERÉS SEGÚN DEPENDENCIA EDUCACIONAL.	151
FIGURA 15. RED DE CIENCIA Y TECNOLOGÍA SIN FILTRO.	156
FIGURA 16. RED DE ASOCIACIÓN CIENCIA Y TECNOLOGÍA CON RANGO DE GRADO 7.	158
FIGURA 17. RED CIENCIA Y TECNOLOGÍA DE PRIMER AÑO.	159
FIGURA 18. GRADO DE CENTRALIDAD EN CIENCIA Y TECNOLOGÍA.	160
FIGURA 19. RESULTADO DE LA RED CON NODOS Y ARISTAS SIGNIFICATIVAS.	162
FIGURA 20. RED CIENCIA Y TECNOLOGÍA 4 AÑO.	164
FIGURA 21. RED FILTRADA DE G2.	166
FIGURA 22. RED DE CIENCIA Y TECNOLOGÍA DE PROFESIONALES.	167
FIGURA 23. RED DE G3 FILTRADA, EN CIENCIA Y TECNOLOGÍA.	169
FIGURA 24. RED DE POLÍTICA CON TRES GRUPOS INCLUIDOS, AL 100%.	171
FIGURA 25. RED SE ASOCIACIÓN EN POLÍTICA, G1.	172
FIGURA 26. RED POLÍTICA CON FILTRO, G1.	175
FIGURA 27. RED ASOCIACIÓN SEMÁNTICA POLÍTICA, G2.	176
FIGURA 28. POLÍTICA G2, CON FILTROS.	178

FIGURA 29. RED EN TEMA POLÍTICA, G3.	179
FIGURA 30. RED FILTRADA DE POLÍTICA, G3.	181
FIGURA 31. RED TOTAL DE DEPORTE.	183
FIGURA 32. RED DEPORTE G1.	184
FIGURA 33. RED FILTRADA DEPORTE, G1.	187
FIGURA 34. RED DEPORTE G2.	188
FIGURA 35. RED FILTRADA DE DEPORTE, G2.	190
FIGURA 36. RED DEPORTE, G3.	192
FIGURA 37. RED FILTRADA EN DEPORTE, G3.	195



ÍNDICE DE GRÁFICOS

GRÁFICO 1. ORGANIZACIÓN DEL CORPUS.	84
GRÁFICO 2. DIFERENCIAS DE DENSIDAD LÉXICA SEGÚN GRUPO.	117
GRÁFICO 3. DIFERENCIAS DIVERSIDAD LÉXICA SEGÚN GRUPO.....	118
GRÁFICO 4. RELACIÓN ENTRE EXTENSIÓN Y DIVERSIDAD TOTAL.	124
GRÁFICO 5. INDICADORES DE RIQUEZA LÉXICA RELACIONADOS CON EXTENSIÓN DEL TEXTO Y TIPO DE ESTABLECIMIENTO.	126
GRÁFICO 6. TOTAL DE PALABRAS Y VOCABLOS OBTENIDOS EN CADA CENTRO DE INTERÉS.	128
GRÁFICO 7. GRÁFICO SEGÚN PROMEDIO DE RESPUESTAS.....	129
GRÁFICO 8. TOTAL DE PALABRAS SEGÚN SEXO.	133
GRÁFICO 9. TOTAL DE VOCABLOS SEGÚN SEXO	134
GRÁFICO 10. PRUEBA U DE MANN WHITMAN PARA MUESTRAS INDEPENDIENTES.	137
GRÁFICO 11. COMPARACIÓN DE PROMEDIOS DE GRUPO SEGÚN CENTRO DE INTERÉS.	138
GRÁFICO 12. ÍNDICE DE COHESIÓN.	139
GRÁFICO 13. TOTAL DE PALABRAS SEGÚN GRUPO.	140
GRÁFICO 14. PROMEDIO DE PALABRAS SEGÚN GRUPO.	143
GRÁFICO 15. NÚMERO DE VOCABLOS SEGÚN GRUPO.....	144
GRÁFICO 16. ÍNDICE DE COHESIÓN EN CENTROS DE INTERÉS SEGÚN GRUPO	147
GRÁFICO 17. TOTAL DE PALABRAS EN CADA CENTRO DE INTERÉS SEGÚN TIPO DE ESTABLECIMIENTO EDUCACIONAL.....	149
GRÁFICO 18. PROMEDIO DE PALABRAS SEGÚN DEPENDENCIA EDUCACIONAL.....	152
GRÁFICO 19. ÍNDICE DE COHESIÓN SEGÚN ESTABLECIMIENTO EDUCACIONAL.	153
GRÁFICO 20. DIFERENCIAS EN EL PROMEDIO DE DISPONIBILIDAD LÉXICA.	155
GRÁFICO 21. DISTRIBUCIÓN DE LAS COMUNIDADES SEGÚN MODULARIDAD.	161
GRÁFICO 22. MODULARIDAD DE CIENCIA Y TECNOLOGÍA G2.	165
GRÁFICO 23. COMUNIDADES EN CIENCIA Y TECNOLOGÍA G3.	168
GRÁFICO 24. COMUNIDADES EN POLÍTICA G1.	174
GRÁFICO 25. COMUNIDADES EN POLÍTICA G2.	177
GRÁFICO 26. DISTRIBUCIÓN DE LAS COMUNIDADES DE POLÍTICA EN G3.	180
GRÁFICO 27. COMUNIDADES EN DEPORTE, G1.	185
GRÁFICO 28. COMUNIDADES DE G2 DISTRIBUIDAS SEGÚN TALLA.....	189
GRÁFICO 29. DISTRIBUCIÓN DE LAS COMUNIDADES EN RED DEPORTE G3.	193



RESUMEN

El objetivo de esta tesis es describir la competencia léxica de estudiantes de periodismo a través de una aproximación que integra disponibilidad léxica, riqueza léxica - entendida bajo los conceptos de diversidad y densidad léxica-, y la asociación léxica. Los enfoques anteriores, en el contexto de este trabajo, se aplican al discurso profesional especializado, que en este caso se materializa en las noticias escritas por estudiantes.

Para lograr una descripción en profundidad de lo que se ha delimitado como competencia léxica, se trabajó con las variables sociocultural, sexo y nivel educacional, utilizando un diseño metodológico mixto. Por otra parte, se incluyó un grupo de periodistas profesionales, con el fin de establecer el nivel que los estudiantes requieren alcanzar en un medio de comunicación escrito.

Los resultados de la investigación señalan que los estudiantes cuentan con un léxico amplio y variado, el cual se expresa en los resultados de riqueza léxica. Sin embargo, dicho léxico no cumple con las expectativas de densidad léxica que presentan los textos de profesionales, evidenciando una problemática que se debe solucionar ya que este indicador se relaciona con la capacidad informativa que el profesional debe entregar en su noticia.

En el ámbito de la disponibilidad léxica, los estudiantes evocan un alto número de vocablos, superior al de los profesionales. Sin embargo, el léxico seleccionado corresponde en mayor grado a palabras poco técnicas y con un alto

nivel valorativo, a diferencia del que presentan los profesionales que se enfocan en palabras concretas de los centros de interés estudiados. Este hecho se comprueba en el estudio de redes semánticas, donde se identifican conceptos claves y las asociaciones más importantes del léxico mental de cada grupo.

Este trabajo aporta, desde una aproximación léxica, una metodología que sirve para identificar y realizar un seguimiento a la evolución de la competencia léxica de las cohortes de estudiantes.



CAPÍTULO 1: INTRODUCCIÓN Y OBJETO DE ESTUDIO

1. INTRODUCCIÓN

“Las decisiones que el periodista toma en el plano léxico de la lengua son pertinentes para la conservación y desarrollo del buen escribir en español” (Romero, 2003). Con lo anterior se entiende que los desafíos para el periodista no sólo se circunscriben a su responsabilidad profesional de informar a la comunidad sobre hechos relevantes -que la autora relaciona con el aspecto práctico de la profesión-, sino que también están relacionados -quíéralo o no- con una labor instructiva y de preservación de la lengua, debido a que, en palabras de Romero (2003) “Los medios de comunicación ... coadyuvan a la mejora de la competencia comunicativa del hablante, del lector”.

La autora indica, además, que el periodista debe conocer su lengua en profundidad para emplearla y enriquecerla, haciendo referencia a la actuación lingüística del profesional y distinguiendo entre *empleo* y *uso* del lenguaje; la autora explica que el empleo del lenguaje es el “indicador de la capacidad del hablante de servirse del código para comunicar con eficacia” (Romero, 2003).

La importancia de la competencia léxica –entendida en primera instancia, según Obegui (2005), como el conocimiento del vocabulario y la capacidad para utilizarlo correctamente-, en el quehacer de los periodistas, es fundamental debido a que su función social es la de ser un comunicador cualificado, a quien se

le exige emplear la lengua de manera eficiente para cumplir con su trabajo. Esta idea es compartida por Martínez Albertos (1989) quien señala que el lenguaje periodístico debe utilizar una lengua fijada, normalizada y estándar. “Una lengua donde exista ya un elenco preciso de modos de narrar, de describir y de comentar” (Martínez Albertos, 1989).

El lenguaje periodístico utiliza como materia prima la noticia. Es noticia aquel hecho “verdadero, actual o inédito, de interés general, que unos expertos cualificados comunican a un número considerable de personas a través de los llamados medios de comunicación de masas: prensa, radio y televisión” (Martínez Albertos, 1989). A estos se les suman, en los últimos 20 años, los medios digitales que en la actualidad cumplen con las características tradicionales de la noticia, agregándole -además- nuevas oportunidades narrativas a través de sus propias características como inmediatez, autonomía, elementos multimediales, entre otros factores.

Por otra parte, Martín Vivaldi (1993) señala que la característica más destacada del estilo periodístico es “la variedad de vocabulario, de los giros y de las construcciones gramaticales”. A juicio del autor, estos rasgos son consecuencia de la necesidad de exactitud y de precisión requeridas por la profesión; de utilizar la palabra exacta en el sitio preciso y en el momento oportuno: “El periodista es un especialista -debe serlo- de la palabra propia y ajustada” (Martín Vivaldi, 1993). En este punto aparece una pregunta natural: ¿cuál y qué tan variado es el vocabulario de los periodistas? ¿Cómo se prepara a un profesional con estas condiciones?

Al respecto, surgen algunas propuestas metodológicas que pueden entregar luces para responder a dichas preguntas. Una muy utilizada y de gran desarrollo científico para el español en los últimos cincuenta años, es el estudio de la Disponibilidad Léxica, campo de investigación dentro de la lingüística aplicada que tiene como objetivo la "recogida y posterior análisis del léxico disponible de una determinada comunidad de habla" (Saralegui y Taberner, 2008, p.745). Natividad Hernández (2004) define disponibilidad léxica como "la línea de investigación encuadrada en la lingüística aplicada que se encarga de recoger, editar y analizar el caudal léxico que puede ser actualizado por los hablantes de una lengua en torno a un tema o contexto dado" (p. 21). La autora realiza hincapié al hablar de "tema", y no de "situación", debido a que la producción del vocabulario que se considera como disponible en este tipo de estudios aparece generalmente en condiciones experimentales muy concretas.

El léxico disponible, en tanto, es el conjunto de palabras que los hablantes tienen en el lexicón mental y cuyo uso está condicionado por el tema concreto de la comunicación. "Lo que se pretende con un estudio de disponibilidad léxica es descubrir qué palabras sería capaz de usar un hablante en determinadas situaciones comunicativas" (Saralegui y Taberner, 2008, p.745).

Por otra parte, bajo las premisas sobre el lenguaje y los géneros periodísticos, expuestos por Martínez Albertos (1989) y Martín Vivaldi (1993), subyace la coincidencia sobre la calidad de la escritura profesional como un elemento central. Al respecto, Humberto López Morales (2002) señala que esta calidad está integrada por una serie de factores, entre los que se puede encontrar

riqueza léxica, madurez sintáctica y esquemas de cohesión y coherencia discursiva; donde la riqueza léxica en un elemento donde converge la amplitud y variedad del vocabulario y está muy apoyada en la disponibilidad léxica del hablante (López Morales, 2002).

En la actualidad existe poca evidencia empírica que dé cuenta sobre la riqueza léxica de los textos periodísticos. En 2016 la Fundación Español Urgente¹, Fundéu BBVA, publicó los resultados del Proyecto Aracné. Este estudio tuvo por objetivo realizar una observación y medición sobre cómo ha cambiado “el lenguaje de la prensa española en los últimos cien años”, atendiendo especialmente la riqueza lingüística, de noticias publicadas en medios de comunicación escrita desde 1914 hasta 2014.

Para abordar la riqueza lingüística los investigadores decidieron trabajar desde el parámetro de la variación léxica. Para ello orientaron su metodología a la tradición léxico-estadística, tomando como unidades de trabajo palabras y vocablos y, posteriormente, estableciendo una relación entre ambas mediante el cálculo *type/token ratio*².

A grandes rasgos, los resultados mostraron que los valores de la riqueza lingüística de los medios de comunicación españoles se mantienen en general muy estables, con diferencias mínimas entre épocas. Los valores además están

¹ La Fundación del Español Urgente, Fundéu BBVA, es una institución sin fines de lucro que tiene como principal objetivo impulsar el buen uso del español en los medios de comunicación. Nacida en el año 2005 fruto de un acuerdo entre la Agencia Efe y el banco BBVA, trabaja asesorada por la Real Academia Española. www.fundeu.es

² En capítulos posteriores de esta tesis se analizará en profundidad esta medida.

muy concentrados, las diferencias respecto a la media nunca superan el 10 % y se mantienen sorprendentemente uniformes en todas las épocas y para todas las longitudes³.

Un estudio anterior fue realizado a fines de la década de 1980. Raúl Ávila del Centro de Estudios Lingüísticos y Literarios de El Colegio de México, junto a un grupo de lingüistas de España, Argentina, Bolivia, Colombia, Costa Rica, Chile, República Dominicana y Estados Unidos se plantearon estudiar el español utilizado por los medios de comunicación de Hispanoamérica y España⁴. Max Echeverría (1997) fue el lingüista que estudió en Chile la riqueza léxica en las noticias, emitidas por el canal TVN, y en el deporte en vivo transmitido por dos radioemisoras nacionales. En total, se analizaron 48 unidades textuales, de 1200 palabras cada una, lo que conforma un total aproximado de 72 000 palabras gráficas (también llamadas por los autores *running words*). Los resultados de esta muestra señalan que en los informativos aparecen 4259 tipos (formas o palabras distintas) y 1965 voces distintas en deportes (Echeverría, 1997).

Los valores de riqueza léxica de los chilenos, a juicio de Echeverría, fueron extraordinariamente altos; cerca de 67 puntos en términos de los índices de Ávila (1994). "El mismo autor ha expresado que el rango superior en sus mediciones va de 66,5 a 70,5 tipos por cada 100 palabras. Sorprende notar la similitud de estos valores de riqueza léxica con los encontrados en noticias por Raúl Ávila en

³ Ver informe completo en el sitio web <http://www.fundeu.es/ aracne/index.html>, revisado el 4 de enero de 2018.

⁴ Proyecto denominado Difusión internacional del español por radio, televisión y prensa: Unidad y diversidad de la Lengua (DIES-RTP).

México y que promediaron 66,5" (Echeverría, 1997). Hay que señalar que este índice está basado en unidades textuales recopiladas a partir de discursos orales, no escritas como el ejemplo anterior.

1.1. ¿ES EL PERIODISMO UN DISCURSO ESPECIALIZADO?

Martín Vivaldi (1993) se cuestiona: "¿Sería admisible, científicamente, la teoría de una posible 'redacción estructural' como técnica expresiva propia del periodismo? La respuesta debe ser afirmativa. El periodismo es un medio específico de comunicación y expresión del pensamiento [...] existe un modo de hacer periodístico, claramente diferenciable de otros". Por lo tanto, se puede decir, cuando hablamos de textos del género periodístico, que estamos dentro de un tipo de discurso especializado. Investigaciones chilenas que toman como objeto de interés estos tipos de estudio son los de Parodi (2005a, 2005b, 2006) con los que se resalta el interés creciente y la vitalidad por la investigación sobre los discursos disciplinarios –Parodi (2005a) adopta la definición de Hyland (2000) sobre discursos especializados–, argumentando sobre el impacto que este ámbito disciplinar tiene sobre la variación lingüística.

De manera complementaria, la formación universitaria de los periodistas es la etapa en la que el estudiante se ve inmerso y se perfecciona en los conceptos y estrategias fundamentales para su futuro ejercicio. En el ámbito de la competencia léxica, especialmente en la elaboración de noticias escritas, es posible dar cuenta de la preocupación por parte de las universidades, cuyas escuelas de periodismo contemplan dentro de sus mallas curriculares un número

no despreciable de cursos relacionados con redacción periodística, análisis sobre actualidad, talleres de prensa y producción de textos, junto a cátedras sobre comunicación en distintos niveles, que completan la formación del periodista.

Sin embargo, la competencia léxica, por sus características de desarrollo, es una de las más complejas de abordar a través del currículo y de los procesos de enseñanza aprendizaje. Sería difícil pensar, por ejemplo, que se dictara una asignatura de vocabulario especializado en política, o un curso para estudiar los términos asociados al deporte. Trabajos como los de Baralo (2005), describen el conocimiento de un ítem léxico como un proceso complejo y gradual en el que se aprende no sólo la forma y el significado, "sino también una intrincada red de relaciones formales y semánticas entre este ítem y otras palabras y morfemas que constituyen subsistemas de diferentes niveles" (Baralo, 2005).

Baralo (2005) aporta, desde sus estudios de enseñanza de lenguas, algunos fundamentos psicolingüísticos, semánticos y de uso para la construcción de la competencia léxica: "El objetivo se centra en describir las redes de relaciones que establecen entre sí las palabras". En este contexto, el estudio de disponibilidad léxica es una puerta de entrada para indagar en un aspecto más complejo y que tiene que ver con la representación mental de las palabras; una representación que se asemeja a una red y a sus conexiones.

Por otra parte, pero en el mismo ambiente universitario, las pautas de corrección de los textos periodísticos abordan fundamentalmente errores ortográficos, redacción y contenido. Sin embargo, un aspecto tan interesante

como es la riqueza léxica de los textos escritos por los estudiantes y su posible evolución a través de los años formativos, no suele ser abordado por los profesores a menos que sean correcciones estilísticas propias de los géneros periodísticos.

Es probable que la naturaleza de los estudios lingüísticos en este ámbito – que se basan principalmente en el establecimiento de distintos índices cuantitativos–, no sea una prioridad para los académicos de estas carreras. Sin embargo, una nueva perspectiva podría abrirse con el desarrollo de investigaciones de corpus, las cuales, gracias a la lingüística computacional, han abierto interesantes líneas de estudios. Un corpus compuesto por noticias periodísticas podría ser un objeto importante de análisis, sobre todo para conocer el perfil léxico de los profesionales que se están formando para la sociedad.

En este punto surgen más interrogantes: ¿cuál es el índice de riqueza léxica de los textos especializados, es decir con un formato y una estructura de noticia, elaborados por estudiantes? ¿Tiene alguna relación este índice con la riqueza de los textos que se publican en los medios de comunicación escritos? ¿Existirá un déficit entre la riqueza de los textos que escriben los estudiantes en comparación con los que escriben los profesionales? ¿Podríamos establecer algunas variables como predictoras del nivel de riqueza léxica en los estudiantes? Éstas son algunas de las interrogantes a las que esta investigación espera aportar.

1.2. OBJETO DE ESTUDIO Y/O PROBLEMÁTICA A ABORDAR

Como podríamos suponer, el periodista debe superar al momento de elaborar una noticia escrita, una serie de dificultades. El profesional debe cumplir con ciertos requisitos y cualidades que han sido expuestos en diversos manuales de redacción periodística: claridad, concisión, densidad, exactitud, precisión, sencillez, originalidad, brevedad, variedad, corrección y propiedad, entre otros (Martín Vivaldi, 1993; Leñero y Marín, 1986; Martínez Albertos, 1989).

Esta investigación aborda la problemática de la competencia léxica, entendida en términos de disponibilidad, riqueza y asociación léxica que presenta un grupo de alumnos de periodismo. Esta competencia se analiza enmarcada en el género discursivo periodístico, mediante textos de noticias escritas relacionadas con tres temáticas –más adelante llamados “centros de interés”-, ciencia y tecnología, deporte y política, debido a que son dominios con los que deben estar familiarizados los estudiantes de periodismo. La investigación espera contribuir en la entrega de un modelo o metodología que sirva para describir el perfil léxico de un grupo determinado, así como la de un sujeto. Con ello se espera aportar con un indicador concreto –en el ámbito de la competencia léxica-, sobre la formación de los estudiantes de esta carrera.

Como ya se ha expresado, existe poca investigación, por ejemplo, sobre la calidad de la escritura en términos de riqueza léxica, de las noticias elaboradas en los medios de comunicación. Es más, en la actualidad, no existe evidencia sobre el real estado de la competencia léxica de los estudiantes de periodismo y

si son capaces o no de cumplir con el nivel que se requiere en los medios. Si bien se han realizado investigaciones sobre competencia léxica en muestras de estudiantes de enseñanza básica y enseñanza media en distintos países (Yu, 2010; Reyes, 2010; Obegi, 2005; Echeverría, Herrera, Moreno y Pradenas, 1987; Echeverría, 1993; Hernández 2004; Reyes, 2010;), llama la atención que no se haya investigado sobre la riqueza de los textos elaborados por alumnos de las carreras de periodismo, quienes por definición y formación deberían tener un alto grado de habilidad en la escritura.

Los resultados que un modelo o metodología de este tipo aporte, podrían utilizarse en la adecuación de los programas de estudios de dichas carreras o en enfocar esfuerzos especiales en mejorar los índices de riqueza léxica de los estudiantes. Otro aspecto en el que puede contribuir este estudio es en comparar y medir la variación de los índices entre un curso y otro; o, incluso, una comparación entre los índices de riqueza léxica de los textos producidos por los estudiantes y de los textos elaborados por profesionales, como una forma de conocer la expectativa de competencia léxica a la que deberían llegar los futuros periodistas.

Relacionado con lo anterior, se puede mencionar que tampoco se han realizado estudios para conocer cuánto y cuál es el vocabulario disponible en estudiantes de periodismo. Tener acceso a información como ésta podría servir, por ejemplo, para ayudar en la adquisición progresiva de vocabulario especializado durante los años de estudio de los alumnos. La importancia de este aspecto radica en solucionar posturas que han señalado a la pobreza léxica como

uno de los vicios más comunes entre los periodistas. La relevancia de abordar un estudio desde la perspectiva del vocabulario de un discurso especializado se explica desde el entendido de que las palabras son las unidades básicas en la construcción del lenguaje, unidades de significado desde donde surgen estructuras mayores como las oraciones, párrafos y los propios textos completos que son formados (Read, 2000, p. 3).

Por otra parte, nuevas investigaciones en el ámbito de la disponibilidad léxica, que incluyen las relaciones semánticas de las palabras disponibles (Hernández, Izura, y Ellis, 2006; Echeverría, Vargas, Urzúa y Ferreira, 2008; Ferreira y Echeverría, 2010), están entregando evidencia sobre relaciones más profundas entre el léxico y la conformación del llamado lexicón mental. Ferreira y Echeverría (2010) señalan la relevancia que tiene para la disponibilidad léxica conocer la naturaleza del lexicón mental, ya que "cada vez que un sujeto evoca una palabra, necesariamente debe recorrer un camino que lo lleva a esta estructura mental para activar, seleccionar y, finalmente, producir un ítem léxico".

Por lo tanto, la presente propuesta de investigación pretende contribuir con evidencia empírica sobre la competencia léxica y el perfil léxico de los futuros periodistas enmarcados en un género discursivo especializado del periodismo.

En definitiva, se aborda la problemática de la competencia léxica en el discurso especializado del periodismo desde tres centros de interés específicos:

ciencia y tecnología, deporte y política, analizando el objeto de estudio desde tres ámbitos:

1) Los estudios de disponibilidad léxica, con el objeto de determinar cuál es el léxico de los estudiantes de periodismo y cuál es el grado de disponibilidad en los centros de interés ciencia y tecnología, política y deporte. Esto, en contraste con la disponibilidad en los mismos centros de interés, de periodistas profesionales en ejercicio, quienes deberían entregar una visión más completa, por su conocimiento y uso de la lengua, de la cantidad del léxico necesario o presente en torno a dichas temáticas.

2) Los estudios de riqueza léxica, mediante los índices de diversidad léxica y densidad léxica, con el propósito de identificar los índices de riqueza de textos elaborados tanto por estudiantes como por profesionales y realizar comparaciones entre los grupos y las temáticas.

3) Análisis descriptivo de las relaciones léxicas obtenidas a partir de las pruebas de disponibilidad léxica de los estudiantes y de los profesionales, con la finalidad de entregar nuevos antecedentes empíricos sobre la naturaleza del lexicón mental en cada grupo.

1.2.3. OBJETIVO GENERAL

El objetivo general de esta investigación es describir la competencia léxica de estudiantes de periodismo en el discurso especializado escrito, a través de un modelo de estudio basado en disponibilidad, riqueza y asociación léxica.

Para lo anterior se han planteado una serie de objetivos específicos que apuntan a explicar cada componente del objetivo general, a su vez estos objetivos específicos se encuentran en concordancia con las preguntas de investigación planteadas en el Marco Metodológico.

1. Medir la disponibilidad léxica de la muestra en los centros de interés específicos: ciencia y tecnología, deporte y política.
2. Comparar la disponibilidad léxica entre estudiantes y profesionales.
3. Evaluar la riqueza léxica de los textos producidos por la muestra, para compararla entre los grupos de estudiantes.
4. Comparar los índices de riqueza léxica de estudiantes y profesionales.
5. Identificar las redes de asociación léxica de la muestra y realizar comparaciones entre los grupos de sujetos y los centros de interés relacionados con las tres temáticas indicadas.
6. Comparar el índice de riqueza léxica de la muestra con el índice de riqueza léxica de una muestra de un corpus de noticias realizadas por profesionales.

CAPÍTULO 2: MARCO TEÓRICO

2. CONCEPTOS FUNDAMENTALES

En este capítulo se articulan los principales fundamentos teóricos que dan sustento a la presente investigación. En el punto uno se encuentran las definiciones fundamentales sobre competencia léxica, junto a dos de sus características relacionadas directamente con el propósito de este trabajo: el perfil léxico y el lexicón mental.

El punto dos aborda la conceptualización del discurso especializado, aspecto central de la investigación debido a que delimita la competencia léxica a un marco específico lo que, entre otras características, aporta novedad a este estudio debido a las pocas investigaciones actuales que estudian la competencia léxica en este ámbito.

En el apartado tres se presentan los aspectos teóricos relacionados con la riqueza léxica, sus definiciones, aproximaciones y problemáticas a través de dos indicadores utilizados en esta investigación: diversidad léxica y densidad léxica

El punto cuatro se encuentra dedicado al ámbito de la disponibilidad léxica, su definición teórica, los principales exponentes de esta corriente teórica-metodológica, así como diversos estudios y elementos que componen esta disciplina.

Finalmente, el punto cinco presenta el componente relacionado con las redes de asociación léxica, mediante el cual la investigación busca integrar de

manera holística aspectos cuantitativos y cualitativos relacionados con la competencia léxica de estudiantes en el discurso especializado escrito.

Como ya se ha enunciado en la introducción, este trabajo se enfoca desde el ámbito disciplinar de la lingüística aplicada, específicamente en el estudio de la competencia léxica enmarcada en el discurso especializado. Si bien, en el ámbito de la competencia léxica se desarrolla un número significativo de estudios ligados a la enseñanza de segundas lenguas, investigaciones en los últimos años, desarrolladas desde esta disciplina indican que la lingüística aplicada ha retomado la idea de patrones lingüísticos empleados en la enseñanza de la lengua materna, enriqueciéndose al conjugar saberes de distintas ciencias y corrientes teóricas, aplicándolos para la solución de un problema lingüístico concreto.

2.1. COMPETENCIA COMUNICATIVA

2.1.1. COMPETENCIA LÉXICA

Diversos autores han intentado abordar la complejidad que supone definir el concepto de competencia léxica. Algunos, como Lahuerta y Pujol (1996), la definen como la capacidad para relacionar formas con significados, utilizándolas de manera adecuada; otros como Marconi (2000) nuevamente hacen uso del término *capacidad*, esta vez referida, por una parte, a la habilidad de acceder a una red de conexiones entre una palabra con otras palabras y expresiones lingüísticas y, por otra parte, ser capaz de proyectar las piezas léxicas en el mundo real al poder, en primer lugar **nombrar**, seleccionado una palabra adecuada en respuesta a un objeto o circunstancia; y, en segundo lugar, poder

aplicar una palabra, seleccionando el objeto o la circunstancia adecuada, en respuesta una palabra dada (Marconi, 2000, p. 14).

Desde una perspectiva didáctica, una concepción amplia de la competencia léxica incluye aspectos como grado de madurez y dominio léxico general de un hablante, los cuales se ponen de manifiesto en su producción lingüística, es decir, en su actuación léxica, ya sea oral o escrita (Núñez y del Moral, 2005.p. 94-95). El nivel de competencia, por lo tanto, podría ser conocido por medio del estudio de esta actuación léxica, utilizando procedimientos estadísticos (Núñez y del Moral, 2010).

El Marco Común Europeo de Referencia para las Lenguas, MCER (2001), desde su enfoque de segundas lenguas, define la competencia léxica como el “conocimiento del vocabulario de una lengua y la capacidad para utilizarlo”. Lo anterior puede ser ligado a lo que propone Baralo (2005), quien señala que la competencia léxica es la base del desarrollo de las habilidades y destrezas comunicativas, aspecto al que López-Mezquita (2007) se adscribe al señalar que el aspecto léxico se consideró como una parte de la competencia comunicativa, bajo la propuesta de Hymes, con su trabajo *On communicative competence*, el cual señala que un hablante, además de contar con la capacidad de formar frases gramaticales correctas, lo que se denomina “competencia gramatical”, necesita dominar la habilidad para utilizar la lengua de forma apropiada de acuerdo a los diferentes contextos sociales. En otras palabras, el hablante no solo debe tener una competencia gramatical para comunicarse de manera eficaz, sino que

necesita, además, conocer cómo usan el lenguaje los miembros de una comunidad lingüística para realizar sus propósitos (López-Mezquita, 2007, p. 30).

Jiménez-Catalán (2002), por su parte, aporta con la propuesta de que la competencia léxica cuenta con "la capacidad de reconocer, aprender, recuperar y relacionar distintas palabras a nivel oral y escrito". A partir de estas premisas se puede señalar que los niveles de competencia léxica atienden, por una parte, a la riqueza de vocabulario y, por otra, al grado de dominio que caracterizan el perfil léxico de un sujeto o de un grupo y al entendimiento de que existe un contexto social que debe ser considerado para lograr el éxito en la comunicación.

Siguiendo esta línea, se abre la posibilidad de abordar el concepto de *categorización* como una forma o mecanismo de organizar la información obtenida a partir del conocimiento y entendimiento de la realidad. Esta idea es tomada por Gómez (2002) quien señala que esta categorización se trata de un proceso mental de ordenamiento cuyo producto son las categorías cognitivas, caracterizándolas como "Conceptos mentales almacenados en nuestro cerebro, que constituyen lo que se denomina *lexicón mental*" (Gómez, 2002).

Bajo la hipótesis de categorización han surgido nuevas investigaciones que buscan entregar otras luces a la pregunta sobre ¿cómo se organiza el lexicón mental? Iniciativas como las de Hernández, Izura y Ellis (2006), Urzúa, Sáez y Echeverría (2006); Ávila y Villena (2010), Ferreira y Echeverría (2010) proponen mostrar las relaciones semánticas de las palabras obtenidas en las pruebas realizadas para las investigaciones de disponibilidad léxica. Es decir, los

investigadores vinculan los estudios de disponibilidad léxica con la relación semántica de las palabras, lo que ha llevado a algunos de ellos a tomar perspectivas e interpretaciones de los resultados con un carácter psicolingüístico y cognitivo.

Para Gómez (2002) el léxico específico, enmarcado en la competencia léxica, designa las voces utilizadas en situaciones de comunicación que implican la transmisión relevante de un campo de conocimiento y de experiencia particular. Este vocabulario va siempre acompañado de una exigencia de exactitud y precisión, lo que es característico de los campos científicos, técnicos y profesionales a los que el léxico específico hace referencia y de los que se nutre. Lo que Gómez propone mediante estas ideas es que la competencia léxica, enmarcada en un discurso especializado, tiene características propias y relevantes, entre las cuales se puede ejemplificar el léxico específico. Éste es uno de los fundamentos que investigadores abocados al estudio del discurso especializado (Cabré, 1993; Parodi, 2005b) utilizan para definirlo como un objeto de estudio científico.

La competencia léxica está constituida por un conjunto de relaciones. Así, se pueden señalar asociaciones fonológicas, morfológicas, léxicas, discursivas, sociocultural, entre otras (Baralo, 2007). Por otra parte, Núñez y del Moral (2010) desde una perspectiva didáctica, señalan que la competencia léxica utiliza como indicadores de la actuación léxica la amplitud y dominio de vocabulario, y la riqueza léxica. Por su parte, Marconi (2000) hace referencia a la tradición de la semántica filosófica para aproximarse a la conceptualización de la competencia

léxica, indicando que dicha línea teórica no se ha hecho cargo del significado léxico, sino que se ha centrado en otros aspectos del significado, aquellos que el autor define como *estructurales* o *composicionales* (Marconi, 2000, p. 19).

Para Marconi, un hablante competente sabe más que lo expresado por los postulados de la significación y lo ejemplifica con la palabra *cuchara*. Es difícil que alguien desconozca lo que una cuchara es: una persona la puede describir, puede explicar cuáles son sus usos, indicar dónde se adquieren y dónde se almacenan; incluso, la mayoría, puede realizar un dibujo de una cuchara que sería reconocible por muchas otras personas. En otras palabras, "Conocer una palabra, ser competente en su uso, es, o incluye, la posesión de todas estas distintas destrezas" (Marconi, 2000, p. 79).

Volviendo a la idea inicial, que dota a la competencia léxica con la capacidad de proyectar piezas léxicas y donde las destrezas son *nombrar*, al ser capaz de seleccionar una palabra adecuada y *aplicar*, al seleccionar un objeto o circunstancia adecuada como respuesta a una palabra dada, Marconi postula que ambas son capacidades independientes, donde la primera subyace a nuestras actuaciones *inferenciales* y la segunda -aplicar- está vinculada al ámbito *referencial*.

El aspecto no referencial de la competencia se ha relacionado con el ámbito del conocimiento o del saber, que en un plano más amplio tiene que ver con la capacidad de "habérselas con una red de conexiones entre las palabras, que subyace a actuaciones como las de inferencia semántica, paráfrasis, definición,

recuperación de una palabra a partir de su definición” (Marconi, 2000), entre otros aspectos, lo que Marconi viene a definir como el “aspecto *inferencial* de la competencia semántica léxica”, el cual sin mencionarlo expresamente estaría relacionado con la idea de la existencia de un lexicón mental, la red o lugar desde donde se recuperan las palabras necesarias; aun así, el autor señala que la capacidad de sacar inferencias semánticas es esencial para la competencia léxica (Marconi, 2000, p. 81). Por su parte, el lado *referencial* de la competencia la define como la capacidad de proyectar elementos léxicos en el mundo real.

2.1.2. VOCABLO Y VOCABULARIO

Las diferencias teóricas entre el significado de léxico y vocabulario, elementos fundamentales para el estudio de la competencia léxica, radican en que mientras el primero se considera como el conjunto de términos lexicales de un individuo, el segundo –el vocabulario- tiene que ver con los términos lexicales que un sujeto emplea, siendo, por lo tanto, más reducido que el léxico de un individuo. Ambos, eso sí, se caracterizan por tener valores cuantitativos, un número determinado de piezas que lo componen, “pero mientras en el vocabulario puede obtenerse la frecuencia de empleo de cada unidad, en el léxico podrá calcularse la probabilidad de aparición” (Núñez y del Moral, 2010).

Müller (1973) considera a la estadística lexical una parte importante de la lingüística cuantitativa y clasifica bajo esta idea todas las operaciones que toman como unidades elementales las *palabras*. Sin embargo, en un sentido más

restringido, la estadística lexical comienza con la introducción de la noción de *vocablo* (Müller, 1973, p. 225).

En el desarrollo teórico de la aproximación léxica, autores han llegado a definir de distintas formas las voces de *vocabulario*, *vocablo* y *léxico* (Müller, 1973; Marconi, 2000; Read, 2011; Núñez y del Moral, 2010). Por una parte, el *vocabulario* se entiende como el conjunto de términos lexicales que emplea un individuo y que se pone de manifiesto en el conjunto de textos orales y escritos que produce, siendo más reducido que el caudal léxico que conoce (Nuñez y del Moral, 2010); en este sentido, por lo tanto, *léxico* se define como “el conjunto de términos lexicales que un individuo posee” (Nuñez y del Moral, 2010).

Müller (1973) señala que para evitar confusiones el término *palabra* es reservado para las unidades elementales, “perfectamente distinguidas por la tipografía y la escritura, que constituyen el texto, es decir, a las ocurrencias de un vocablo cualquiera. El *vocablo* es una unidad del léxico, la *palabra* una unidad de texto; leemos una palabra en el texto, pero en el diccionario encontramos un vocablo” (Müller, 1973:226).

Aun así, trabajos como el de López-Mezquita, considera los términos léxico y vocabulario de forma intercambiable, como sinónimos, sin ninguna diferencia apreciable (López-Mezquita, 2007). En su tesis, la autora se enfoca en la escasez de pruebas de vocabulario; a su juicio, conocer los niveles de competencia léxica es desde donde se pueden detectar las carencias, realizar diagnósticos y sugerir contenidos léxicos que tiendan a mejorar los vocabularios de los estudiantes. Hay

que considerar, eso sí, que dicho trabajo se enfoca en la enseñanza de una segunda lengua, a diferencia de la presente investigación que se realiza con el español como lengua materna.

2.1.3. PERFIL LÉXICO

Se ha hablado sobre un perfil léxico y se han esbozado sus características, pero no se ha llegado a una delimitación de su significado. Para entender el concepto de *perfil léxico* se recurrirá, en primer lugar, a la definición de sus constituyentes. Según la vigésimo segunda edición del Diccionario de la RAE (2014), *perfil* significa "Conjunto de rasgos peculiares que caracterizan a alguien o algo". Por su parte, *léxico* cuenta con cinco definiciones, entre estas se incluye a "vocabulario, conjunto de las palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico dado, etc.". Entendiéndolo así, se podría decir que el *perfil léxico* está compuesto por los rasgos peculiares que caracterizan el vocabulario de un individuo o de un grupo en un contexto determinado.

Sin embargo, al hablar de *perfil léxico* como un constructo teórico, la conceptualización se complejiza. Echeverría (1982) señala que "un perfil nos muestra las áreas fuertes y deficitarias de un sujeto, y además nos sugiere posibles vías de desarrollo". Dicho autor caracteriza los componentes de lo que, a su juicio, debería contener un perfil en el léxico general: "Para mí más interesante, es necesario conocer el rango de léxico pasivo dominado por el sujeto. Para ello, analizamos tres aspectos: a) nivel general de vocabulario

pasivo, b) nivel general de vocabulario de orientación científica, y c) perfil léxico especializado” (Echeverría, 1982).

Por otra parte, Marconi (2000) se refiere a la competencia semántica, indicando que la mayor parte de los hablantes tiene conocimiento y competencia en el uso de un importante número de palabras, pudiendo llevar adelante diversas actuaciones frente a una palabra determinada, tal como se explicó con el ejemplo de cuchara, expuesto anteriormente. Sin embargo, existen diferencias en el alcance del uso y conocimiento sobre una palabra que diversos hablantes pueden tener, y Marconi lo ejemplifica con la palabra macaco, explicando que un zoólogo puede conocer muchas proposiciones verdaderas en que esa palabra aparece, mientras que para un hablante común un macaco no es más que un mono.

Marconi, citando a Putman (1975), señala que las diferencias de riqueza, articulación y sutileza entre las competencias de varios hablantes se deben a los efectos de la “división del trabajo lingüístico”. Sin embargo, difiere al indicar que, si bien Putnam distinguía entre hablantes muy competentes en el uso de ciertas palabras y hablantes menos competentes, la distinción que propone tiene que ver principalmente con la competencia individual: “Un hablante puede ser muy competente referencialmente, o sea, en la aplicación de determinadas palabras y ser, a la vez, profundamente ignorante de la naturaleza y propiedades de su referente” (Marconi, 2010, p. 81). En otras palabras, señala que los hablantes difieren unos de otros: donde unos pueden tener una competencia referencial excelente y, a la vez, un conocimiento pobre, así como otros pueden ser

acreedores de un conocimiento excelente, pero de una competencia referencial pobre.

2.1.4. LEXICÓN MENTAL

Se ha hecho mención al *lexicón mental* y se ha definido, someramente, como el espacio o lugar donde se articula y reside el conocimiento de tipo léxico (Lahuerta y Pujol, 1996). El lexicón mental ha sido objeto de estudio y constante preocupación de la psicolingüística; aun así, y pese a los avances que se han obtenido gracias a esta disciplina, la mayoría de los autores coinciden en que se sabe muy poco sobre las capacidades cognitivas y el modo en que funciona (López-Mezquita, 2007).

Aitchison (1994) define al lexicón mental como el conjunto de palabras que un individuo conoce. Una de las características de este conjunto es que posee una estructura precisa, que sigue un complejo y elaborado conjunto de interconexiones que permiten relacionar un gran número de palabras y recuperarlas cuando las necesitamos. Gracias al tipo de organización que tiene este también llamado "almacén mental", se puede tener acceso rápido y eficiente, y cuando sea necesario, a la gran cantidad de datos que se encuentran organizados en la mente. Por ejemplo, un hablante nativo que cuente con unas 60 mil palabras en su léxico puede tardar menos de un segundo en realizar la búsqueda de un término o comprobar su existencia o no existencia (Aitchinson, 1994).

Por otra parte, el contenido del lexicón no es fijo. Se trata de un sistema en continuo cambio a medida que se aprenden o crean palabras, se altera la pronunciación o se extiende el significado a lo que ya se conocía (López-Mezquita, 2007). El mecanismo por el cual se conforma el lexicón mental se basa en tres tipos de tareas fundamentales: etiquetado, que ocurre en los dos primeros años de vida; empaquetado, que consiste en la clasificación de los objetos; y, finalmente la construcción de redes, que tiene lugar gradualmente y continúa a lo largo de toda la vida de la persona (Aitchinson, 1994; López-Mezquita, 2007).

López Morales (1995) proponía que en el lexicón mental existe una serie de términos que no se actualizan a menos que sean necesarios para comunicar una información muy específica. Con esto el autor se refiere al léxico disponible, cuyo estudio se emprende a partir de la disponibilidad léxica, aspecto que será revisado con mayor profundidad en los próximos puntos de este trabajo.

2.2. DISCURSO ESPECIALIZADO

El estudio en torno al discurso especializado se encuentra en pleno apogeo (Parodi 2005a,2005b, 2006, 2007) y está ampliamente aceptado por los estudiosos del lenguaje; sin embargo, se trata de un campo de estudio relativamente nuevo.

Dentro de las primeras complejidades que ha debido sortear es que el propio término de "*discurso especializado*" no siempre se ha utilizado en el mismo sentido. Parodi (2005a) define el discurso especializado mediante una serie de rasgos que lo caracterizan: "Un conjunto de textos que se distinguen y se agrupan

por una coocurrencia sistemática de rasgos lingüísticos particulares en torno a temáticas específicas no cotidianas” (Parodi, 2005a, p.26). Junto a lo anterior, se exige que sus participantes tengan una experiencia previa disciplinar, en otras palabras, una formación especializada dentro de un dominio conceptual particular de la ciencia y la tecnología. Una tercera característica implica que sus textos revelan principalmente una función comunicativa referencial y circulan en situaciones particulares (Parodi 2005a).

Un aspecto que no se desprende directamente de la anterior definición, pero que sí aparece en los llamados *rasgos prototípicos* que Parodi (2005a) propone es el empleo de un léxico especializado. Este rasgo, no solo es relevante, sino que es uno de los principales señalados por otros autores (Cabré, 1993) al caracterizar el discurso especializado.

Cabré (1993) planteaba una problemática que Parodi (2007) dio por zanjada al definir el discurso especializado como una *hiperonimia* de los llamados discurso académico y discurso profesional. Los desacuerdos fundamentales surgían de la consideración de aspectos comunes entre la lengua general y este lenguaje especializado. La divergencia de opiniones se polariza en dos sentidos: el primero es tratar estos subconjuntos como lenguas de especialidad, es decir, como lenguas en sí mismas; el segundo se relaciona con tratarlos como lenguas cuya única especificidad es el léxico (Cabré, 1998a).

No obstante, dejando a un lado los matices de diferentes corrientes teóricas y resaltando los puntos coincidentes, se puede llegar a la siguiente

definición de lenguaje especializado: "Se trata de conjuntos *especializados*, ya sea por la temática, la experiencia, el ámbito de uso o los usuarios; se presentan como un conjunto con características interrelacionadas, no como fenómenos aislados; mantienen la función comunicativa como predominante, por encima de otras funciones complementarias" (Cabré 1993).

En primer lugar, el lenguaje especializado está caracterizado pragmáticamente por las variables temáticas, usuario y situación de comunicación, que implican a su vez unas peculiaridades lingüísticas y textuales. Frente a la lengua general, los lenguajes especializados se desarrollan en función de una temática determinada; son especiales en cuanto al contenido de su discurso, ya que transmiten un conocimiento específico (Parodi, 2007).

En cuanto a los usuarios y la función, los lenguajes especializados son sistemas semióticos complejos por lo que el hablante habrá de contar con una formación especial para poder utilizarlo en la conceptualización, clasificación y comunicación dentro de un dominio determinado. Desde este punto de vista, el lenguaje de especialidad será un subconjunto de la lengua global, usado por profesionales, que puede utilizar varios códigos al mismo tiempo y tiene fines comunicativos dentro de un ámbito específico del conocimiento. Exigencias tales como un mayor grado de exactitud en la expresión implican, por necesidad, una selección a nivel gramatical, semántico y también sintáctico que no sólo estará en función de las exigencias de precisión sino también en función de la ya mencionada situación comunicativa. Todo ello dará lugar a una variación en el

seno mismo de estos lenguajes para cuya descripción se ha propuesto la utilización de un eje horizontal y otro vertical (Cabré 1998a).

La variación horizontal se refiere a la temática y la perspectiva. Dado que la temática especializada es una característica definitoria de estos lenguajes, en función de ésta se puede establecer una primera clasificación de los mismos. Pero además de la temática, habrá que tener en cuenta la perspectiva u óptica desde la que se aborda, ya que el conocimiento científico puede ser tratado de forma trivial, sin atenerse a las estructuras conceptuales del dominio y, por lo tanto, como conocimiento general.

En este sentido, se distingue entre el discurso especializado (con distintos grados de especialización) dirigido a los especialistas; el discurso didáctico destinado al personal en formación y el discurso divulgativo enfocado hacia el público general. Estos distintos niveles de competencia se reflejarán en el mayor o menor uso de terminologías específicas; así, la comunicación especializada requiere que la terminología se adapte a cada tipo discursivo, que viene determinado tanto por la cantidad de información compartida entre emisor y receptor como por la finalidad del texto. Dependiendo de estos dos factores la presencia, importancia y forma de la terminología serán diferentes. Finalmente, hay que añadir que a mayor grado de abstracción y hermetismo discursivo le corresponderá un mayor número de rasgos pertenecientes a los lenguajes artificiales (Cabré, 1993), ya que la necesidad de precisión es mayor.

En cuanto a las características comunes que presentan los lenguajes especializados, las características interrelacionadas a las que se refería Cabré, pueden aglutinarse en torno a su tendencia restrictiva que es la que los sitúa como "tipos mixtos, a medio camino entre los lenguajes artificiales y las lenguas naturales" (Cabré 1993). Es restrictiva en cuanto a su adquisición, que debe hacerse de forma consciente, y su flexibilidad, que se ve constreñida incorporando lenguajes artificiales que hacen referencia a conceptos y funciones únicos del dominio. Su materialización, en la gran mayoría de las ocasiones, es a través del medio escrito por lo que se trata de una manifestación bastante controlada. En lo referente a las funciones, éstas suelen estar reducidas a lo estrictamente descriptivo, clasificativo y comunicativo y, por último, la intención suele ser informativa o incluso directiva.

En este sentido, Cabré (1993) afirma que se pueden establecer una serie de puntos de reflexión que nos permitan profundizar en el análisis del texto especializado desde la perspectiva de los aspectos lingüísticos, funcionales y pragmáticos. Así, desde el punto de vista de los aspectos lingüísticos, se observa una selección léxica con las siguientes características: i) léxico común a textos generales y especializados; ii) léxico fronterizo entre la lengua común y la especializada; y, finalmente iii) léxico claramente específico del texto especializado.

Por tanto, los lenguajes especializados hacen uso de los artificiales definidos como un lenguaje cuyas normas y lexicón están explícitamente establecidos a priori por un colectivo determinado (Cabré 1993). Son el resultado

de una restricción máxima de las propiedades de la lengua general de acuerdo con reglas establecidas para regular dichas restricciones. Las disciplinas científicas tienen la necesidad primordial de ordenar y clasificar los fenómenos que estudian; este orden debe reflejarse también en la denominación de tales fenómenos u objetos. Dependiendo de la naturaleza de los objetos estudiados y del propósito de tales conocimientos, las distintas ciencias han desarrollado diferentes criterios de clasificación. Estas normas establecidas por los organismos competentes son muy específicas y no dan gran margen de variabilidad al usuario. La función a desempeñar por estas lenguas "fijas" es la de apoyar a la lengua natural mediante una mayor especificidad de referencia, economía de expresión y mayor adecuación a la comunidad de usuarios (Cabré 1993).

2.2.1. LA NOTICIA COMO DISCURSO ESPECIALIZADO

En una concepción tradicional, se puede definir la noticia -o texto periodístico- como un hecho verdadero, inédito o actual, de interés general, que se comunica a un público que puede considerarse masivo, una vez que -este hecho- ha sido recogido, interpretado y valorado por los sujetos promotores que controlan el medio utilizado para la difusión (Martínez Albertos, 2004; Hernández, 2016).

Naisbitt (citado en Fontcuberta, 1987) señala que existen condicionantes sociales que buscan el desarrollo de medios especializados, anunciando en la década de 1980 la decadencia de los medios de información general. "Hace más de un decenio desaparecieron varias revistas de información general: *Life*, *Look*

y *The Saturday Evening Post* (...) el mismo año en que desaparecieron esas importantes revistas de gran circulación surgieron 300 nuevas de interés particular” (Fontcuberta, 1987, pp. 52-53). Al respecto, Fontcuberta señala que no son los medios los que se especializan sino los contenidos. “En todo caso, los nuevos medios facilitan la difusión de los contenidos especializados, aunque es la prensa la que ha llevado más lejos la tendencia a la especialización” (Fontcuberta, 1987, p.53). Existen dos preguntas fundamentales que el periodista debe realizarse al iniciar la redacción de cualquier noticia: ¿qué quiero decir?, y ¿a quién? En este sentido, los acontecimientos o hechos sobre los cuales las noticias deben informar son intrínsecamente complejos, se producen dentro de un contexto determinado y con la posibilidad de ser enfocados desde distintas perspectivas.

2.2.2. LA ESTRUCTURA DE PIRÁMIDE INVERTIDA, EL LEAD o ENTRADILLA

La estructura de la noticia ha sido concebida para entregar el máximo de información en el menor tiempo o espacio posible (Fontcuberta, 1987). El *lead*, también conocido como *entradilla*, forma parte de la llamada *pirámide invertida* en el periodismo. La estructura del *lead* surge en la guerra de Secesión norteamericana, y modificó de manera sustantiva la estructura de los propios periódicos. Los corresponsales de guerra pedían preferencia en las oficinas de telégrafo para despachar con prontitud las últimas informaciones ocurridas en el campo de batalla y para ganar tiempo evitaban todo tipo de comentarios,

opiniones y detalles excesivos de los hechos, informando sólo lo más relevante de los acontecimientos: qué, quién, cuándo, dónde, por qué y cómo.

La presencia del concepto de *lead* en este trabajo cobra relevancia debido a que una noticia informativa, escrita bajo la estructura de pirámide invertida, con una entradilla o lead que responda a las seis preguntas mencionadas con

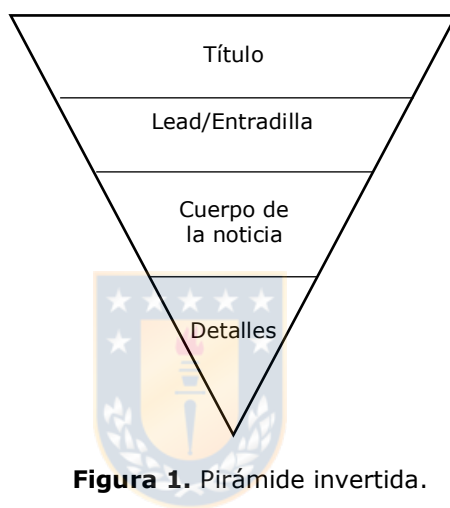


Figura 1. Pirámide invertida.

anterioridad debiera constituir un cuerpo de información robusto elaborado en pocas palabras.

En la **Figura 1** se presenta cómo se estructura una noticia en pirámide invertida. Retomando la idea del párrafo anterior, esta investigación se enfocará en estudiar la riqueza léxica comprendida entre el título y el *lead* de la noticia informativa.

2.3. RIQUEZA LÉXICA

El estudio del léxico cuenta con dos vías de aproximación: las cualitativas y las cuantitativas. Dentro de las primeras –cualitativas-, se encuentran los estudios de frecuencias y de referencia, que están muy unidos a los campos semánticos y a los relacionales; estos estudios se ocupan, por ejemplo, de hiponimias, sinonimias, antonimias y homonimias. En el ámbito de las aproximaciones cuantitativas encontramos, entre otros, lo relativo a los léxicos ‘básicos’, los ‘disponibles’ y la ‘riqueza léxica’ (López Morales, 2002).

Por su parte, Read (2011) define la riqueza léxica como un constructo en el cual emergen aspectos como diversidad léxica, sofisticación léxica y diversidad léxica, y, eventualmente número de errores. Housen, Kuiken y Vedder (2012) desde el ámbito de la adquisición de las segundas lenguas, incluyen el concepto de riqueza léxica bajo la idea de *complejidad*, definiendo este último como “la habilidad de usar una gama amplia y variada de estructuras y vocabulario sofisticados en la L2” (Housen et al, 2012).

Una visión comparativa la ofrecen Bulté y Housen (2012), quienes realizan una crítica al indicar que, en general, los investigadores no desarrollan por completo el concepto teórico de riqueza léxica, sino que se enfocan en describir las medidas empleadas (Berton, 2014; Bulté y Housen, 2012). Es más, los autores son enfáticos al señalar que para obtener medidas significativas y válidas sobre complejidad lingüística se debe establecer, en primer lugar, un nivel

teórico sobre lo que es complejidad y su manifestación de forma concreta, tal como lo ejemplifican en la **Figura 2**.

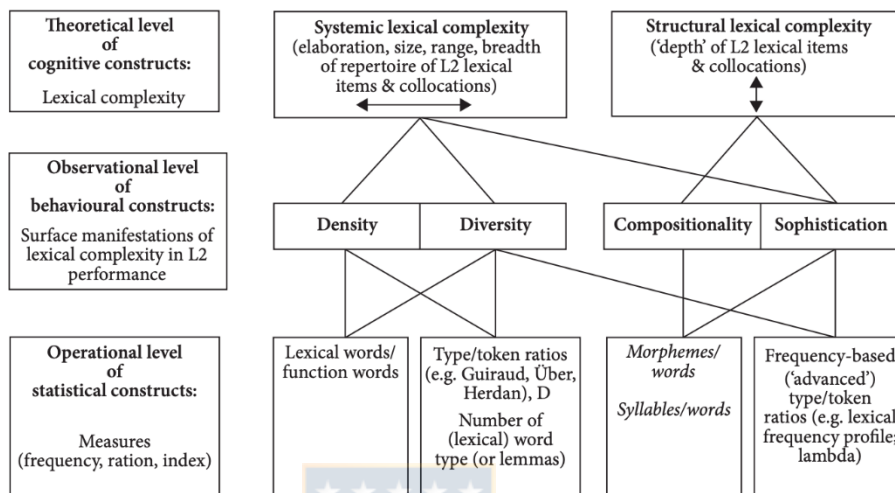


Figura 2. Complejidad léxica en diferentes niveles.

López Morales (2002), por su parte, releva el concepto de *calidad de la escritura* en el cual se integra una serie de factores entre los que destacan riqueza léxica, madurez sintáctica, esquemas de cohesión y coherencia discursiva. El autor explica a qué se asocia cada factor: la amplitud y variedad del vocabulario se encuentran apoyadas en la *disponibilidad léxica* del hablante; por otra parte, la *madurez sintáctica*, se encuentra asociada al grado de entrenamiento combinatorial de oraciones simples en el discurso; mientras, los esquemas de *cohesión* y la *coherencia* discursiva dependen esencialmente del orden que quiera dársele conscientemente a los elementos constitutivos del discurso. Este trabajo se delimitará en los estudios cuantitativos de la riqueza léxica abordando los aspectos de *densidad* y *diversidad* léxica.

Los primeros estudios sobre riqueza léxica aparecieron en 1954, con los trabajos estadísticos de Giraud. Si bien estas investigaciones se han ido refinando cada vez más, los postulados iniciales básicos se mantienen, definiendo a la riqueza léxica como “la relación existente entre el número de palabras y de vocablos de un determinado texto como elemento básico del análisis” (López Morales, 2002).

Aquí, nuevamente se aprecia la diferencia esencial entre *palabra* y *vocablo*, dos factores fundamentales sobre los que se construyen los estudios de riqueza léxica basados en aproximaciones cuantitativas. La primera -palabra-, se refiere al material gráfico comprendido entre dos espacios en blanco de un texto; mientras que vocablo, se define como las palabras diferentes que aparecen en un texto, sin contar las repeticiones.

Para dar sustento a este tipo de estudios la estadística léxica ha sido el primer peldaño para el desarrollo de estas investigaciones. La estadística léxica, o léxico-estadística, abarca el conjunto de operaciones, a veces sumamente complejas, que toman como unidades de trabajo las palabras y los vocablos: la palabra como unidad del texto, y el vocablo, como unidad del léxico (López Morales 2002).

Müller (1973) propuso la idea de que la estructura de un vocabulario comprende elementos cuantitativos simples: el número de vocablos del texto y la frecuencia de cada uno de ellos. Junto a estos elementos, agrega aspectos cualitativos, tales como la naturaleza gramatical de los vocablos y las relaciones

de asociación (gramaticales o semánticas, paradigmáticas y sintagmáticas) que existen entre vocablos. Para él, cuantificar vocabulario de un texto es proceder a dos operaciones distintas que pueden ser sucesivas o simultáneas. La primera se relaciona con el recuento de las palabras que componen el texto y cuyo número, representado por N, entregará una medida de la extensión del texto. En segundo lugar, el recuento de los vocablos empleados en el texto y cuyo número, representado por V, mide la extensión del vocabulario, idea fundacional que aún persiste en la actualidad (López Morales, 2002; Johansson, 2008; Núñez y del Moral, 2010; Reyes, 2010).

Para Müller la norma lexicológica, en términos de las decisiones metodológicas que se toman respecto al tratamiento del problema, es la que debe dar las reglas para delimitar la palabra y el vocablo. "Por lo general, se adopta la solución de sentido común que rige a los diccionarios, a pesar de las objeciones lingüísticas que se podrían formular" (Müller, 1973). De esta manera se obtienen, entre otros resultados, índices que indican las ocurrencias de las palabras, es decir, las características estadísticas de las entradas, el lugar de aparición en el texto, el vocabulario común de una lengua y la jerarquización por grupos de mayor a menor presencia en la lengua (López Morales, 2002).

López Morales (2002) indica que uno de los resultados más novedosos de la léxico-estadística consiste en confirmar que el vocabulario de los hablantes de una comunidad de habla es relativamente limitado, debido a que son muy pocas las palabras que presentan una frecuencia alta. Comúnmente, el individuo

concentra sus necesidades de expresión en una cantidad relativamente reducida de entradas.

Los estudios realizados en Francia hace casi cincuenta años mostraron que, independientemente de los términos de la especialidad laboral de cada individuo, un hablante culto usa cuatro o cinco mil vocablos, mientras que el no culto puede manejarse entre dos y tres mil.

Reyes (2010), por su parte, realiza un recuento de los indicadores de riqueza léxica que se han considerado para realizar descripciones cuantitativas de las muestras. Entre ellos se encuentran:

i) Obtención del porcentaje de vocablos; es decir, una prueba que mide la proporción de palabras diferentes del total de palabras de una composición. Para hallar este índice de riqueza léxica se cuenta el número total de palabras del discurso y el número de vocablos (palabras diferentes) con el fin de reconocer la proporción entre ambas cifras.

ii) Intervalo en que las palabras nocionales (N/PN) en el texto. Este índice considera la cantidad total de palabras del escrito y la divide por el conjunto formado por nombres, adjetivos, verbos y adverbios, para obtener el intervalo de aparición en el texto de estas palabras. "Se dirá que este índice aumenta conforme disminuya el intervalo de aparición de palabras nocionales; por el contrario, cuanto más dilatado sea el intervalo, menor será la riqueza léxica de la muestra en cuestión. Así, un intervalo de voces de contenido léxico de 2,1 detecta

menor riqueza léxica que si se consigue un intervalo entre lexías de 1,2" (Reyes, 2010).

iii) Índice hápax, que detecta el empleo de palabras que sólo se utilizan una vez en el texto analizado. La variedad de hápax se obtiene con la aplicación de la fórmula V/V_1 , donde V es el número total de vocablos y V_1 , la suma de las palabras de una ocurrencia. Para determinarlo se divide la cifra total de vocablos por aquellos que sólo se emplean una vez en un texto determinado, de modo que cuanto más pequeño sea el índice, más riqueza léxica caracteriza al texto que se confronta con otro. Reyes (2010) explica que conforme el índice va aumentando, la variedad léxica va disminuyendo; es decir, una medida de 1,34 revela mayor cantidad de unidades de una sola aparición en un texto que otra de 1,37.

La riqueza léxica, entonces, podría utilizarse para hacer referencia de una o más características del vocabulario usado en una muestra de lengua, sea oral o escrita, y puede medirse de diferentes formas según como se entienda el concepto. En particular, el objetivo de las medidas de riqueza léxica es intentar cuantificar la variedad y el tamaño del vocabulario usado (Laufer y Nation, 1995: 307).

2.3.1. DIVERSIDAD Y DENSIDAD LÉXICA

Read (2011) señala que la riqueza léxica es una medida estadística y supone que la buena escritura tiene características léxicas como la variedad de palabras diferentes, en lugar de un número limitado de palabras utilizadas repetitivamente. La primera característica de la riqueza léxica se relaciona con la

diversidad léxica la que se refiere al número de palabras diferentes utilizadas en un texto, un rango mayor indica una diversidad mayor (López Morales, 2002; Cushing, 2002; Johansson, 2008; McCarthy y Jarvis, 2007 y 2010). La medida que se aplica en este caso es el type-token ratio o TTR, donde **type**⁵ corresponde a las palabras distintas del texto –más arriba definidas como *vocablo*– y **token**⁶, al número total de palabras. Existe, además, otra clasificación que para efectos de este trabajo no es significativa, se trata del hápax legomena, que corresponde a *palabras* cuya aparición es única dentro del texto (Daller, van Hout y Treffers-Daller, 2003; Matsuda, Sadowsky y Sabaj, 2001).

La diversidad léxica fue definida en la década de 1980 como la cantidad relativa de repetición de vocabulario. Trabajos más recientes, como el de Malvern, Richards, Chipere y Durán (2004), la definen como “the range of vocabulary and avoidance of repetition”⁷. Esta definición es planteada por otros autores quienes indican la diversidad –o variedad– en el uso de una palabra dentro del texto (McCarthy y Jarvis, 2007; Jarvis, 2013). Es más, el concepto de riqueza léxica ha sido utilizado a menudo como sinónimo de diversidad léxica (Arnaud, 1984; Engber, 1995; Laufer y Nation, 1995; Wimmer y Altmann, 1999; Laufer, 2003; Jarvis, 2013).

⁵ López-Mezquita la define como “la unidad usada para contar y responder a preguntas tales como cuántas palabras se precisan para leer determinado libro o, cuántas palabras hay en un diccionario” (2005, p.76).

⁶ También llamadas *running words*. Un token es “la unidad utilizada para contar y responder a preguntas tales cuántas palabras hay en una página o en una línea, o a qué velocidad se lee, o cuántas palabras puede un hablante medio producir en un minuto” (López-Mezquita, 2005, p.76).

⁷ “El rango de vocabulario, evitando la repetición”.

Una segunda característica de la riqueza léxica tiene que ver con el porcentaje de palabras léxicas –o de contenido semántico– (verbos, nombres, adjetivos y algunos adverbios) comparado con las llamadas palabras gramaticales –o funcionales– (artículos, preposiciones, conjunciones, entre otros). Esta medida se conoce como densidad léxica (Read, 2011; Johansson, 2008, Gregory-Signes, Clavel-Arroitía, 2015). La densidad léxica es, además, un indicador de la calidad informativa del texto: un alto índice de densidad léxica significa que tiene a su haber más palabras de contenido, por lo tanto, dicho texto entrega más información.

Tanto la diversidad como la densidad léxica son medidas cuya ventaja radica en su sencilla operacionalización y medición gracias a los avances en técnicas computacionales de análisis, junto a los estudios de corpus (Johansson, 2008).

2.3.1.1. ÍNDICE DE RIQUEZA LÉXICA

Las medidas más comunes de la diversidad léxica son el número de palabras diferentes y el *type-token ratio* (TTR). Se le llama *token* a cada palabra que compone el texto, mientras que *type* es cada palabra diferente, que puede también ocurrir varias veces. El número de palabras diferentes (Vermeer, 2000; Fusté-Herrman, Silliman, Bahr, Fasnacht y Federico, 2006) es un indicador importante, ya que tiene en cuenta el tamaño del vocabulario empleado por el informante al producir un texto (Malvern et al. 2004). No obstante, Malvern et al. (2004) encuentran que no es una medida simple, pese a ser solamente un

cómputo de las palabras diferentes presentes en la muestra. Su problema es la dependencia de la extensión de los textos, ya que un texto más largo contendrá más palabras diferentes, aunque no se puede predecir en qué proporción.

Para abordar el análisis de riqueza léxica, se siguió la línea tradicional de la léxico-estadística, que en primer lugar busca obtener el porcentaje de vocablos; es decir, la proporción de palabras diferentes del total de palabras de una composición, lo que otorga una visión de la diversidad léxica de un texto y entrega, según López Morales (2002) un indicador grueso del texto. Para calcular este índice se contabiliza el número total de palabras del discurso y el número de vocablos (entendidos como palabras diferentes), con el fin de reconocer la proporción entre ambas cifras en cada uno de los textos.

Porcentaje de vocablos (PV). Este procedimiento requiere que se divida el total de vocablos (V) entre el total de las unidades léxicas comprendidas en el texto (N) y luego que se multipliquen por cien:

$$PV=(V/N)*100$$

Este índice, también es conocido como TTR –o Type/Token Ratio-, con la diferencia que no se expresa en porcentaje, sino que con el resultado de la fórmula:

$$LV=V/ N$$

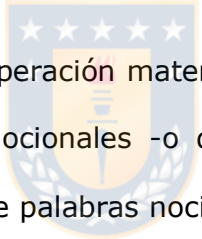
(LV) = Diversidad Léxica.

(V) = Total de vocablos.

(N) = Extensión total del texto.

Un segundo cálculo, propuesto por las investigaciones López Morales (2002) y Reyes (2010), para complementar el concepto de riqueza léxica, está dado por el Intervalo de aparición palabras de contenido nocional (IAT), el cual considera la cantidad total de palabras del texto y la divide por el conjunto formado por nombres, adjetivos, verbos y adverbios, también conocidas como palabras nocionales (PN) o palabras con contenido semántico:

$$IAT = N / PN$$



El resultado de esta operación matemática refleja cifras relacionadas con la proporción de palabras nocionales -o de contenido- en el texto. En otras palabras, a mayor número de palabras nocionales, menor es el intervalo; lo que se interpreta como un mejor índice de riqueza léxica. Se trata de una medida que resulta efectiva especialmente cuando se trabaja con textos de un grado académico particular y se pretende identificar la relación de un sujeto con el resto del grupo. En este caso la relación que se explora es la diferencia entre los dos grupos de estudiantes y la riqueza léxica de los profesionales. (López Morales, 2002)

Finalmente, el índice de densidad léxica fue propuesto por Ure en 1971. En sus hallazgos señaló que usualmente más del 40% de un texto corresponde a palabras con contenido semántico, mientras que en el discurso oral este porcentaje no supera dicha cifra. Lo anterior sería reflejo de que la información y

las ideas se presentan con mayor concentración en el lenguaje escrito que en el oral (Read, 2000). Para obtener este índice se aplicará al corpus de las noticias la siguiente fórmula matemática:

$$DL = PN/N$$

(DL) = Densidad Léxica.

(PN) = Total de palabras nocionales.

(N) = Extensión total del texto.

Johansson (2008) centra su investigación en la correlación de las medidas de diversidad y densidad léxica y su sensibilidad respecto a la modalidad (textos escritos u orales). Su corpus estuvo compuesto por 316 textos orales y escritos en L1 sueco, narrativos y expositivos. Los participantes se distribuyen entre cuatro edades: estudiantes de 10, 13 y 17 años y universitarios adultos con al menos dos años de estudios universitarios y una tesina en su historial. Cada grupo estuvo formado por 20 personas, menos los adultos que fueron 19. La extensión de los textos varió entre 50 y 650 palabras. Los resultados demostraron que la diversidad léxica es más sensible al nivel de competencia del participante respecto a la densidad. Además, la diversidad léxica también tiende a crecer junto con la edad, tendencia menos evidente si se considera la densidad léxica (Johansson, 2008, p. 70). Finalmente, la autora afirma que ambos aspectos de la riqueza léxica discriminan entre niveles de competencia, aunque no se aprecian diferencias dentro de los grupos 10-13 y 17-adultos. Sin embargo, observando el desarrollo a largo plazo, esto es, entre los primeros dos grupos (10 y 13 años) y

los últimos dos (17 años y adultos), las diferencias son evidentes, aunque de nuevo más marcadas en términos de diversidad que de densidad. Por esta razón, Johansson (2008, p. 77) afirma que la diversidad léxica se presta mejor para detectar diferencias entre grupos de edades distintas.

Las limitantes a la aproximación estadística de la riqueza léxica han sido largamente revisadas (Johansson, 2008; McCarthy y Jarvis, 2007; Torruella y Capsada, 2013) y se relacionan a la dependencia del índice de riqueza con la extensión del texto ($N=token$). En otras palabras, la riqueza léxica está ligada con la proporción de *types* (V) y el total de palabras del texto, al aumentar N también aumenta V, pero no de la misma manera debido a que la producción de nuevas palabras tiende a disminuir en la medida que la extensión de texto aumenta.



2.4. DISPONIBILIDAD LÉXICA

Las investigaciones en Disponibilidad Léxica surgieron en Francia durante la década de 1950. El objetivo de esta disciplina, estrechamente ligada a la Lingüística Aplicada, era elaborar un método básico de enseñanza del francés destinado a extranjeros. La idea era establecer el francés elemental.

Este proyecto fue encargado al equipo de lingüistas conformado por Georges Gougenheim, René Michéa, Paul Rivenc y Aurelein Sauvageot. La propuesta era establecer el léxico "elemental" de dicha lengua. La idea inicial era la enseñanza de la lengua a partir del vocabulario frecuente. El equipo recurrió al recuento léxico, basado en la frecuencia de uso; la hipótesis era que las palabras

más utilizadas –las más frecuentes- eran las que debían enseñarse en primer lugar. Esta aproximación entregó las primeras dudas cuando en los recuentos no aparecieron palabras necesarias para la vida cotidiana; palabras que se referían a realidades concretas que no aparecían en los listados basados en las frecuencias porque su uso estaba condicionado por el tema de la conversación o de los escritos cuantificados. Al respecto Echeverría, Herrera, Moreno y Pradenas (1987) señalan:

“Invariablemente, los índices más altos [de frecuencia] corresponden a palabras que pertenecen a inventarios cerrados, como preposiciones, conjunciones verbos auxiliares y modales, etc. y a otras de débil carga semántica que podrían calificarse como nombres de clase por el hecho de poder aplicarlas a gran cantidad de realidades que tienen un nombre específico en la lengua.” (1987, p. 55)

El estudio de frecuencia léxica del equipo de Gougenheim, realizado en textos franceses orales y escritos, dio como resultado una lista de frecuencias decrecientes, donde las palabras más “usadas” correspondían a palabras gramaticales; en segundo lugar, se encontraron verbos; en tercer lugar, adjetivos; y en cuarto lugar algunos sustantivos de carácter general. Sin embargo, dicha lista incluyó pocas palabras concretas. Echeverría (1987) señala que algunas palabras como *codo*, *muela*, *metro*, *auto*, etc., que designan realidades objetivables y presentes en el mundo inmediato de los hablantes, no figuran en la lista de frecuencias del francés fundamental. Lo que revela

Echeverría es que las palabras más frecuentes de la lengua no son términos concretos con fuerte carga semántica, a pesar de que, muchas veces, tales términos designan objetos tangibles de la realidad y del entorno cotidiano de los de hablantes.

La evidencia del trabajo del equipo francés hizo notar la poca estabilidad de las "palabras concretas" frente a la gran estabilidad de las "palabras gramaticales" y de las "palabras de semanticismo extendido" (Bartol, 2006). Estas últimas, corresponden a aquéllas que están presentes en cualquier texto, mientras que las primeras –palabras concretas-, sólo aparecen en textos cuya naturaleza lo permita. René Michéa las denominó "atemáticas", frente a "palabras temáticas" (Bartol, 2006).

Echeverría (1987) explica que las palabras concretas encuentran cabida en textos cuyos temas hacen referencia a objetos designados por tales palabras; pero, aun así, la aparición de dichas palabras de forma reiterada no está asegurada "porque la referencia a los objetos puede hacerse también por los mecanismos de pronominalización y por el uso de superordenados" (Echeverría, 1987). Lo anterior implica que un hablante no utiliza ciertas palabras en su discurso, no por el hecho de que las desconozca, sino porque las condiciones en las cuales se da el discurso pueden hacer superflua tal utilización. Sin embargo, el hablante tiene a su disposición, en su competencia lingüística y en su lexicón mental, un sinnúmero de palabras que utiliza poco en sus discursos tanto orales como escritos. Éstas son las palabras disponibles, a las que Gougenheim definió como aquellas que "están a nuestra disposición en la mente, pero no la

pronunciamos o escribimos efectivamente solo cuando las circunstancias nos llevan a ello". La disponibilidad léxica, por lo tanto, estaría definida en palabras de René Michéa como:

"Una palabra disponible es una palabra que, sin ser particularmente frecuente, siempre está lista para ser usada y se presenta en forma inmediata y natural a la mente en el momento en que se la necesita. Es una palabra que, formando parte de las asociaciones de ideas usuales, existe en potencia en el hablante, en cuanto entran en juego dichas asociaciones" (Michéa, citado en Echeverría 1987)

Basados en esta evidencia, los lingüistas franceses se dieron cuenta de que para captar estas palabras se necesitaban pruebas asociativas -conocidas actualmente como "centros de interés"-, que estimularan el lexicón mental de los informantes para que produjeran estas palabras que están disponibles, pero que sólo aparecen cuando las circunstancias comunicativas lo exigen.

Paredes (2012), en una revisión de los desarrollos teóricos que la disponibilidad léxica ha alcanzado en los últimos años, señala que los estudios se abordan desde la perspectiva general de la variación lingüística. Por otra parte, los avances teóricos señalan que la disponibilidad léxica "se ha servido de -pero también está sirviendo a- otras disciplinas, como la semántica, la psicolingüística o la lingüística cognitiva" (Paredes, 2012). Estas iniciativas se suman a los ya tradicionales aportes en sociolingüística y a la dialectología.

Es así como se pueden apreciar estudios de disponibilidad léxica que han realizado valiosos aportes a la semántica, en los que se abordan asuntos como los campos semánticos, las relaciones de sinonimia y antonimia, y de relaciones de inclusión, entre otros (Paredes, 2012). Otro aporte ha sido en el análisis de asociaciones léxicas (Ávila y Villena, 2010; Echeverría, 2002; Echeverría, Vargas, Urzúa y Ferreira, 2008) en el que se contribuye con una clasificación de las unidades que aparecen en los listados según su mayor o menor compatibilidad semántica con el estímulo de origen.

Un tercer ámbito es el que propuso Hernández et al. (2006), quien considera que los centros de interés son categorías cognitivas –aunque no todas con las mismas características–, y las asocia con niveles de prototipicidad. Por su parte, Ferreira y Echeverría (2010) entregan una nueva visión al aplicar el carácter asociativo de la organización del léxico a un estudio comparativo entre L1 y L2.

Así, se puede continuar con un largo listado sobre las iniciativas que han partido desde los estudios de disponibilidad léxica hacia la búsqueda de un conocimiento más profundo de las diversas implicancias del léxico.

No se puede olvidar resaltar la gran iniciativa que coordina el lingüista Humberto López Morales, el Proyecto Panhispánico de Disponibilidad Léxica, que tiene como objetivo Elaborar diccionarios de Disponibilidad Léxica para las diversas zonas del mundo hispánico. La homogeneidad de criterios permitirá

establecer comparaciones de tipo lingüístico, etnográfico y cultural, dibujar áreas de difusión y, en general, servirá de punto de partida para análisis posteriores.

2.4.1. CENTROS DE INTERÉS

Echeverría (2002) señala que la noción de *centro de interés* fue tomada de la pedagogía tradicional y corresponde a agrupaciones de palabras en torno a ideas de *conjunto, ambiente o situación*, en las que los hablantes ponen en juego asociaciones de ideas que recurren tanto a conocimientos como a experiencias vividas. Estos centros alcanzan un alto grado de coherencia semántica, en el sentido que se les puede limitar a un reducido número de clases lógicas; otros, en cambio presentan un débil grado de coherencia lógica porque el número de clases lógicas que lo componen es difícil de delimitar.

Algunas críticas al modelo desarrollado por el equipo francés, relacionadas con los centros de interés, fueron recopiladas por Sánchez-Saus Laserna (2011), quien señala que ya en 1970 lingüistas como Julliard indican como uno de los aspectos más débiles de la metodología de disponibilidad léxica la subjetividad con la que se elige un determinado centro. Otra de las críticas señala que los resultados cuantitativos de las encuestas de disponibilidad léxica originales no eran material para análisis estadísticos debido a que el francés fundamental trabajaba con muestras cerradas de veinte palabras por centro.

Sin embargo, cabe señalar que los avances y desarrollos en este ámbito han permitido adentrarse en los estudios de disponibilidad léxica con un mayor

grado de certeza y si bien la metodología guarda semejanza con aquella de la década de 1950, ha sido actualizada y perfeccionada abordando aspectos como los socioculturales, sexo, edad y rendimiento, entre otros aspectos

2.4.2. ÍNDICE DE DISPONIBILIDAD LÉXICA

El léxico disponible se obtiene a partir de una prueba de disponibilidad léxica. Este instrumento implica que en dos minutos los sujetos involucrados como muestra en la investigación entregan su léxico disponible en cada uno de los centros de interés.

Posteriormente, se realiza una depuración de los listados. En este punto se debe realizar un análisis teórico y metodológico sobre la disponibilidad léxica, para tomar decisiones respecto a temas como el tratamiento de errores ortográficos, utilización de plurales u otros aspectos que puedan aparecer en las listas de léxico disponible (elementos considerados más adelante, en el capítulo Marco Metodológico). Una vez resueltas estas situaciones se procede al cálculo del Índice de Disponibilidad Léxica, IDL.

El cálculo de IDL, según Echeverría et al. (1987), debe considerar la frecuencia de las menciones de cada palabra en un grupo determinado, es decir, cantidad de veces que la palabra es citada por los encuestados en la posición 1, 2, 3, etc. (f); además, debe considerar el número de sujetos implicados en la muestra, ya que la probabilidad de que ocurra una palabra va en función del grupo. Un tercer aspecto fundamental para este cálculo es el lugar en el que fue mencionada la palabra. Echeverría et al. (1987) señala que "No está igualmente

disponible un término que nos viene a la mente de inmediato que aquel que aparece luego de una dilatada reflexión”. Por este motivo es que en la fórmula la posición debe ponderarse con un mayor peso que las ubicadas en el centro o al final del listado.

López Morales, en Echeverría et al (1987) propone que a la fórmula que calcula el IDL se le agregue un factor de ponderación lambda (λ) elevado a n , donde n es la posición en la cual es citada la palabra y $\lambda=0,9$,

$$D(p) = \frac{f_1}{N_1} + \frac{N_2 f_2}{N_1 N_2} + \frac{\lambda^2 N_3 f_3}{N_2 N_3} + \dots + \frac{\lambda^{n-1} N_n f_n}{N_{n-1} N_n}$$

El problema detectado por Echeverría et al (1987) con esta fórmula es que debido a que la frecuencia en cada posición se divide por el número de respuestas para la posición, ocurre que a menor respuestas en una posición determinada, mayor será el valor asignado a la frecuencia en esa posición y, por lo tanto, en las últimas posiciones la fórmula de López Morales entrega una importancia exagerada. (Echeverría et al. 1987) Por este motivo es que se adopta la siguiente fórmula

$$IDL(p) = \frac{f_1 + \lambda f_2 + \lambda^2 f_3 + \dots + \lambda^{i-1} f_i}{N}$$

Donde N equivale al total de menciones en primera posición o, lo que igual, al total de sujetos de la muestra (Echeverría, 1987).

2.4.3. ESTUDIOS DE DISPONIBILIDAD LÉXICA EN CHILE

En Chile los estudios de disponibilidad léxica se inician con Echeverría en la década de 1980, donde después de analizar la disponibilidad léxica en estudiantes chilenos, en relación a diferentes centros de interés, concluyen que la información obtenida es muy interesante para el ámbito pedagógico, porque permite obtener una base de datos, con los cuales se pueden realizar recomendaciones a las autoridades educativas, a los autores de textos y maestros.

Entre otros hallazgos, se obtuvo evidencia de que los alumnos de enseñanza básica y media en nuestro país no tienen un léxico disponible suficiente (Valencia y Echeverría, 1999), lo que inevitablemente trae problemas a la hora de comprender un texto o entender lo que un profesor quiere comunicar (Ferreira, Salcedo, y Del Valle, 2014).

En 2006, un equipo de investigadores (Urzúa, Sáez y Echeverría, 2006), realizaron una investigación en la cual se aplicó una encuesta de disponibilidad léxica a alumnos y docentes de la carrera Ingeniería Civil Matemática perteneciente a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Concepción. La investigación tuvo como objetivo saber cuánto y cuál es el léxico disponible de los alumnos, en centros de interés directamente relacionados con las materias que en esta carrera se tratan.

En la investigación se confirmaron las hipótesis planteadas, concluyendo que el léxico del alumno aumenta durante sus años de estudio y que el léxico del

profesor siempre es superior al del alumno. Además, el léxico del alumno tiende a ser básico y el del profesor a ser más específico.

Extraer la disponibilidad léxica de los estudiantes sigue siendo una preocupación para los investigadores. Ferreira, Salcedo y Del Valle (2014) realizaron un estudio de disponibilidad léxica en el ámbito de las matemáticas, que tuvo como objetivo cuantificar y describir el léxico disponible en matemáticas de los alumnos de enseñanza media de la ciudad de Concepción, Chile, en cuatro centros de interés (números, álgebra, geometría, y probabilidades, datos y azar/estadística). Utilizaron un modelo de disponibilidad léxica, a través de un análisis cuantitativo, considerando tres índices: promedio de respuestas (XR), total de palabras diferentes (PD) y el índice de cohesión (IC). Como resultado de esta investigación evidencian un aumento en el promedio de palabras y palabras diferentes a medida que aumenta el nivel educativo y tipo de establecimiento, observándose sólo una leve diferencia entre los establecimientos subvencionados y particulares. También que hay mayor cohesión léxica en los establecimientos municipales que en los subvencionados, y de los centros de interés estudiados, el centro con mayor promedio de palabras y mayor índice de cohesión en *geometría*. Los autores concluyen que los cursos adquieren el léxico de manera similar a medida que aumenta su nivel educativo. Estos resultados guardan relación con lo reportado en las investigaciones de Urzúa, Sáez y Echeverría (Urzúa, Sáez y Echeverría, 2006).

2.5. REDES DE ASOCIACIÓN LÉXICA

Echeverría (2001) señala que en la lingüística aplicada se distingue entre léxico frecuente y vocabulario disponible, el primero constituido por los vocablos de mayor ocurrencia en un conjunto de textos; y el segundo, por los términos que acuden con facilidad a nuestra mente en una situación determinada. Obtener unos u otros considera sus propias técnicas y los resultados muestran lo descubierto por el grupo de franceses explicado en puntos anteriores: las listas ordenadas de palabras frecuentes difieren completamente de las listas ordenadas por índices de disponibilidad. Las preguntas asociadas a estos resultados ¿Por qué ciertos términos aparecen antes que otros?, o ¿Por qué hay algunos que muestran una asociatividad con respecto de otros?, y más sencillo aún ¿Por qué un hablante selecciona un determinado término para referirse a un tema? Son interrogantes que han perdurado en el tiempo y han dado pie para que disciplinas como la psicolingüística, la lingüística cognitiva, e incluso el desarrollo de complejas teorías de asociación detrás de múltiples *softwares* relacionados con inteligencia artificial, busquen respuestas sobre la naturaleza del léxico en la mente o, más bien, sobre cómo se organiza en el cerebro el conocimiento que tiene el hablante sobre el vocabulario.

Johnson Laird (1983) propone que el conocimiento de las relaciones semánticas no es la totalidad de la competencia semántica, por ese motivo es que las teorías del significado, u otros equivalentes, para representar el significado resultan inadecuadas, "a menos que una teoría ponga el lenguaje en

relación con el mundo...no es una teoría del significado completa” (Johnson Laird, 1983, citado en Marconi, 2000).

Johnson Laird (1983) propone que el conocimiento de las relaciones semánticas no es la totalidad de la competencia semántica, por ese motivo es que las teorías del significado, u otras equivalentes para representar el significado, resultan inadecuadas, “a menos que una teoría ponga el lenguaje en relación con el mundo [...] no es una teoría del significado completa” (Johnson Laird, 1983, citado en Marconi, 2000).

El autor ha insistido en que las relaciones semánticas no determinan las condiciones de verdad y en que para dar cuenta cabalmente de la competencia semántica hace falta una teoría de cómo los elementos léxicos se proyectan en el mundo real, entendiendo esto como un proceso cognitivo y no como el supuesto de que se da semejante proyección. Entender el lenguaje, entonces, consiste en construir modelos mentales de las proposiciones que se entienden donde el aspecto referencial de la competencia léxico-semántica es una destreza cognitiva de la mente humana.

Las redes semánticas entre palabras son construcciones cognitivas que tienen lugar en el lexicón de todo hablante, dentro de su competencia plurilingüística; por lo que constituyen una motivación y una gran facilitación del aprendizaje léxico, tanto para el reconocimiento de una palabra en el *input* al que está expuesto, como para su recuperación en una situación de producción lingüística en el *output* (Baralo, 2007). Vienen, por lo tanto, de una tradición

teórica que estudia las representaciones del significado léxico y que, en este caso, ha sido de importante influencia para el campo de la inteligencia artificial (Marconi 2000, p. 31).

Por otra parte, bajo la premisa de las teorías conexionistas, se ha desarrollado investigación sobre las bases neuronales del léxico. Echeverría (2001) indica que los estudios sobre afasias permitieron avanzar en la confección de un posible "mapa de funciones lingüísticas alojadas en el cerebro", estudios en los que se han logrado avances gracias a las nuevas tecnologías que han permitido el estudio de la función cerebral en sujetos vivos, con métodos no invasivos como la resonancia magnética nuclear, la tomografía por emisión de positrones y otros, más innovadores aún, como la utilización de softwares para la investigación en neurolingüística.

Caramazza (1999), citado en Echeverría 2001, indica que las evidencias obtenidas de los estudios neuropsicológicos y de imagenología neuronal apuntan a una conclusión compartida por la comunidad científica: "El lexicon mental está organizado en subsistemas neuronales relativamente autónomos en el hemisferio izquierdo, cada uno de los cuales está dedicado a procesar un aspecto diferente del conocimiento léxico".

Con lo anterior, se sustenta lo indicado por Johnson Laird, sobre que las teorías del significado no son suficientes para explicar las capacidades inferenciales y referenciales de la competencia léxica, ya que, según lo reportado

por Caramazza, el significado y la forma de las palabras estarían representados en áreas distintas del cerebro.

Si bien los avances en el conocimiento neurológico han dado nuevas perspectivas al estudio del lexicón mental, ha sido por medio de la psicolingüística desde donde se ha intentado explicar los lazos asociativos que las palabras presentan entre sí.

Una red semántica o esquema de representación en red es una forma de representación de conocimiento lingüístico en la que los conceptos y sus interrelaciones se representan mediante un grafo. Las redes semánticas son usadas, entre otras cosas, para representar mapas conceptuales y mentales.

“Una red semántica representa el conocimiento como un grafo en forma de red. Una idea, un evento, una situación u objeto tiene casi siempre una estructura compuesta; esto se representa en una red semántica mediante una correspondiente estructura de nodos que representan unidades conceptuales, y aristas direccionadas que representan las relaciones entre las unidades” (Lehmann, 1992, p. 2, citado en Echeverría, 2010).

Relaciones semánticas asociadas a fenómenos como hiponimia e hiperonimia, régimen semántico, y sinonimia y antonimia, entre otras se insertan en una red cognitiva que las personas tejen a lo largo de la experiencia y del aprendizaje (Baralo, 2007). Por ejemplo, los hablantes saben que blanco se asocia con claro, y se opone a negro y oscuro.

Urrutia (2003) señala que la concepción de redes semánticas con juegos de activación y propagación da cuenta de modelos híbridos que simbolizan estructuras de redes, las cuales cambian en forma dinámica según sus activaciones. Agrega que las redes semánticas pueden ser consideradas como predecesoras de modelos conexionistas de cognición, puesto que comparten elementos con la arquitectura neural en lugar de redes distintas.

En este sentido, es a través del Modelo Conexionista con el cual se podría explicar el proceso cognitivo del acceso léxico a través de las redes neuronales compuestas de nodos y conexiones entre dichos nodos.

El modelo postula que las unidades de palabras usadas con mayor frecuencia están conectadas más estrechamente con los nodos de nivel inferior, como los de rasgos fonológicos y ortográficos. "Cuando se activa un nodo o se realiza una conexión la activación se propaga en todas direcciones y se incrementan las representaciones que se parecen visual, fonológica y semánticamente, entre otras características, a la palabra buscada" (Urrutia, 2003).

En un grafo o red semántica los elementos semánticos se representan por nodos. Dos elementos semánticos entre los que se da la relación semántica que representa la red estarán unidos mediante una línea, flecha, enlace o arista. Cierta tipo de relaciones no simétricas requieren grafos dirigidos que usan flechas en lugar de líneas (Echeverría, 2010).

Dicho de otra forma, este modelo señala que la organización de las palabras está dada por la fuerza de las conexiones entre los nodos, los que, a su vez, se basan en asociaciones que se producen entre nodos de palabras y nodos de rasgos.



CAPÍTULO 3: MARCO METODOLÓGICO

3.1. DISEÑO DE INVESTIGACIÓN

Este trabajo se basó en un diseño mixto –cuantitativo-cualitativo–, no experimental y transversal, cuyo propósito fue realizar comparaciones y entregar descripciones del fenómeno estudiado –la competencia léxica–, tomando en cuenta las bondades de ambos enfoques metodológicos. El objetivo de la investigación fue describir la competencia léxica de estudiantes de periodismo, enmarcada en el discurso especializado escrito, el cual se operacionaliza en la forma de noticias informativas escritas, en tres dominios temáticos.

Para llegar a una descripción de competencia léxica se ha propuesto utilizar tres subestudios, que han sido empleados, según la literatura especializada, para aproximarse tanto a la descripción como a la configuración de un determinado nivel de competencia léxica.

Se realizó una aproximación cuantitativa para lograr la determinación de los objetivos referidos a riqueza léxica y a disponibilidad léxica. En tanto, las redes de asociación semántica a partir del léxico disponible si bien se nutren de datos duros obtenidos cuantitativamente, son el resultado de un enfoque cualitativo, ya que con estas redes no se busca la cuantificación numérica de los datos obtenidos, sino que la vinculación que existe entre los conceptos que crean la idea general frente a un centro de interés determinado y cómo factores asociados a la preparación profesional afectan a dicho imaginario.

3.1.1. ESTRUCTURA

La estructura del trabajo se llevó a cabo realizando tres tipos de estudios, los cuales fueron integrados posteriormente analizando sus relaciones. Para ello se efectuaron comparaciones para los aspectos medidos de la siguiente forma:

a) Disponibilidad léxica: léxico latente de centro de interés, Índice de Cohesión de cada centro; Índice de Disponibilidad Léxica resultante entre estudiantes agrupados por curso y entre estudiantes y profesionales.

b) Riqueza léxica: se efectuaron diversas comparaciones, la primera fue una general entre la riqueza léxica de los textos elaborados por los estudiantes de periodismo y los profesionales en los tres dominios; una segunda comparación entre los estudiantes agrupados por curso, finalmente, una tercera comparación entre los índices obtenidos según temática (ciencia y tecnología, política y deporte).

Como se indicó al inicio del capítulo, esta investigación se planteó como un estudio descriptivo de carácter mixto y, como tal, buscó especificar propiedades, características y rasgos importantes del fenómeno analizado (Hernández, Fernández y Baptista, 2003).

Según lo propuesto en la problemática sobre las relaciones semánticas, el análisis cualitativo de las redes semánticas, realizadas a partir de la disponibilidad léxica, intenta dar respuesta a interrogantes tales como: de qué forma se organiza el léxico disponible en los sujetos estudiados, qué relaciones semánticas se establecen entre los términos evocados, cómo evolucionan en el tiempo (entre

los cursos), cuáles son las diferencias entre estudiantes y profesionales, entre otras.

3.2. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

3.2.1. OBJETIVO GENERAL

El objetivo general de esta investigación es:

- Describir el perfil de la competencia léxica de estudiantes de periodismo en el discurso especializado escrito, a través de un modelo de estudio basado en disponibilidad, riqueza y acceso léxico.

3.2.2. OBJETIVOS ESPECÍFICOS

Para cumplir con el objetivo general se plantearon una serie de objetivos específicos que apuntan a cubrir cada componente del objetivo general, a su vez estos objetivos específicos se encuentran en concordancia con las preguntas de investigación planteadas en el siguiente punto.

1. Medir la disponibilidad léxica de la muestra en los centros de interés específicos: ciencia y tecnología, deporte y política.
2. Comparar la disponibilidad léxica entre estudiantes y profesionales.
3. Evaluar la riqueza léxica de los textos producidos por la muestra, para compararla entre los grupos de estudiantes y profesionales.
4. Identificar las diferencias de la riqueza léxica entre los estudiantes y profesionales. Identificar las redes semánticas de la muestra y

realizar comparaciones entre los grupos de sujetos y los centros de interés.

3.2.3. PREGUNTAS DE INVESTIGACIÓN

Para adentrarse en los objetivos específicos enunciados más abajo se proponen las siguientes preguntas:

- ¿Cuál es la disponibilidad léxica de la muestra estudiada?
- ¿Existen diferencias estadísticas en el IDL entre los grupos de estudiantes y profesionales?
- ¿En términos de diversidad y densidad léxica, cuál es la diferencia en cada uno de los centros de interés?
- ¿Cuál es la diferencia de los índices de riqueza presentada por los textos de la muestra de estudiantes en comparación con el corpus profesional?
- ¿Cómo se organizan las redes de asociación semántica en cada grupo de la muestra?

3.2.4. HIPÓTESIS DE TRABAJO

- H₁: El sexo no es una variable que influya en la riqueza léxica de los textos
- H₂: No existen diferencias significativas en la diversidad y densidad léxica entre hombres y mujeres, incluyendo las tres temáticas asociadas a la investigación.

- H₃: Existen diferencias significativas entre los grupos (estudiantes de primer año, estudiantes de cuarto año y profesionales) y la diversidad y densidad léxica de sus textos.
- H₄: Existen diferencias significativas en la riqueza léxica de los estudiantes de primer año según la dependencia del establecimiento educacional del cual egresaron de enseñanza media.

3.3. PARTICIPANTES/MUESTRA

Dada las características mencionadas del estudio, el cual consta de tres partes, la muestra la componen, desde el ángulo de la riqueza léxica, tres corpus temáticos de noticias informativas elaboradas por estudiantes de periodismo pertenecientes a dos niveles distintos y por profesionales, configurándose tres grupos en total: G1, alumnos de primer año; G2, alumnos de cuarto año y G3, profesionales.

A la vez, desde un segundo ángulo de la investigación, estos sujetos divididos en los mismos grupos participaron como informantes de las encuestas de disponibilidad léxica, conformando una muestra no probabilística propositiva. El foco de dicho grupo está definido según el interés de la investigación – estudiantes de periodismo y profesionales periodistas– sin embargo, no es una muestra aleatoria.

Finalmente, el último ángulo, se nutre del léxico obtenido en las encuestas de disponibilidad léxica, utilizando, específicamente, el índice de disponibilidad de

cada léxico, para configurar la red de asociación semántica de cada grupo de sujetos presente en la investigación.

Como se puede observar, los tres ángulos que componen la investigación se relacionan tanto con relación a quienes componen la muestra de informantes que participan, así como las producciones que realizan: las noticias según las temáticas requeridas, las pruebas de disponibilidad léxica y el léxico, derivado de estas últimas, para establecer las asociaciones semánticas.

Tabla 1. Cruce según sexo y grupo.

Tabla cruzada SEXO*Grupo al que pertenece					
		Grupo al que pertenece			Total
		G1	G2	G3	
Sexo	hombre	25	9	5	39
	mujer	33	11	15	59
Total		58	20	20	98

En resumen, los participantes en el estudio fueron 98 personas: 58 estudiantes de Primer Año, 20 de Cuarto Año y 20 profesionales. La **Tabla 1** ordena, además, a los grupos según sexo. Por otra parte, se obtuvieron datos sobre la procedencia administrativa del establecimiento educacional en el cual cursaron su enseñanza media. De los participantes, 28 cursaron sus estudios secundarios en un establecimiento cuya dependencia administrativa correspondió a privada; 26 a establecimientos municipalizados y 44 a establecimientos subvencionados.

Por lo tanto, la investigación queda conformada por las siguientes variables predictoras⁸ (independientes):

Tabla 2. Variables independientes y sus valores.

Variable	Etiqueta	Valores
SEXO	Género.	0= Hombre
		1= Mujer
GRUPO	Grupo al que pertenece la muestra.	1= G1 Primer año.
		2= G2 Cuarto año.
		3 = G3 Profesionales.
EDUC	Perteneencia administrativa de institución educacional de la cual egresó en Enseñanza Media.	1 = Privado.
		2 = Municipal.
		3 = Subvencionado.
TEMA	Temáticas en las cuales de realizaron las noticias.	1 = Ciencia y Tecnología.
		2 = Política.
		3= Deporte.

A continuación, se presenta en detalle cada una de estas muestras:

3.3.1. DETALLES DE LA MUESTRA EN CADA SUB-ESTUDIO

3.3.1.1. ESTUDIO DE DISPONIBILIDAD LÉXICA

La muestra en este estudio está compuesta por:

⁸ Debido a que este estudio es una investigación no experimental, lo más adecuado es utilizar la terminología "variables predictoras"; mientras que las variables independientes utilizan la terminología "variables criterio", también denominadas respuesta o resultado. (Herrera, Martínez y Amengual, 2011)

i) 78 Estudiantes de periodismo de la Universidad de Concepción.

ii) 20 Periodistas profesionales.

iii) La disponibilidad léxica fue obtenida mediante las pruebas tradicionales de disponibilidad léxica, utilizando cinco centros de interés, tres corresponden a las temáticas de los textos: ciencia y tecnología, política y deporte; mientras que los otros dos corresponden a los centros de interés analizados en el Proyecto Panhispánico de disponibilidad léxica: partes del cuerpo y medios de transporte.

3.3.1.2. ESTUDIOS DE RIQUEZA LÉXICA

Esta muestra está compuesta por tres corpus fundamentales producidos por cada grupo de sujetos, y divididos, además, en los tres ámbitos temáticos de estudio.

i) Corpus de noticias elaboradas por estudiantes de primer año de periodismo, con los dominios: política, deporte y ciencia y tecnología, compuesto por 121 textos.

ii) Corpus de noticias elaboradas por estudiantes de cuarto año de periodismo, con los dominios: política, deporte y ciencia y tecnología, compuesto por 44 textos.

ii) Corpus de noticias publicadas en medios de comunicación escritos por periodistas profesionales, en tres dominios: política, deporte y ciencia y tecnología, compuesto por 97 noticias.

3.3.1.3. ESTUDIO SOBRE RELACIONES LÉXICAS

La elaboración de las redes semánticas a partir del léxico disponible, en los tres ámbitos temáticos, se efectuaron mediante el programa Gephi. Para ello se trabajó con los resultados de las encuestas de disponibilidad léxica de los participantes de la muestra que en total reunió a más de ocho mil palabras.

3.4. PRUEBAS Y MEDICIONES

Para realizar el estudio sobre riqueza léxica, se propusieron dos vías: en primer lugar, evaluar la riqueza léxica de los textos producidos por los estudiantes y los profesionales y, en segundo lugar, identificar si existen diferencias -y de existir cuáles son- entre ambos grupos.

La riqueza léxica de los textos se obtuvo a través dos aproximaciones cuantitativas, previamente abordadas en el Marco Teórico, cuales son **la diversidad léxica**, es decir la relación entre el total de vocablos y los vocablos distintos de un texto; y, por otra parte, la **densidad léxica** de los textos periodísticos, es decir, la relación entre palabras con contenido semántico y el total de palabras del texto.

3.4.1. VARIABLES INDEPENDIENTES EN RIQUEZA LÉXICA

Las variables dependientes en los estudios de riqueza léxica fueron operacionalizadas de la siguiente manera:

Diversidad léxica: implica la relación entre la extensión del texto y el total de palabras diferentes de dicho texto. Variable cuantitativa.

Densidad léxica: relación entre las palabras de contenido semántico y la extensión del texto. Es una variable cuantitativa

3.4.1.2. PROCEDIMIENTO PARA MEDIR LAS VARIABLES

DIVERSIDAD LÉXICA Y DENSIDAD LÉXICA

La evaluación de la diversidad léxica siguió la línea tradicional de estudio, es decir se utilizó la fórmula *type/token ratio*, que en palabras sencillas significa obtener la relación entre el total de palabras distintas (*type*) y el total de palabras (*token*).

Este procedimiento se efectuó en cada uno de los textos, según la temática del corpus y según grupo muestral de sujetos. Además, los textos fueron analizados en cada grupo según extensión de la noticia, evitando así lo que señala la evidencia de diversas investigaciones, que los índices de relación entre las palabras distintas y el total de palabras disminuyen a medida que aumenta la extensión de un texto, tema que fue abordado en el Marco Teórico y que se refiere a la distorsión que puede ocurrir con los datos al comparar textos con distintas extensiones.

3.4.1.2.1 LEAD

Por otra parte, se planteó la idea de analizar la diversidad léxica del *lead* de las noticias. La relevancia de este punto radica en que, según la estructura piramidal invertida de una noticia, el periodista debe entregar la información más relevante del hecho noticioso en la parte que comprende el titular de la noticia y

el primer párrafo. Esto quiere decir que en el inicio de un texto –en este caso de una noticia escrita- la mayor cantidad de palabras distintas se encuentra dada en los primeros párrafos, decreciendo a medida que aumenta la extensión del texto, ya que los vocablos comienzan a repetirse con mayor frecuencia y la generación de nuevos es cada vez menos frecuente.

Para lograr este objetivo se realizó una estimación sobre cuántas palabras conformaban el *lead* de una noticia, a través de los siguientes pasos (Figura 3. Proceso para calcular extensión del lead (elaboración propia).):

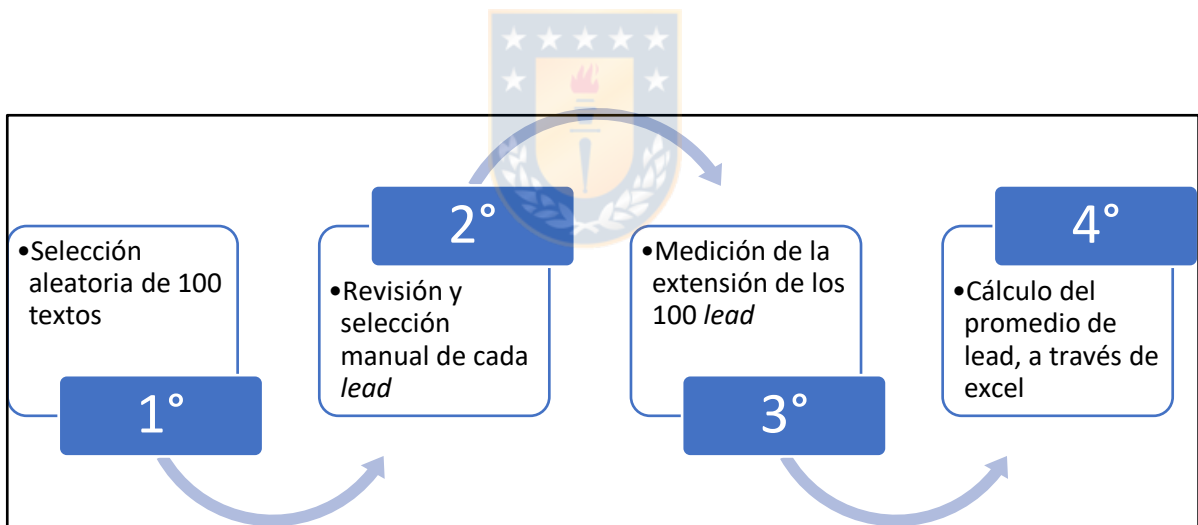


Figura 3. Proceso para calcular extensión del lead (elaboración propia).

El proceso arrojó como resultado que, en promedio el lead está compuesto por, aproximadamente 100 palabras, por este motivo es que, junto con la división en cuatro grupos según la extensión de los textos, se midió la diversidad de las 200 primeras palabras de cada texto, como un indicador de la diversidad léxica.

Hay que recordar que, en este grupo de palabras, el periodista debe dar respuesta a las cinco interrogantes (qué, cuándo, cómo, por qué, dónde), con las cuales se elabora un texto informativo.

3.4.1.2.2 DENSIDAD LÉXICA

En el caso de la densidad léxica, como ya se indicó, se utilizaron los mismos textos del corpus de noticias. Una vez obtenido el valor de la diversidad léxica, se procedió a ingresar a un script⁹, desarrollado para los fines de esta investigación, el listado de más de 300 palabras que no se consideraron como nocionales para no utilizarlas en el cálculo (Anexo 01). Este paso fue fundamental debido a que la densidad léxica responde a la relación de palabras con contenido semántico y el total de palabras de cada texto.

Al enfrentarse al análisis de una gran cantidad de textos surge la problemática, especialmente en el español, de cómo cuantificar y/o etiquetar una gran cantidad de palabras que pertenecen a categorías como verbos, adjetivos y sustantivos -entre otros-. El desarrollo de la lingüística de corpus y, especialmente, el desarrollo de nuevas herramientas tecnológicas ha permitido que con diversos *softwares* (WordSmith, SketchEngine, Antco, entre otros) se puedan efectuar estudios más finos relacionados con el léxico. Aun así, es necesario la *educación* de dichos programas, con el propósito de que puedan reconocer y marcar las categorías gramaticales a las que corresponde cada

⁹ El funcionamiento de script se explicará más adelante en este capítulo.

palabra, para luego realizar las operaciones matemáticas necesarias que entregan los índices de diversidad o densidad léxica.

3.4.1.2.3 DIVERSIDAD LÉXICA

Para el análisis del indicador de la diversidad léxica se trabajó en la elaboración de tres estrategias que permitieran minimizar el riesgo de distorsión en los datos debido a la extensión de los textos. Por lo tanto, junto con realizar una división de los textos, como se expondrá más adelante en el punto Características del corpus para el análisis de la riqueza léxica, se definió establecer pruebas complementarias que permitieran obtener datos con la menor distorsión posible. Las medidas que se establecieron para esto se presentan en la Figura 4. Medidas utilizadas para analizar riqueza léxica

Total de palabras	Contabiliza el total de palabras del texto. También llamado extensión o <i>token</i>
Palabras únicas	Es el total de palabras del texto sin contar sus repeticiones. También llamadas <i>type</i>
Palabras nocionales	Corresponde al total de palabras que quedan en el texto una vez eliminadas las stop word. En este caso son también conocidas como palabras nocionales o de contenido semántico.
Densidad léxica	Es el resultado de la relación entre el total de palabras del texto (extensión o tokens) y el número de palabras nocionales.
IAT	El Intervalo de aparición de palabras estima la aparición de palabras nocionales en un contexto de palabras funcionales: a mayor intervalo menor riqueza.
Diversidad léxica (DL)	Relación entre el total de palabras distintas y la extensión del texto.
DL muestra	Resultado de un muestreo aleatorio de la DL de un texto, con el propósito de revisar posibles distorsiones.
DL Lead	Diversidad léxica de las primeras 200 palabras de la noticia.

Figura 4. Medidas utilizadas para analizar riqueza léxica.

La propuesta para esta investigación, como ya se ha explicado, fue trabajar con un script desarrollado en el mismo contexto de la tesis, es por lo que, junto con la posibilidad de realizar un conteo automático del léxico en cada una de las noticias escritas, dicho *script* se enriqueció con la utilización de un listado de palabras que no se consideran como nocionales, gracias al cual se eliminan de las noticias estas palabras que no representan valor semántico en el texto. Posteriormente, se aplica la fórmula matemática que entrega el resultado de la densidad léxica de cada texto y el promedio de densidad léxica en el corpus en general.

Al realizar los procedimientos para obtener la densidad léxica, el resultado fue, en primer lugar, la disminución de palabras en la extensión de cada texto, evidenciando que cada noticia fue limpiada de las palabras sin contenido semántico. Una vez obtenida la cantidad total de vocablos de cada texto, se efectuaron las operaciones para obtener la proporción entre las palabras nocionales y la extensión total del texto aplicado a todo el corpus, divididos según la extensión de cada noticia, tal como en la medición de diversidad léxica.

3.4.1.3. CÁLCULO AUTOMÁTICO DE LA DIVERSIDAD LÉXICA Y LA DENSIDAD LÉXICA

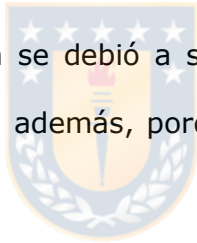
Con el objeto de facilitar el cálculo de la diversidad léxica y de la densidad léxica se desarrolló un *script* que procesa los textos en forma automática y permite realizar las operaciones necesarias para obtener los dos índices nombrados.

A continuación, se enunciarán las operaciones que realiza el *script* brevemente y se explicará cómo se utiliza éste desde el punto de vista del usuario.

3.4.1.3.1 OPERACIONES QUE REALIZA EL SCRIPT

El *script* se construyó utilizando Python 3 y NLTK¹⁰ (*Natural Language Toolkit*, disponible en <http://www.nltk.org/>). Dadas las características del lenguaje de programación utilizado, el *script* es multiplataforma, aunque para su desarrollo se utilizó un computador con Windows 10 de 64 bits. Su funcionamiento también se testeó en Linux, específicamente en la versión 16.04 de Ubuntu.

La elección de Python se debió a su gran versatilidad y potencia como lenguaje de programación y, además, porque NLTK se enfoca precisamente en éste.



El *script* recibe como entrada los textos que se quieren procesar y el listado de *stop words*. Los primeros deben estar en una carpeta que se encuentre en el mismo directorio del *script* (más adelante se explicará cómo se ingresa el nombre de ésta) y el segundo debe estar en un archivo de texto que también debe compartir la ubicación del *script*. Las operaciones realizadas en este paso buscan preparar los textos para su posterior procesamiento. Por ejemplo, se transforma la codificación en UTF-8 si no fuera ésta la que tuvieran, se eliminan del texto caracteres como signos de puntuación que no se van a considerar en el

¹⁰ *The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning* (<https://www.techopedia.com/definition/30343/natural-language-toolkit-nltk>).

procesamiento posterior y se transforman las palabras del texto en una lista¹¹ de Python para poder trabajar con ellas posteriormente.

A continuación, el *script* realiza las operaciones necesarias para calcular el número de palabras totales del texto, de palabras únicas (se considera sólo una instancia) y de palabras nocionales. Además, aplica a los datos las fórmulas necesarias para calcular el índice de densidad léxica, el intervalo de aparición de palabras nocionales y la diversidad léxica. En este último caso, calcula la diversidad léxica para el texto completo, para un número determinado de muestras aleatorias de palabras del texto (en el paso siguiente se explicará cómo el usuario puede definir la cantidad de palabras aleatorias y elegir el número de muestras), el valor promedio de los resultados de este número x de muestras y , finalmente, la diversidad léxica de las primeras palabras del texto, cuyo número también definido por el usuario, se busca que corresponda aproximadamente al *lead* de las noticias.

En el siguiente punto se explicará cómo opera el *script* desde el punto de vista del usuario.

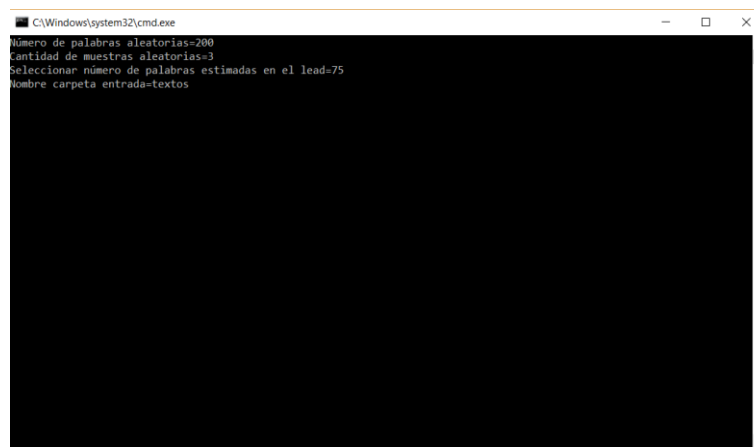
3.4.1.3.2 CÓMO UTILIZA EL SCRIPT EL USUARIO

El *script* puede ejecutarse desde cualquier ubicación dentro del computador. La única exigencia es que el junto al *script* se ubique el directorio

¹¹ “Las listas son conjuntos ordenados de elementos (números, cadenas, listas, etc). Las listas se delimitan por corchetes ([]) y los elementos se separan por comas” (<http://www.mclibre.org/consultar/python/lecciones/python-listas.html>).

que contiene los textos a procesar, los que deben estar en un archivo de texto plano cada uno.

Al ejecutar el *script* se abre una consola de comandos en la que se pide al usuario que ingrese el número de palabras que el *script* utilizará para realizar el cálculo de la diversidad léxica. Posteriormente, dado que se trata de muestras aleatorias y que por ende variarán en el valor que arrojen como resultado, el *script* ofrece la posibilidad de determinar cuántas muestras se utilizarán, manteniendo siempre el número de palabras determinado en el paso anterior. Luego, para realizar un cálculo diferente al de los pasos anteriores, solicita al usuario indicar el número aproximado de las palabras del texto que corresponden al *lead* de la noticia, que corresponden a las primeras palabras de un texto. Por último, se le pide al usuario que indique el nombre de la carpeta en que están ubicados los diferentes textos a procesar. Este último paso permite que el usuario pueda manejar diferentes carpetas con distintos textos y elegir con precisión sobre cuál quiere trabajar. En la Figura 5 se puede ver lo descrito:



```
C:\Windows\system32\cmd.exe
Número de palabras aleatorias=200
Cantidad de muestras aleatorias=3
Seleccionar número de palabras estimadas en el lead=75
Nombre carpeta entrada=textos
```

Figura 5. Preguntas que el script realiza al usuario.

Luego, el *script* realiza los procesamientos necesarios y entrega los resultados en forma separada para cada texto procesado. De aquí en adelante, la explicación se realiza enfocándose en un único texto procesado, con independencia de que el *script* puede trabajar simultáneamente con tantos como se desee.

Primero indica el número de palabras totales del texto y luego la cantidad de palabras únicas; es decir, para este último paso se considera sólo una instancia de cada palabra y no las repeticiones de la misma.

A continuación, entrega el número de palabras nocionales del texto. Para el cálculo de esto último se utiliza el mismo principio de las *stop words*¹². Éste se basa en la incorporación de un listado de palabras que deben excluirse de la búsqueda y hay muchos disponibles en internet para las distintas lenguas, aunque como señalan en Clips (2017) no hay un listado definitivo de *stop words*. Para la presente investigación, según ya se adelantó, se utilizó un listado adaptado de *stop words* que incluye más palabras de 300 que no son consideradas como nocionales. En todo caso, tal cual se señaló, para este caso tampoco es posible afirmar que este listado sea definitivo; por el contrario, es perfectible.

Luego, basándose en la cantidad de palabras que quedan tras excluir las del listado recién mencionado, que son consideradas como palabras nocionales, el *script* calcula el índice de densidad léxica dividiendo el total de palabras

¹² Las *stop words* se pueden traducir como *palabras de parada* o *palabras vacías*. El Centro de Investigación en Lingüística Computacional y Psicolingüística (Clips, por su nombre en inglés) las define de la siguiente forma: “*Stop words are words that are so common that they are often filtered out prior to, or after, processing of natural language data (text)*” (Clips, 2017. <https://www.clips.uantwerpen.be/pages/stop-words>).

nocionales por el total de palabras del texto y entrega dicho resultado al usuario. Posteriormente, calcula el intervalo de aparición de palabras nocionales (IAT) y lo presenta.

Finalmente, los últimos resultados que arroja el *script* son las cuatro variantes del índice de diversidad léxica que calcula. La primera de éstas es simplemente la diversidad léxica del texto procesado, considerando el texto completo y sin atender a la cantidad de palabras determinada para obtener una muestra aleatoria. Luego, sí considera el número de palabras de la muestra aleatoria y realiza este procesamiento para la cantidad de muestras aleatorias a calcular según el número que el usuario haya elegido en el paso descrito más arriba y entrega los valores obtenidos para cada una de estas muestras. A continuación, presenta el promedio de los valores obtenido para cada una de las muestras en el paso anterior. Por último, entrega el valor de la diversidad léxica para la cantidad de palabras que se definió como representativas del *lead* de la noticia, es decir, las primeras palabras del texto.

```
C:\Windows\system32\cmd.exe
Total de palabras (004_p_c.txt) = 474
Palabras únicas (004_p_c.txt) = 243
Palabras nocionales (004_p_c.txt) = 201
Densidad léxica (004_p_c.txt) = 0.4240506329113924
LAI (004_p_c.txt) = 2.3582089552238807
Diversidad léxica (004_p_c.txt) = 0.5126582278481012
Diversidad léxica muestra valores (004_p_c.txt) = [0.645, 0.62, 0.64]
Diversidad léxica muestra promedio (004_p_c.txt) = 0.635
Diversidad léxica 75 primeras palabras (004_p_c.txt) = 0.6266666666666667

Añadir salida al registro(s/n)=s
```

Figura 6. Resultados que entrega el script en la consola.

Como se puede observar en la última línea de la Figura 6, el *script* realiza una última pregunta al usuario, en relación con si quiere añadir la salida al registro. Lo anterior merece una explicación breve. Los resultados no sólo se entregan a través de la consola de comandos, lo que sería bastante incómodo ya que habría que copiarlos uno por uno para poder procesarlos en alguna aplicación de análisis de datos. Por lo mismo, el *script* también crea un archivo de texto plano llamado *salida.txt* que entrega los resultados ordenados y tabulados de izquierda a derecha, esto es separados en columnas: el primer texto procesado es el de la izquierda y los posteriores se ordenan hacia la derecha. A continuación, en la **iError! No se encuentra el origen de la referencia.** se presenta una muestra del archivo de salida para cuatro textos procesados simultáneamente.

Número de palabras aleatorias = 200

Cantidad de muestras aleatorias = 3

Nombre texto	Total palabras	Palabras únicas	Palabras nocionales	Densidad léxica	AT	Diversidad léxica (DL)	DL muestra	DL lead
01_p_c.txt	316	156	123	0,39	2,57	0,49	0,55	0,71
02_p_c.txt	315	170	138	0,44	2,28	0,54	0,61	0,67
03_p_c.txt	386	210	165	0,43	2,34	0,54	0,63	0,68
04_p_c.txt	474	243	201	0,42	2,36	0,51	0,64	0,63

12/01/2018 12:58

Figura 7. Resultado de un archivo de salida del script (Los valores solo consideran dos decimales para facilitar su presentación).

Para finalizar, la pregunta que se muestra en la Figura 7, permite añadir las distintas salidas que se vayan obteniendo cada vez que se utilice el *script* a un archivo de texto plano llamado *registro.txt*, que almacena el registro histórico de todas las salidas que se han obtenido para facilitar su consulta posterior. Al igual que en cada archivo de salida, los datos se almacenan con la fecha y hora de la ejecución del *script*.

3.4.1.4. CARACTERÍSTICAS DEL CORPUS PARA EL ANÁLISIS DE LA RIQUEZA LÉXICA

El corpus completo, que reúne los textos de las tres muestras de sujetos, está conformado por 262 noticias del género informativo. Dentro de las características de este tipo de noticias se encuentra su estructura, la cual corresponde a una pirámide invertida (definida en el capítulo del Marco Teórico) lo que significa, a grandes rasgos, que la información más importante de la noticia se encuentra entre el titular y el primer párrafo del texto.

Como ya se señaló, en el estudio participaron estudiantes de la Carrera de Periodismo de la Universidad de Concepción de primer y cuarto año, junto a periodistas profesionales en ejercicio. Los textos de los estudiantes fueron recopilados en el marco de las asignaturas *Producción de textos*, para primer año y *Producción de contenidos en tiempo real I y II*, para cuarto año. En el primer caso los textos corresponden a tareas evaluadas en el marco de la asignatura, en la cual los estudiantes debían redactar una noticia según la temática asignada por el docente, basada en datos obtenidos desde distintas fuentes y referencias de Internet.

En el caso de la asignatura de cuarto año, los textos fueron elaborados por los estudiantes para ser publicados en la revista *Entrelíneas*, que semanalmente publican los alumnos en papel digital, simulando así la dinámica de un medio real, en la cual se debe respetar el estilo, espacio (longitud del texto) y formato, entre otros aspectos. Otro punto importante es que en este caso los estudiantes deben publicar sus textos a partir de hechos y datos reales, obtenidos desde fuentes principales (por ejemplo, mediante entrevistas de hechos recientes y de interés).

El corpus elaborado en base a los textos de los estudiantes de primer año se compone de 121 noticias; de éstas, 50 corresponden a la temática ciencia y tecnología, 45 a política y 26 a deporte. El corpus construido en base a los textos de los alumnos de cuarto año está conformado por un total de 43 textos: 12 sobre ciencia y tecnología, 20 noticias sobre política, y 12 correspondientes al dominio deporte. En total participaron 58 estudiantes de primer año y 20 estudiantes de cuarto año.

Finalmente, como se indica en el Gráfico 1. Organización del corpus. se encuentra el corpus de noticias elaboradas por 20 periodistas. A los profesionales se les solicitó que seleccionaran tres noticias elaboradas por ellos en cada temática con el único requisito de que hubieran sido publicadas en los medios en los que trabajan. Este corpus quedó conformado de la siguiente manera: 97 textos en total, de los cuales 30 corresponden al dominio ciencia y tecnología; 32 a política y 35 a deporte.

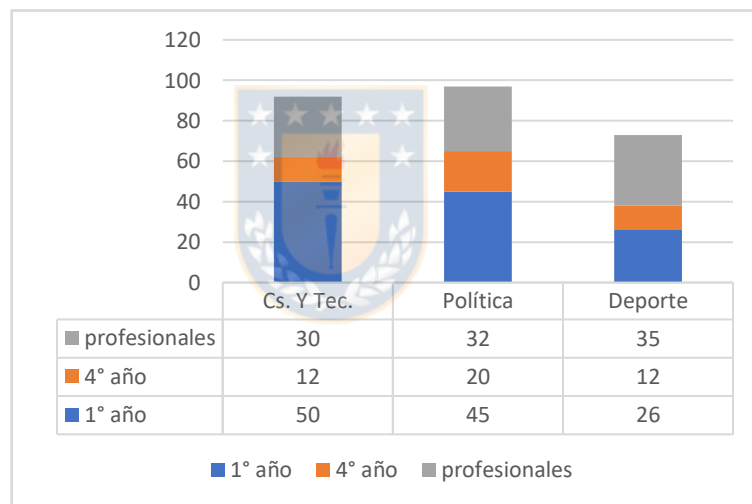


Gráfico 1. Organización del corpus.

En un principio se esperaba que tanto estudiantes como profesionales fueran capaces de realizar un texto por cada ámbito temático. Sin embargo, este aspecto solo pudo ser cumplido por los profesionales, quienes en su mayoría poseían registro de sus trabajos. Distinto fue el caso con los estudiantes quienes no lograron entregar a cabalidad los tres textos cada uno. En general, como

fueron noticias solicitadas en el contexto del aula y para evaluación, hubo estudiantes, particularmente los de primer año, que no cumplieron con los requisitos establecidos en la asignatura, entregando textos enmarcadas en solo dos ámbitos temáticos.

Los textos recolectados en el corpus de estudiantes fueron escritos en el marco de asignaturas dedicadas a fomentar la producción escrita. En el caso del corpus de primer año, los textos no tenían límite de extensión, por lo que la longitud de las noticias varió entre 300 y 2000 palabras; el requisito, sin embargo, era que se redactara con la estructura de pirámide invertida, es decir, que la información relevante se expusiera entre el titular de la noticia y su primer párrafo. Debido a que se trata de tareas propias de la asignatura ocurrió que no todos los estudiantes cumplieron con presentar un texto en cada temática: de los 58 estudiantes 50 escribieron sobre ciencia y tecnología, 25 sobre política y 25 sobre deporte.

Así también ocurre con las noticias redactadas por los 20 estudiantes de cuarto año para su periódico. En este caso la extensión de los textos varió entre 250 y 800 palabras. La explicación para esta variación radica en que, al ser textos publicados en un periódico estudiantil, cuyas características con similares a un diario normal los estudiantes deben ceñirse a un espacio limitado. Finalmente, las noticias elaboradas por los profesionales tuvieron una extensión desde las 300 hasta las 900 palabras.

Todos los textos fueron traspasados a formato **.txt*, con el objetivo de limpiar el archivo de datos que impediría su posterior estudio en el *script* desarrollado. El análisis se realizó comparando, en primer lugar, las temáticas (ciencia y tecnología, política y deporte) por cada uno de los grupos (alumnos de primer año, alumnos de cuarto año y profesionales). El objetivo de este proceso fue explorar las diferencias que se podrían encontrar en cada uno de los campos temáticos por cada grupo de sujetos tanto a nivel de diversidad como de densidad léxica.

Para cautelar que la extensión de los textos no distorsionara las comparaciones, problemática recurrente en los estudios cuantitativos de los textos la cual que fue explicada en el Marco Teórico, el corpus fue ordenado en distintos grupos según su extensión. A cada uno de estos grupos de textos ordenados según el número total de palabras los denominaremos *cluster*, y la nomenclatura para diferencia a cada uno especificará, en primer lugar, si representa al *cluster* de entre 250 y 500 palabras (X1); entre 501 y 750 palabras (X2); entre 751 y 950 palabras (X3); finalmente, de 951 palabras y más (X4), seguido del grupo muestral al que pertenece y, finalmente, la temática a la que corresponde, tal como se muestra en la Figura 8. El propósito de lo anterior fue realizar comparaciones entre textos que cuenten con la una extensión similar, como una forma de controlar la variación en los resultados.

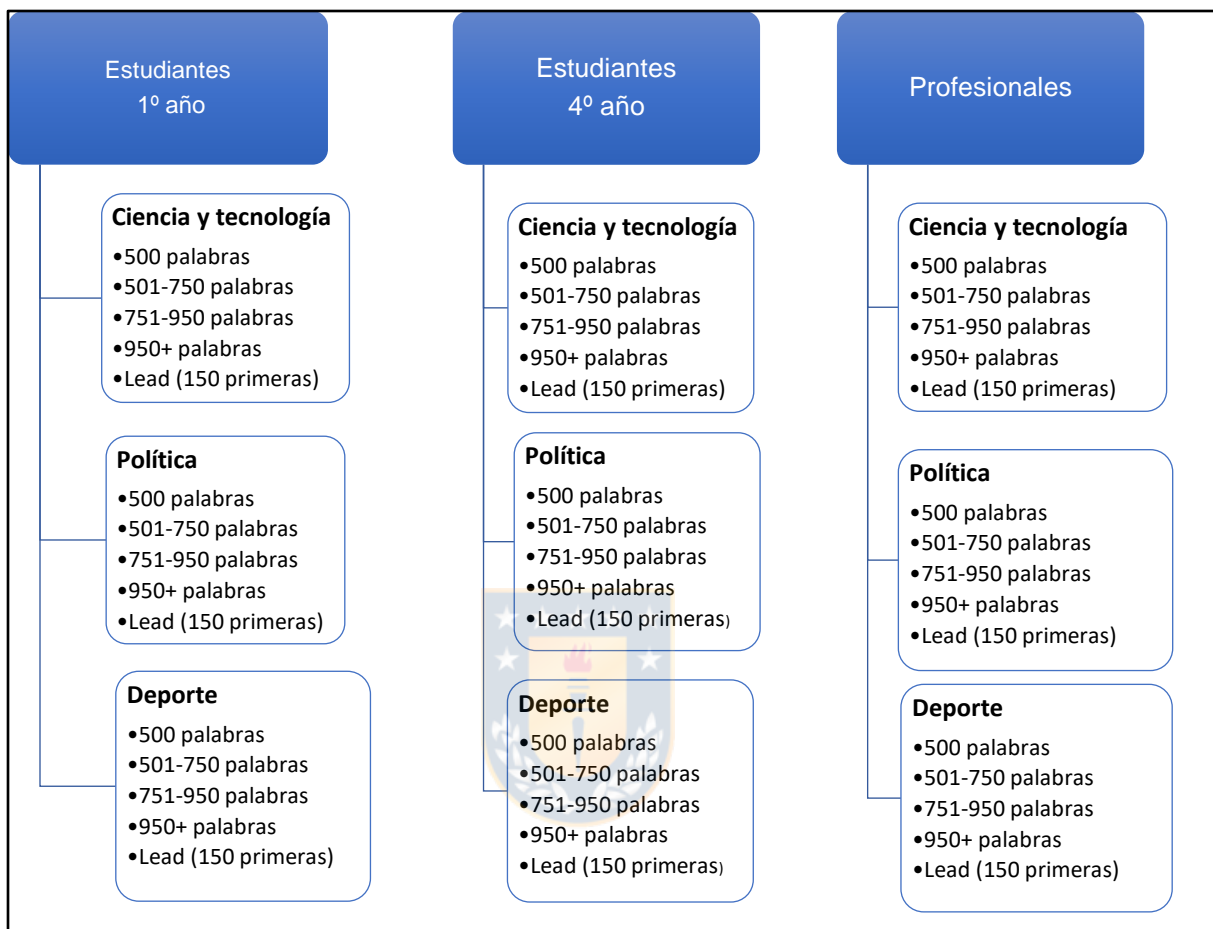


Figura 8. Esquema de análisis de diversidad y densidad léxica.

Para evaluar la densidad léxica se utilizó el mismo script que, como ya se explicó, para este cálculo consideró el listado de *stop words*, también llamadas en español palabras vacías, palabras gramaticales o funcionales (López Morales, 2002), debido a que carecen de significado por sí solas, dentro de ellas se encuentran categorías como artículos, preposiciones, conjunciones y pronombres.

El objetivo de utilizar un listado de palabras vacías en el script, previo a realizar el análisis numérico, fue limpiar de palabras gramaticales los textos,

dejándolos sólo con palabras de contenido semántico: verbos, adverbios, sustantivos, adjetivos calificativos. Además, se quitaron los verbos ser, estar y tener, junto a sus flexiones más comunes, debido a que su sobreutilización en los textos periodísticos, delatan pobreza léxica (Anexo 01: Listado de stop words para el análisis de densidad léxica).

Según lo indicado en el marco teórico, la diversidad léxica corresponde a la relación entre palabras distintas y el total de vocablos de un texto, mientras que la densidad léxica implica la relación entre las palabras con contenido semántico que contiene el texto y su extensión.

3.4.2. DISPONIBILIDAD LÉXICA

Como se ha mencionado en el Marco Teórico, la disponibilidad léxica surgió en Francia como un intento de crear un método para la enseñanza del francés elemental para extranjeros. Con el tiempo y el desarrollo de diversas investigaciones la disponibilidad léxica ha servido para la recogida y posterior análisis del léxico disponible de una determinada comunidad de habla (Saralegui y Taberner, 2008).

En esta investigación se ha considerado el léxico disponible como un elemento que puede aportar datos relevantes para la competencia léxica de estudiantes, sobre todo al enmarcar este tipo de estudio al ámbito del discurso especializado que desarrollan estudiantes de periodismo y periodistas.

Para realizar esta área de la investigación se siguió la línea tradicional, mediante una **prueba escrita de léxico disponible**, cuyo objetivo es medir el

léxico latente de cada sujeto en los centros de interés específicos para este estudio (ciencia y tecnología, política y deporte) además de dos centros que tradicionalmente se abordan en el Proyecto Panhispánico de Léxico Disponible (partes del cuerpo, medios de transporte).

Como se indicó al inicio de este capítulo, la muestra que participó en las encuestas de disponibilidad léxica fue la misma que actuó como informante y generadora de las noticias analizadas en el corpus de riqueza léxica: 58 estudiantes de primer año, 20 de estudiantes de cuarto año y 20 periodistas.

Cada participante que formó parte de la muestra compuesta por estudiantes debía completar una primera hoja con datos personales: nombre, edad, año de ingreso a la universidad, ciudad/comuna de procedencia. Mientras que en la muestra compuesta por profesionales los datos personales fueron: nombre, edad, universidad, año de egreso y ciudad/comuna donde reside. Luego, en las hojas posteriores, aparecía como encabezado cada centro de interés en los cuales debía responder cada sujeto (Anexo 02: Pruebas de disponibilidad léxica).

La prueba se tomó siguiendo el siguiente protocolo:

1. El investigador se presenta, señala que se trata de una actividad enmarcada en una investigación doctoral y reparte las pruebas -en formato papel- boca abajo en cada escritorio, solicitándoles no dar vuelta los documentos.

2. Se les solicita completar solo los datos personales y, una vez terminado, no dar vuelta la página hasta que todos los participantes estén listos.
3. Se les explica que cada hoja posterior tiene un concepto y que la actividad consiste en escribir en cada uno todas las palabras que asocien a ellos.
4. Se les indica que para lo anterior tienen un tiempo de dos minutos, los cuales serán indicados por el investigador.
5. Se señala que no importa la cantidad de palabras que haya anotado, ni si cometió faltas de ortografía o errores al escribir.
6. Se consulta si existen dudas.
7. Finalmente, se procede a la toma de la prueba cronometrada.

3.4.2.1. PROCEDIMIENTO PARA MEDIR LA DISPONIBILIDAD

LÉXICA

Cada prueba fue transcrita a una planilla *Excel*, ordenadas según grupo muestral y centro de interés. Posteriormente, los datos fueron configurados bajo el formato **.csv* y **.txt* según lo requerido para su procesamiento.

El procedimiento se realizó transcribiendo, en primer lugar, los datos obtenidos en cada hoja de respuesta de la prueba de disponibilidad léxica¹³, que

¹³ Ver anexo 02, con el diseño de las pruebas para los estudiantes y para los profesionales. Ambas difieren en los datos requeridos debido a que a los profesionales ya no están en la universidad, cambiando dicha pregunta por el medio en el que se desempeña.

grafías diferentes (por ejemplo: fútbol, football, futbol), otras en plural o singular (diputado, diputados), así también con diferente género (presidente, presidenta); además de errores ortográficos, en el caso en que los sujetos escribieron más de una palabra para referirse a un tema (e.g. en el centro de interés *política*, un sujeto indicó como vocablo "bien común"), estos fueron procesados con un guion de unión entre las grafías involucradas ("bien-común").

Todas las palabras fueron transcritas en su forma original, posteriormente, en un nuevo archivo, los datos fueron estandarizados siguiendo las reglas de la Figura 10 :

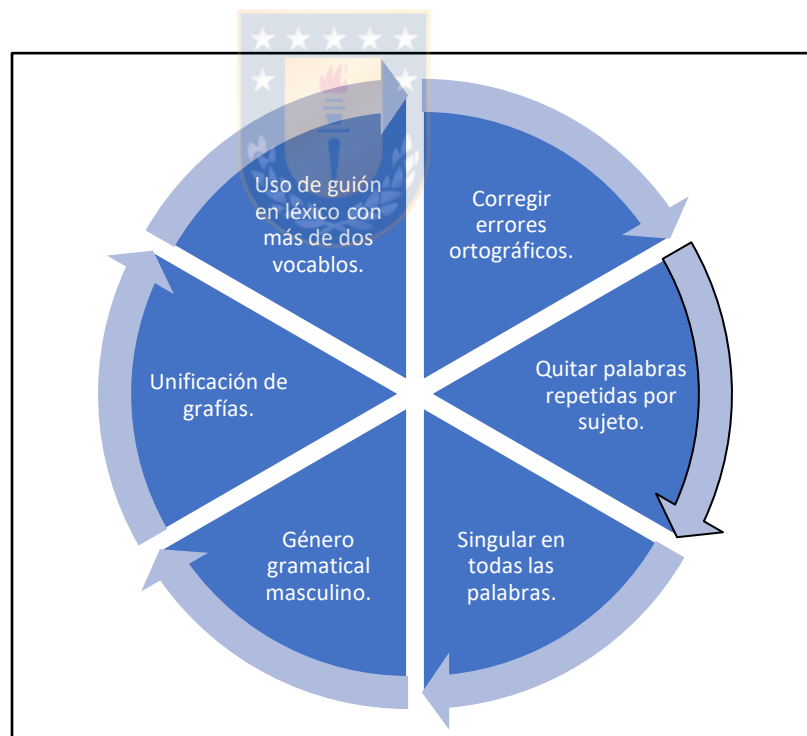


Figura 10. Depuración del corpus de léxico disponible

Para efectuar el análisis y cálculo de los datos ya depurados se utilizaron dos herramientas: *Lexmath* y *Dispogen II*.

La plataforma web de *Lexmath*, un sistema hipermedia adaptativo que forma parte del proyecto Fondecyt 1120911 "Disponibilidad Léxica Matemática en estudiantes de enseñanza media y su aplicación en hipermedios adaptativos", auspiciado por la Comisión Nacional de Ciencia y Tecnología, CONICYT. Entre sus objetivos se encuentra el de cuantificar y describir el léxico de estudiantes educación media en el área de la matemática.

Si bien el área temática y población hacia la que está dirigida esta herramienta no se ajusta a los propósitos de este trabajo, *Lexmath* ha sido validada estadísticamente y sus resultados publicados en diversas revistas científicas de corriente principal (Salcedo y del Valle, 2016; Salcedo, Ferreira y Barrientos, 2013), por lo tanto, es una herramienta confiable para obtener los datos cuantificadores sobre léxico latente que se asocien a cualquier centro de interés, al entregar -sin discriminar la temática-, los índices asociados a la disponibilidad léxica.

También se utilizó el *software* *Dispogen II*, creado por el lingüista Max Echeverría en 2005. Se trata de una aplicación realizada en MatLab, programa especializado en cálculos de matrices y en análisis estadístico de datos de multivariantes. *Dispogen II* tiene la ventaja de realizar los cálculos de índices y frecuencias, entre los primeros el índice de cohesión.

Dispogen II calcula los índices de disponibilidad léxica a través de la fórmula de Strassburguer y López Chávez, creada en 1987. Ésta tiene la ventaja de trabajar con un exponente, cuya función es asintótico a cero; es decir, nunca habrá un resultado cuyo valor sea cero. De este modo, los valores resultantes mantienen su capacidad discriminante (López Chávez y Strassburger, 1991).

Ambos softwares entregan datos comunes, como el índice de cohesión, el promedio de respuestas, y el número de palabras distintas. Se utilizaron ambos ya que *Lexmath* ofrece la posibilidad de estudiar cada centro y grupo por separado, además genera un archivo especial para ser leído por *Gephi*, el cual analiza redes de asociación semánticas. *Dispogen*, por su parte, entrega el índice de disponibilidad léxica personal (que no está presente en *Lexmath*) y la posibilidad de agrupar a todos los sujetos de la muestra en un solo archivo.

3.4.2.2. INDICADORES EN ÍNDICE DE DISPONIBILIDAD LÉXICA

Los resultados de la disponibilidad léxica se encuentran relacionados con tres *estadígrafos léxicos*, que corresponden a las medidas cuantitativas que más aportan a la determinación de la riqueza léxica en el ámbito de la disponibilidad léxica (Urzúa, Sáez y Echeverría, 2006):

- **Promedio de respuestas** (XR), indica el promedio de respuestas en un centro de interés de un grupo en particular.
- **Total de palabras diferentes** (PD): cantidad total de palabras diferentes conocida por la muestra.

- **Índice de cohesión (IC)**, es una interpretación de los índices anteriores, dado a que mide el grado de coincidencia de las respuestas.
- **Índice de Disponibilidad léxica individual (IDLI)**: grado de aportación de un individuo dentro de los listados generales de disponibilidad.

3.4.3. REDES DE ASOCIACIÓN

El análisis de las redes se efectuó a partir de los resultados obtenidos por los 98 sujetos que participaron de la muestra, al igual que en el análisis de la riqueza léxica y el de la disponibilidad léxica, en la muestra participaron tres grupos: G1, compuesto por estudiantes de primer año; G2, con estudiantes de cuarto año y G3, por periodistas profesionales.

Luego de utilizar los softwares para el cálculo de disponibilidad léxica, Lexmath y Dispogen II, los resultados de las asociaciones obtenidas fueron analizada, en primer lugar, por el software Dispografo¹⁴, el cual, entre otras herramientas, permite seleccionar cada centro de interés y realizar una “poda” de nodos y aristas que resulten poco significativa al investigar las relaciones semánticas de un centro. Además, permite la selección de variables de interés para el estudio; cabe señalar que las variables que se han definido como predictoras, corresponden a sexo, grupo y dependencia administrativa de la institución educacional de la cual egresó el sujeto.

¹⁴ Software desarrollado por Echeverría et al. (2008).

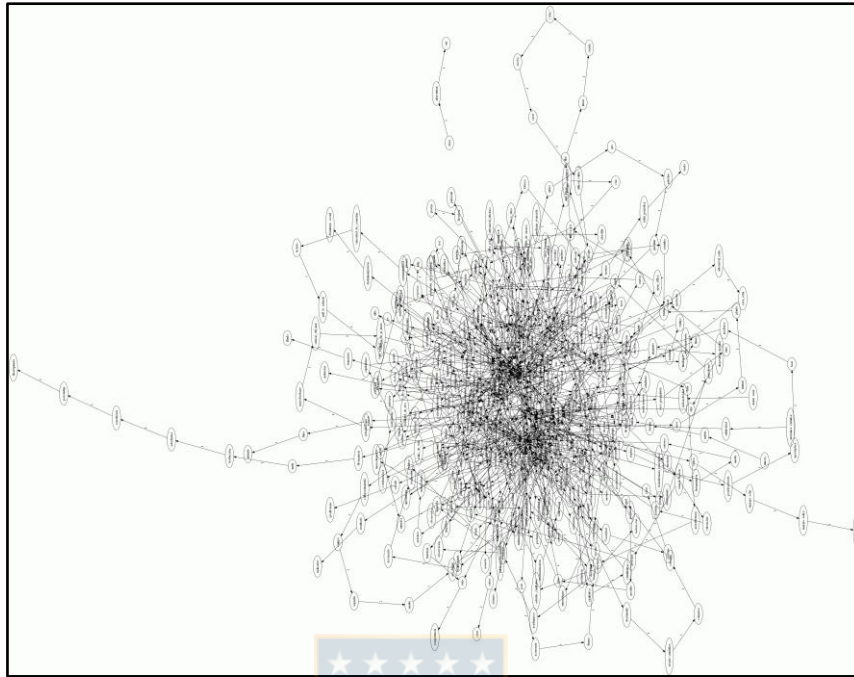


Figura 11. Imagen del centro de interés ciencia y tecnología, proveniente de Dispografo.

Dispografo, como resultado del procesamiento, entrega una salida con dos archivos. El primero corresponde a una imagen en *.gif*, la cual permite visualizar los nodos y aristas de la red. Se trata de una imagen fija –no se puede manipular–, cuyo ordenamiento corresponde a una colección de puntos u objetos llamados nodos, donde cada nodo es un concepto del corpus de disponibilidad léxica (ver Figura 11). La imagen también está compuesta por aristas, que representan las relaciones semánticas entre los vocablos disponibles.

El segundo archivo emanado por Dispografo es uno de tipo **.dot*, el cual permite, entre otros usos, transformar los datos en gráficos para su análisis en otro tipo de softwares de análisis gráficos.

Estos últimos archivos, separados y ordenados según centro de interés, en primer lugar, y, luego, según las variables indicadas en el párrafo anterior, fueron analizados en el *software* libre Gephi 0.9.2, herramienta desarrollada para la visualización y el estudio de redes.

Al importar los archivos *.dot en Gephi, se obtiene una red de nodos ordenados de manera aleatoria (Figura 12).



Figura 12. Red del centro de interés ciencia y tecnología, en Gephi.

Entre las utilidades que entrega Gephi se encuentra la posibilidad de ordenar los nodos según diversos algoritmos que permiten al investigador explorar las asociaciones de los nodos desde diversas perspectivas.

3.4.3.1. INDICADORES

El análisis de las redes semánticas se basa en la *Teoría de Grafos*, disciplina de las matemáticas discretas y de las ciencias de la computación, que se ha utilizado para estudiar una gran variedad de características y propiedades que surgen de las diferentes formas de relación para un conjunto de objetos. Los grafos son una herramienta que permite modelar relaciones de esta naturaleza (Pelegrín, Cánovas y Fernández, 2009).

Para el caso de las redes semánticas resulta de gran utilidad debido a que las asociaciones léxicas funcionan como una red, activando vocablos -*nodos* en la teoría de grafos-, los cuales se relacionan con otros, esta asociación es definida como una arista, que implica la comunidad o relación con otros nodos.

Para el análisis se utilizarán algunas métricas y estadísticos que permitirán no solo ordenar la red de una manera más amigable, sino que relevarán distintos tipos de asociaciones presentes en una red semántica.

- I. **Distribución de grado:** el grado de un vértice o nodo se refiere al número de conexiones asociadas a un nodo.
- II. **Coeficiente de *clustering*:** También definido como coeficiente de agrupamiento, se interpreta como la probabilidad de que dos nodos escogidos aleatoriamente que compartan un vecino tengan un enlace entre ellos. Como ejemplo, un nodo perteneciente a una red altamente conectada presenta un coeficiente de *clustering* cercano a 1.

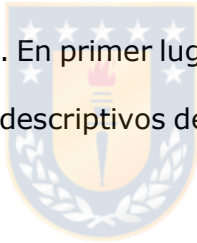
III. **Comunidades:** se define como la división de los nodos de una red en subgrupos dentro de los cuales las conexiones, o aristas, entre los nodos que la forman son densas o numerosas, pero entre las comunidades estas aristas son débiles o escasas.



CAPÍTULO 4: RESULTADOS Y ANÁLISIS DE DATOS

4.1. ANÁLISIS DESCRIPTIVOS DE DATOS

La finalidad de los análisis descriptivos es definir, clasificar o categorizar fenómenos de interés, ver cómo son y cómo se manifiestan, intentando especificar propiedades, características y perfiles de personas, grupos, comunidades, etc. (Herrera, Martínez y Amengual, 2011). A continuación, se presentan los resultados obtenidos con los análisis descriptivos de los tres ámbitos de esta investigación. En primer lugar, se encuentran los datos de riqueza léxica, luego, los estadísticos descriptivos de la disponibilidad léxica y, finalmente, los de asociación semántica.



Posteriormente, se presentan estadísticos inferenciales, con los cuales se presentan las relaciones obtenidas entre los tres ámbitos de estudio. Para llevar adelante este análisis se utilizó el software de análisis estadístico IBM SPSS 23.

4.1.1. RIQUEZA LÉXICA

El primer objetivo de este punto es exponer los resultados obtenidos en un nivel descriptivo de la riqueza léxica de los textos, con la idea de obtener una mirada general del corpus. Se expondrán los datos ordenados según género, grupo y pertenencia de establecimiento educacional en cada una de las temáticas

estudiadas. Sin embargo, y para entregar una primera aproximación en términos numéricos globales, es importante indicar que el corpus de los estudiantes de primer y cuarto año está compuesto por 106.286 *tokens*, 12.698 correspondían a *types* o palabras únicas y de ellas, 12.541 a palabras nocionales (PN) o con contenido semántico. En el caso de los profesionales el corpus está compuesto por 37.989 *tokens*, de ellos 6.751 son palabras únicas y 6.602 corresponden a palabras nocionales.

Como se indicó en el capítulo anterior, no todos los estudiantes cumplieron con entregar un texto de cada tema, es por ello por lo que a continuación se presentan los detalles del corpus según grupo ordenados por cada temática.



Tabla 3. Resumen de casos según tema y grupo.

	Grupo	Casos					
		Válido		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Corpus	G1	47	81%	11	19%	58	100%
Ciencia	G2	11	55%	9	45%	20	100%
	G3	20	100%	0	0%	20	100%
Corpus	G1	38	65,5%	20	34,5%	58	100%
Deporte	G2	20	100%	0	0%	20	100%
	G3	20	100%	0	0%	20	100%
Corpus	G1	55	94,8%	3	5,2%	58	100%
Política	G2	20	100%	0	0%	20	100%
	G3	20	100%	0	0%	20	100%

Como se indica en la Tabla 3, en el corpus "Ciencia", 47 de los 58 alumnos de primer año que participaron en el estudio entregaron los textos requeridos y, por ende, forman parte del corpus; en el caso de los estudiantes de cuarto año 11 de los 20 alumnos cumplieron y, por último, los 20 profesionales entregaron sus noticias sin excepción. El corpus "Deporte" quedó conformado por los textos de 38 estudiantes de primer año, los 20 de cuarto año y los 20 profesionales; éste fue el corpus en el que participó un menor número de estudiantes de primer año. Finalmente, en el corpus "Política" se recibieron las noticias de 55 estudiantes de primer año y las de los 20 participantes de cuarto año, así como los 20 que pertenecían al grupo de los profesionales.

En tanto, el ordenamiento según sexo en cada una de las temáticas quedó establecido según lo indicado en la Tabla 4:

Tabla 4. Resumen de casos según sexo y corpus.

	Sexo	Casos					
		Válido		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Corpus Ciencia	hombre	31	79,5%	8	20,5%	39	100%
	mujer	47	79,7%	12	20,3%	59	100%
Corpus Deporte	hombre	30	76,9%	9	23,1%	39	100%
	mujer	48	81,4%	11	18,6%	59	100%
Corpus Política	hombre	37	94,9%	2	5,1%	39	100%
	mujer	58	98,3%	1	1,7%	59	100%

En el estudio participaron 39 hombres y 59 mujeres. El desglose en este nivel de la muestra, como se presenta en la Tabla 4, señala que, de los casos válidos, es decir de quienes entregaron sus textos y se incluyeron en el corpus "Ciencia", 31 son hombres (79,5%) y 47 mujeres (79,7%). En "Deporte", el 76,9% son varones (30 participaron) y el 81,4% damas (48); mientras que en el corpus "Política", los 37 hombres que facilitaron sus noticias corresponden al 94,9% de los casos válidos y las 58 mujeres que participaron corresponden al 98,3% de la muestra.

Un tercer punto que se incluyó fue el tipo de establecimiento educacional (privado, municipal o subvencionados) del cual egresaron los sujetos que participaron de la muestra.

Tabla 5. Resumen de casos según tipo de institución y corpus.

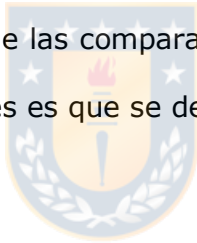
	Tipo de Institución Educativa	Casos					
		Válido		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Corpus Ciencia	privado	23	2,1%	5	17,9%	28	100%
	municipal	24	92,3%	2	7,7%	26	100%
	subvencionado	31	70,5%	13	29,5%	44	100%
Corpus Deporte	privado	28	100%	0	0%	28	100%
	municipal	16	61,5%	10	38,5%	26	100%
	subvencionado	34	77,3%	10	22,7%	44	100%
Corpus Política	privado	28	100%	0	0%	28	100%
	municipal	26	100%	0	0%	26	100%
	subvencionado	41	93,2%	3	6,8%	44	100%

La Tabla 5 resume cómo se ordena el corpus según el tipo de institución educativa del que egresó el sujeto que aportó con su noticia según el corpus temático. En "Ciencia" fueron 23 los sujetos de establecimientos privados que contribuyeron con sus textos, 24 sujetos provenientes del sistema municipal y 31 del subvencionado. En "Deporte" del total de sujetos, 28 provenían de instituciones educativas privadas, 16 de establecimientos con dependencia administrativa municipal y 34 subvencionados. En "Política" 28 participantes eran de colegios privados, 26 de establecimientos municipalizados y 41 sujetos de instituciones subvencionadas.

Tabla 6. Recuento según extensión de textos.

		Frecuencia	Porcentaje
Válido	X1= 250-500	145	55,6%
	X2= 501-750	63	24,1%
	X3= 751-1000	34	13%
	X4= 1001+	19	7,3%
Total		261	100

Finalmente, la Tabla 6 presenta el recuento del corpus según el *cluster*¹⁵ de extensión al cual pertenece. Esta exposición de datos cobra relevancia para comprender el real alcance de las comparaciones, debido a que se quiere evitar distorsiones en las mediciones es que se debe conocer a qué grupo de extensión corresponde cada noticia.



¹⁵ La tabla se ha dejado con la cantidad de palabras a la que corresponde cada grupo con el fin de facilitar la lectura y comprensión del cluster.

4.1.1.1. RESULTADOS GLOBALES DE RIQUEZA LÉXICA

i) VARIABLE SEXO

A continuación, se presentan los datos descriptivos, ordenados según la variable "sexo" de los resultados obtenidos en el análisis global de los textos. Esto quiere decir que se ha realizado el análisis del total de los 261 textos que componen el corpus de noticias, en sus tres ámbitos temáticos.

Tabla 7. Estadígrafos descriptivos sobre Densidad léxica

Sexo		Estadístico	Error estándar
Hombre	Media	0,4221	0,00719
	Varianza	0,005	
	Desviación estándar	0,07264	
Mujer	Media	0,4361	0,00696
	Varianza	0,008	
	Desviación estándar	0,08802	

La Tabla 7 presenta que los datos obtenidos de la variable Densidad léxica, indicador que mide la relación de palabras con contenido semántico dentro del texto. Los resultados indican que el promedio obtenido por los hombres (0,42) no difiere en gran medida del obtenido por las mujeres (0,43). Sí existen mayores diferencias en la varianza (hombres 0,005; mujeres 0,008), aunque la desviación estándar de ambos tiene sólo una diferencia de 0,01.

Tabla 8. Estadígrafos descriptivos de Diversidad léxica total.

Sexo		Estadístico	Error estándar
hombre	Media	0,4988	0,00612
	Varianza	0,004	
	Desviación estándar	0,06178	
mujer	Media	0,491	0,00497
	Varianza	0,004	
	Desviación estándar	0,06292	

La Tabla 8 muestra las medias obtenidas por hombres y mujeres, para la variable diversidad léxica (la que mide la relación entre el número de palabras diferentes y el total de los textos). El promedio (media) presentado por los hombres fue de 0,49 al igual que en las mujeres. Lo anterior significa que los promedios de riqueza léxica en el indicador diversidad léxica total de las noticias (es decir, sin diferenciar entre la extensión de los textos ni la temática), tanto hombres como mujeres no presentan grandes diferencias, lo mismo ocurre en la varianza, donde ambas varían en 0,004; y en su desviación estándar, la cual corresponde a 0,06.

Tabla 9. Estadígrafos descriptivos de Diversidad léxica aleatoria.

Sexo		Estadístico	Error estándar
Hombre	Media	0,6063	0,0038
	Varianza	0,001	
	Desviación estándar	0,03842	
Mujer	Media	0,5963	0,00302
	Varianza	0,001	
	Desviación estándar	0,0382	

Por otra parte, el indicador "diversidad léxica aleatoria", en la Tabla 9, muestra los puntajes promedio que obtuvieron tanto hombres como mujeres, cuando la diversidad léxica fue medida al azar, con 200 palabras, en cada noticia.. Las medias vuelven a presentar valores similares en ambos sexos: 0,60 en hombres y 0,59 mujeres; así como sus varianzas y la desviación estándar.

Tabla 10. Estadígrafos descriptivos de Diversidad léxica del *lead*

Sexo		Estadístico	Error estándar
Hombre	Media	0,6961	0,00615
	Varianza	0,004	
	Desviación estándar	0,06206	
Mujer	Media	0,6784	0,00486
	Varianza	0,004	
	Desviación estándar	0,06148	

Una de las preguntas de investigación está orientada a analizar la riqueza léxica que pueden alcanzar los estudiantes en el *lead* de las noticias. Como se ha explicado tanto en el marco teórico como en el metodológico, el *lead* es la parte de la noticia informativa que contiene la información más importante del hecho que el redactor quiere dar a conocer, por lo tanto, en este contexto surgió la idea de realizar un análisis a esta estructura en específico. Los indicadores descriptivos (Tabla 10) señalan que la media de diversidad es mayor en los hombres (0,69) que en las mujeres (0,67), con similares valores tanto en su varianza como en su desviación estándar.

Con lo anterior, se puede aventurar que se confirmaría una de las hipótesis planteadas en el marco metodológico, la cual postula que el género no es una variable que influya en la riqueza léxica. Esta hipótesis será contrastada más adelante mediante una prueba estadística apropiada a la normalidad de la muestra y al tipo de variable.

i.i) PRUEBA NO PARAMÉTRICA PARA SEXO

Una vez presentado los resultados para esta variable, se efectuó la prueba estadística U Mann-Withney, apropiada para el estudio de muestras independientes cuyos resultados no tienen una distribución normal, y que calcula las diferencias estadísticas entre las medianas de los resultados y así explorar la existencia de diferencias estadísticas entre hombres y mujeres.

Tabla 11. Resumen del contraste de hipótesis.

	Hipótesis nula	Prueba	Sig.	Decisión
1	Las medianas de Densidad léxica son las mismas entre las categorías de Sexo.	Prueba de la mediana para muestras independientes	,204	Conserve la hipótesis nula.
2	Las medianas de Diversidad léxica total texto son las mismas entre las categorías de Sexo.	Prueba de la mediana para muestras independientes	,990	Conserve la hipótesis nula.
3	Las medianas de Diversidad léxica medido aleatorio son las mismas entre las categorías de Sexo.	Prueba de la mediana para muestras independientes	,157	Conserve la hipótesis nula.
4	Las medianas de Diversidad léxica lead son las mismas entre las categorías de Sexo.	Prueba de la mediana para muestras independientes	,136	Conserve la hipótesis nula.
Se muestran significaciones asintóticas. El nivel de significación es ,05.				

La propuesta es que entre hombres y mujeres no existen diferencias estadísticamente significativas en la riqueza léxica (Tabla 11). Debido a que la prueba señala por defecto que la hipótesis nula considera que no hay diferencias significativas entre sus medianas.

ii) *VARIABLE “GRUPO”*

A continuación, se exponen los estadígrafos descriptivos analizados según “Grupo”. Al igual que en el punto anterior, esta variable fue analizada con la totalidad del corpus, sólo organizado mediante los tres grupos que participaron con sus noticias, sin distinción de la temática de éstas.

Tabla 12. Estadígrafos descriptivos Densidad léxica según grupo.

Grupo		Estadístico	Error estándar
G1	Media	0,3697	0,00363
	Varianza	0,002	
	Desviación estándar	0,03989	
G2	Media	0,4355	0,00975
	Varianza	0,004	
	Desviación estándar	0,06468	
G3	Media	0,5045	0,00676
	Varianza	0,004	
	Desviación estándar	0,06662	

Al describir la variable densidad léxica según el grupo que produjo el corpus (Tabla 12), se obtiene que el G1 (estudiantes de primer año) arroja un promedio de 0,36 –el más bajo entre los tres grupos–, seguido por G2, con un 0,43 como media de densidad léxica y, finalmente, el grupo conformado por los profesionales, G3, tiene una media de 0,5 en su densidad léxica.

Tabla 13. Estadígrafos descriptivos Diversidad léxica según grupo.

Grupo		Estadístico	Error estándar
G1	Media	0,4713	0,0045
	Varianza	0,002	
	Desviación estándar	,03989	
G2	Media	0,5332	0,01123
	Varianza	0,54	
	Desviación estándar	0,006	
G3	Media	0,5047	0,00609
	Varianza	0,004	
	Desviación estándar	0,05998	

En cuanto a la diversidad léxica total (Tabla 13), que mide la relación de la diversidad con la extensión total de las noticias, los estadígrafos descriptivos señalan que el G1 tiene un promedio de diversidad total del texto de 0,47; en cambio el G2, es en esta oportunidad el grupo que mejor puntaje de diversidad tiene con un promedio grupal de 0,53; luego en G3, los profesionales muestran una media en diversidad léxica de 0,5.

Tabla 14. Estadígrafos descriptivos Diversidad léxica aleatoria según grupo.

Grupo		Estadístico	Error estándar
G1	Media	,5921	,00298
	Varianza	,001	
	Desviación estándar	,03277	
G2	Media	,6320	,00508
	Varianza	,001	
	Desviación estándar	,03373	
G3	Media	,5959	,00408
	Varianza	,002	
	Desviación estándar	,04022	

Los resultados de la diversidad léxica aleatoria en la Tabla 14 muestran leves diferencias con relación a la medida en la tabla anterior (diversidad léxica total, Tabla 13). Las medias muestran que en G1, la diversidad léxica es de 0,59; en el G2, es de 0,63 y la de G3 es de 0,59.

Tabla 15. Estadígrafos descriptivos Diversidad léxica *lead* según grupo.

Grupo		Estadístico	Error estándar
G1	Media	0,6773	0,00545
	Varianza	0,004	
	Desviación estándar	0,05999	
G2	Media	0,7436	0,00591
	Varianza	0,002	
	Desviación estándar	0,03918	
G3	Media	0,6688	0,00593
	Varianza	0,003	
	Desviación estándar	0,05842	

En el caso de la diversidad léxica del *lead* (Tabla 15), los promedios aumentan y superan en los tres grupos la media de 0,65. El G1 presenta en este indicador el promedio más alto obtenido de todas las variables de diversidad, alcanzando una media de 0,67. El G2, por su parte, también presenta el promedio más alto, con una media de 0,74; mientras G3 presenta una media de 0,66. Lo interesante de este estadígrafo descriptivo es que si bien el G3 presentaba los promedios más altos en las mediciones anteriores, en este indicador queda rezagado al relacionarlo con los otros grupos. Aun así, se debe señalar que este grupo presenta promedios similares en todos los indicadores de diversidad, situación que no ocurre con los otros dos grupos que varían considerablemente en unos indicadores de otros.

Esta situación podría ser resultado de la práctica y experiencia en la redacción de noticias escritas. El grupo compuesto por estudiantes de primer año

de la carrera de periodismo está, por primera vez, elaborando textos de hechos noticiosos bajo una estructura determinada. Son capaces de elaborar un *lead* con una diversidad rica, así como textos que, en general, tienen una variedad léxica en la que seis de cada 10 palabras son diferentes, demostrando que son capaces de generar una alta diversidad en sus textos. Sin embargo, en el indicador de densidad léxica, es decir cuando la variedad se mide a través de palabras distintas, pero esta vez con contenido semántico y su relación con la extensión total del texto, sus promedios caen sustancialmente y no superan la diversidad de cuatro palabras diferentes por cada 10 presentes en la noticia. Esta probable explicación se puede contrastar con lo que ocurre en el G2, estudiantes de cuarto año, quienes presentan mejores índices que sus pares de primer año en todos los indicadores.

Por otro lado, hay que tener a consideración algunos aspectos de formalidad en la elaboración de las noticias redactadas por ambos grupos. Mientras que al G1 se le solicitó el texto como parte de una tarea de un curso de producción de textos; el G2 debió elaborar sus noticias para una asignatura cuyo objetivo es publicar, semanalmente, una revista de noticias miscelánea, en un ambiente similar al que se vivencia en un medio de comunicación.

ii.i) PRUEBAS NO PARAMÉTRICAS SEGÚN GRUPO

Mediante la prueba estadística no paramétrica de Kruskal-Wallis se puede comprobar que los grupos que se están estudiando son distintos significativamente. Constituye una alternativa no paramétrica al uso del análisis

de varianza de un factor (Anova1) y, así como en ésta, las muestras pueden ser de tamaños diferentes.

Se pudo comprobar que existen diferencias significativas entre los grupos y la riqueza léxica de sus textos, tanto en la variable diversidad léxica como en la variable densidad léxica, al desechar en ambos casos la hipótesis nula “no existen diferencias significativas en la riqueza léxica en los textos producidos por estudiantes de primer año, de cuarto y profesionales”, hecho que se comprueba al medir los cuatro indicadores de riqueza léxica.

Tabla 16. Resumen de contraste de hipótesis.

	Hipótesis nula	Prueba	Sig.	Decisión
1	Las medianas de Densidad léxica son las mismas entre las categorías de Grupo al que pertenece.	Prueba de la mediana para muestras independientes	,000	Rechace la hipótesis nula.
2	Las medianas de Diversidad léxica total texto son las mismas entre las categorías de Grupo al que pertenece.	Prueba de la mediana para muestras independientes	,000	Rechace la hipótesis nula.
3	Las medianas de Diversidad léxica medido aleatorio son las mismas entre las categorías de Grupo al que pertenece.	Prueba de la mediana para muestras independientes	,000	Rechace la hipótesis nula.
4	Las medianas de Diversidad léxica lead son las mismas entre las categorías de Grupo al que pertenece.	Prueba de la mediana para muestras independientes	,000	Rechace la hipótesis nula.
Se muestran significaciones asintóticas. El nivel de significación es ,05.				

La Tabla 16 contiene el resumen de contrastes de hipótesis de la muestra según el grupo al cual pertenecen. Según lo presentado, existen diferencias

estadísticas significativas no atribuibles al azar y, por ende, se rechazan las hipótesis nulas indicadas en la misma tabla, tras comparar las medianas de cada grupo.

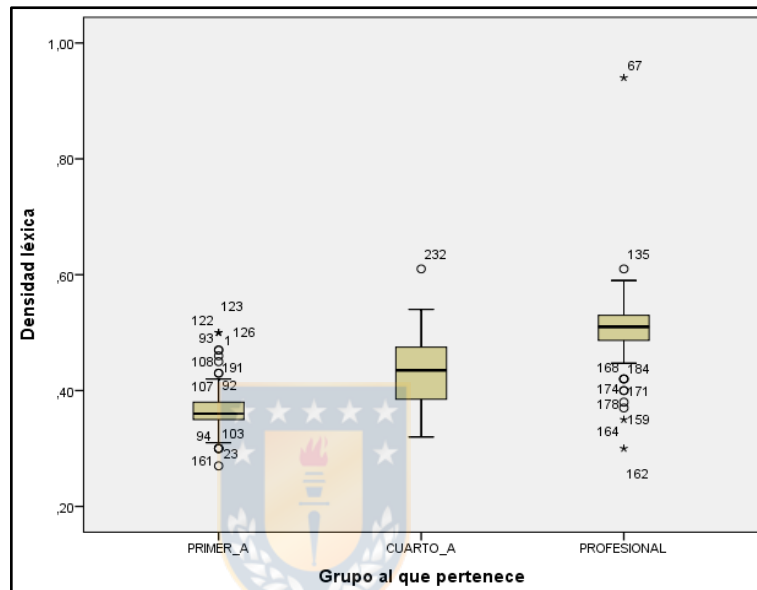


Gráfico 2. Diferencias de densidad léxica según grupo.

El **Gráfico 2** presenta las diferencias que existen entre los tres grupos que participaron con sus noticias escritas. Los datos indican que, efectivamente, existen diferencias significativas entre las muestras y que la variable predictora "Grupo" es un indicador de que los estudiantes de primer año cuentan con una riqueza léxica, en el plano de la densidad, menor a la de los estudiantes de cuarto año y de los profesionales. Por otra parte, evidencia una alta dispersión en el resultado de la densidad dentro de sus propios grupos, principalmente entre los estudiantes de primer año y los profesionales. Aun así, los puntajes límites, es

decir los más altos y más bajos, tienen una mejor puntuación entre los profesionales.

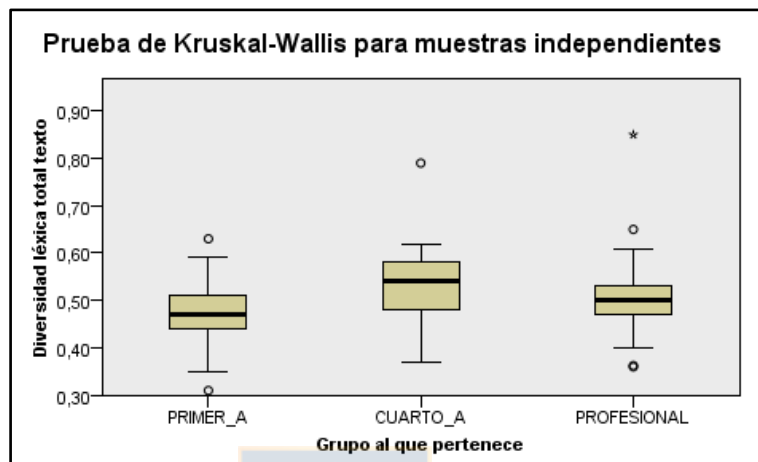


Gráfico 3. Diferencias Diversidad léxica según grupo.

Con el Gráfico 3 se expone, nuevamente, que existe una diferencia estadísticamente significativa entre las medias de cada grupo y que dicha diferencia no corresponde al azar. En esta oportunidad, a diferencia del indicador de densidad léxica, lo que se aprecia es que el grupo conformado por estudiantes de cuarto año presentan un mayor puntaje. Por su parte, los profesionales, si bien no presentan una diversidad mayor que el grupo anterior, sí presentan una menor dispersión en sus puntajes evidenciando a un grupo más cohesionado.

iii) VARIABLE TIPO DE ESTABLECIMIENTO EDUCACIONAL

En Chile, la dependencia administrativa de las instituciones que componen el sistema educacional se divide en tres grandes grupos. El primero de ellos es el sistema público, en el cual las instituciones educacionales (colegios, escuelas, liceos) dependen de fondos del estado para funcionar. Estos fondos son administrados por los municipios, a través de las direcciones de educación municipal, que tienen por objetivo procurar las condiciones óptimas para un eficiente y eficaz desarrollo del proceso educativo en estos establecimientos. El otro gran grupo (el que, debido a las modificaciones de la Ley de Educación en Chile, está próximo a desaparecer) es la educación particular subvencionada. Las instituciones que dependen de este tipo de administración son responsabilidad de un ente sostenedor que recibe aportes o subsidios del Estado, los cuales deben ser cofinanciados por la comunidad que asiste a dichos establecimientos. El tercer grupo está compuesto por las instituciones privadas que se sostienen con el pago particular de la matrícula y aranceles.

Debido a las características descriptivas de esta investigación se propuso abordar el aspecto socioeconómico como una variable que influye en la competencia léxica de los estudiantes. Sin embargo, debido a que el constructo socioeconómico es una variable compleja, que aborda elementos financieros del grupo familiar de cada sujeto, escolaridad de los padres, junto a otro tipo de datos, se decidió constatar solo la dependencia administrativa de la institución de egreso de los participantes de la muestra, entendiendo que esto puede significar algún tipo de sesgo en el resultado del efecto de esta variable.

Tabla 17. Estadígrafos descriptivos Densidad léxica.

Sistema educacional		Estadístico	Error estándar
Privada	Media	0,46	0,01
	Varianza	0,008	
	Desviación estándar	0,09	
Municipal	Media	0,40	0,009
	Varianza	0,006	
	Desviación estándar	0,078	
Subvencionada	Media	0,42	0,006
	Varianza	0,004	
	Desviación estándar	0,06	

La Tabla 17 presenta las medias de la muestra ordenadas según el tipo de establecimiento educacional. Tal como se explicó anteriormente, se encuentra dividida en los tres tipos de dependencia administrativa. La densidad léxica que presenta el grupo de egresados del sistema privado muestra una media de 0,46; la sigue en promedio los egresados del sistema subvencionado con una media de 0,42; y luego los egresados del sistema municipal, con una media de 0,4. En el caso del efecto de la variable dependencia administrativa de los establecimientos educacionales de egreso de enseñanza media, sólo se considerará esta tabla de estadígrafos descriptivos ya que, como se verá más adelante, en el punto Pruebas no paramétricas, no hay una diferencia significativa en las otras variables dependientes para medir la riqueza léxica¹⁶.

¹⁶ Más detalles del resto de los estadígrafos se encuentran en el Anexo 03.

iii.i) PRUEBA NO PARAMÉTRICA PARA TIPO DE ESTABLECIMIENTO EDUCACIONAL

Tabla 18. Contraste de hipótesis Tipo de establecimiento educacional.

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de Diversidad léxica total texto es la misma entre las categorías de Dependencia establecimiento educacional.	Prueba de Kruskal-Wallis para muestras independientes	,041	Rechace la hipótesis nula.
2	La distribución de Densidad léxica es la misma entre las categorías de Dependencia establecimiento educacional.	Prueba de Kruskal-Wallis para muestras independientes	,000	Rechace la hipótesis nula.
3	La distribución de Diversidad léxica medido aleatorio es la misma entre las categorías de Dependencia establecimiento educacional.	Prueba de Kruskal-Wallis para muestras independientes	,078	Conserve la hipótesis nula.
4	La distribución de Diversidad léxica lead es la misma entre las categorías de Dependencia establecimiento educacional.	Prueba de Kruskal-Wallis para muestras independientes	,404	Conserve la hipótesis nula.
Se muestran significaciones asintóticas. El nivel de significación es ,05.				

El resumen de la prueba no paramétrica de Kruskal-Wallis en la Tabla 18 señala que las diferencias estadísticamente significativas se encuentran en los indicadores diversidad léxica total y densidad léxica. Sin embargo, hay que reconocer que el nivel de significación de la variable diversidad léxica total se encuentra muy cercana al límite de significación de 0,5 utilizado en las investigaciones en ciencias sociales y humanidades, por lo tanto, no es de extrañar que en la medida diversidad léxica aleatoria, la prueba no alcance la significancia necesaria para establecer estadísticamente una diferencia significativa.

iv) VARIABLE EXTENSIÓN

Para comprender las comparaciones realizadas en esta variable se debe señalar, en primer lugar, que los análisis se efectuaron con los corpus X1 (con menos de 500 palabras) y X2 (entre 501 y 750 palabras) debido a que en los restantes la cantidad de texto fue insuficiente y, de trabajar con sus promedios, podrían alterar los resultados, induciendo al error en la interpretación.

Como se ha visto desde la revisión teórica de esta investigación, la extensión del texto constituye una variable predictora ante la cual las variables dependientes son sensibles. No es de extrañar que en cada uno de los cluster propuestos se hayan presentado diferencias significativas en los diversos indicadores de riqueza léxica, tal como se muestra en la siguiente tabla:

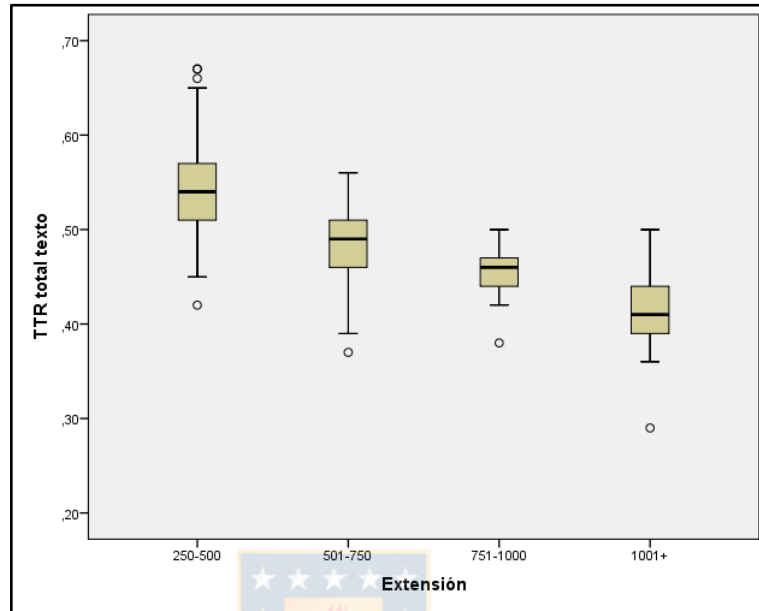


Tabla 19. Estadísticos descriptivos Diversidad léxica según extensión.

Extensión		Estadístico	Error estándar
X1 (250-500)	Media	,5426	,00407
	Varianza	,002	
	Desviación estándar	,04898	
X2 (501-750)	Media	,4846	,00513
	Varianza	,002	
	Desviación estándar	,04075	
X3 (751-1000)	Media	,4582	,00450
	Varianza	,001	
	Desviación estándar	,02622	
X4 (1001+)	Media	,4153	,01186
	Varianza	,003	
	Desviación estándar	,05168	

La Tabla 19 confirma los hallazgos de estudios previos que indican que en textos de menor extensión los índices de diversidad léxica son mayores que en textos más extensos. Mientras que en X1 la media supera el 50% de variedad de palabras, en X4 apenas supera la relación de cuatro palabras diferentes por cada diez presentes en el texto. El descenso de este indicador es consistente en la medida que aumenta la extensión del texto. Este hecho queda mejor explicado en el Gráfico 4.

Gráfico 4. Relación entre extensión y diversidad total.



Un comportamiento similar ocurre con el indicador Densidad léxica, el cual disminuye en textos más extensos, aunque de una manera menos dramática. El promedio de los textos del *cluster X1* superan el 40% de variedad de palabras, mientras que esta relación disminuye a medida que el texto es más extenso. El error en estos estadígrafos es bajo, lo que se ve reflejado, a la vez, en la varianza de cada *cluster*, lo que implica que se trata de una medida muy estable, cuyos valores se encuentran concentrados en torno a su media.

Tabla 20. Estadígrafos descriptivos Densidad léxica según extensión.

Extensión		Estadístico	Error estándar
X1 (250-500)	Media	,4224	,00354
	Varianza	,002	
	Desviación estándar	,04257	
X2 (501-750)	Media	,3937	,00449
	Varianza	,001	
	Desviación estándar	,03566	
X3 (751-1000)	Media	,3785	,00418
	Varianza	,001	
	Desviación estándar	,02439	
X4 (1001+)	Media	,3574	,01028
	Varianza	,002	
	Desviación estándar	,04483	

Al igual que la diversidad, la densidad léxica es también susceptible a la extensión de los textos. La Tabla 20 muestra las medias obtenidas por cada grupo de extensión, indicando una varianza –es decir el grado en el que los resultados se alejan de la media del grupo– que no supera el 0,002 graficando una baja dispersión de los resultados.

Una mirada más completa se presentará a continuación. En el siguiente gráfico de líneas (Gráfico 5) se muestra la relación entre las medidas de riqueza léxica obtenidas del corpus y el tipo de establecimiento educacional del cual egresó el sujeto que escribe la noticia, esta vez tomando como valor la extensión de los textos.

Se puede observar que el promedio en los cuatro indicadores de riqueza léxica –es decir en densidad léxica, diversidad léxica total, diversidad léxica aleatoria promedio y diversidad léxica del *lead*– es superior entre quienes egresaron de un establecimiento educacional particular, decreciendo el promedio en los egresados de educación municipal y volviendo a aumentar entre quienes egresaron de educación particular subvencionada. Se debe indicar que en los datos obtenidos del *cluster* X4, el promedio de cada indicador no refleja la tendencia del resto de los *clusters*, ya que los datos de X4 corresponden solo al 7,3% del total del corpus.

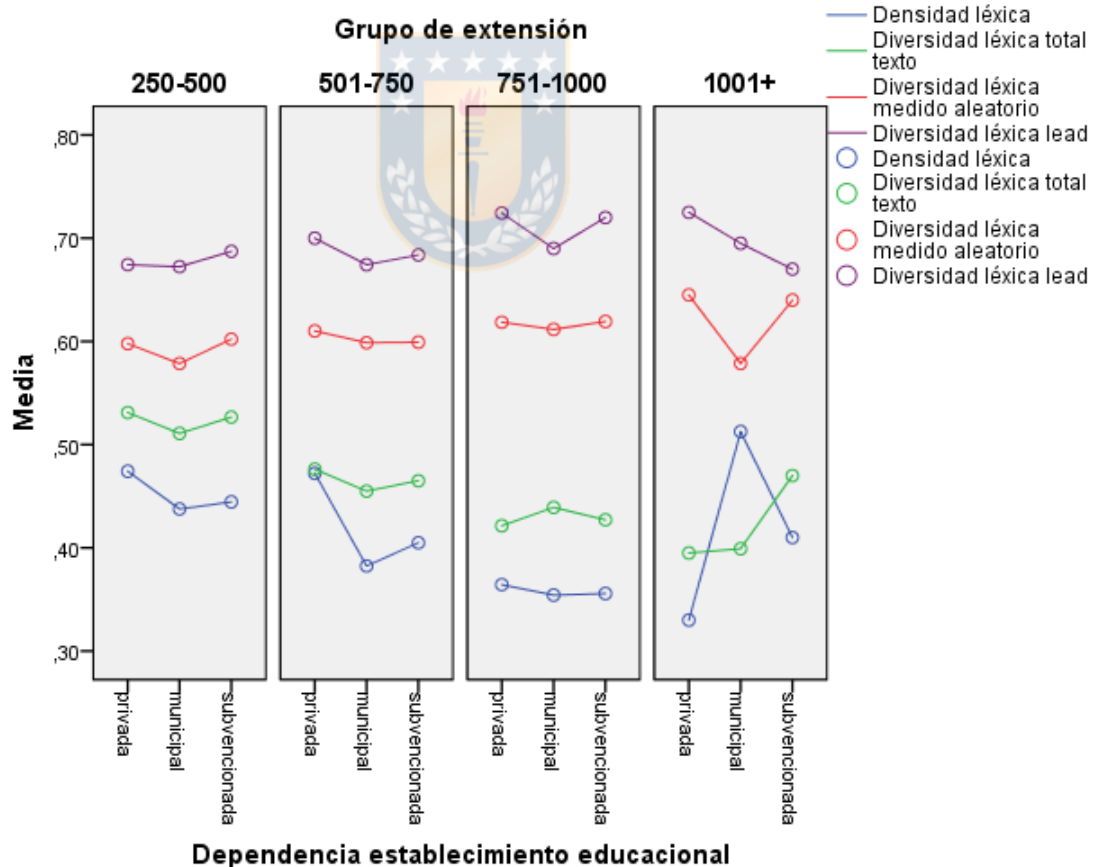


Gráfico 5. Indicadores de riqueza léxica relacionados con extensión del texto y tipo de establecimiento.

4.1.2. DISPONIBILIDAD LÉXICA

A continuación, se presentan los datos obtenidos de las encuestas de disponibilidad léxica. Como se ha indicado en el marco metodológico, la muestra que realizó esta prueba fue la misma que participó con sus noticias para la medición de riqueza léxica, por lo que no se volverá a abordar dicho aspecto.

4.1.3.1. RESULTADOS GENERALES

Para comenzar el análisis, se describirán los datos globales sobre disponibilidad léxica. En primer lugar, se debe recordar que fueron cinco los centros de interés estudiados, tres de ellos (ciencia y tecnología, política y deporte) corresponden a las temáticas solicitadas como textos de noticias escritas para la medición de riqueza léxica; los dos siguientes son partes del cuerpo y medios de transporte, que corresponden a centros de interés estudiados previamente en diversas investigaciones (Baralo, 2007; Echeverría, Herrera, Moreno y Pradenas, 1987; Hernández, 2004).

En general se obtuvieron 7989 palabras, de ellas 2339 corresponden a vocablos, es decir, palabras que no se repitieron, o como se indica en el estudio de riqueza léxica, palabras únicas.

i) *TOTAL DE PALABRAS*

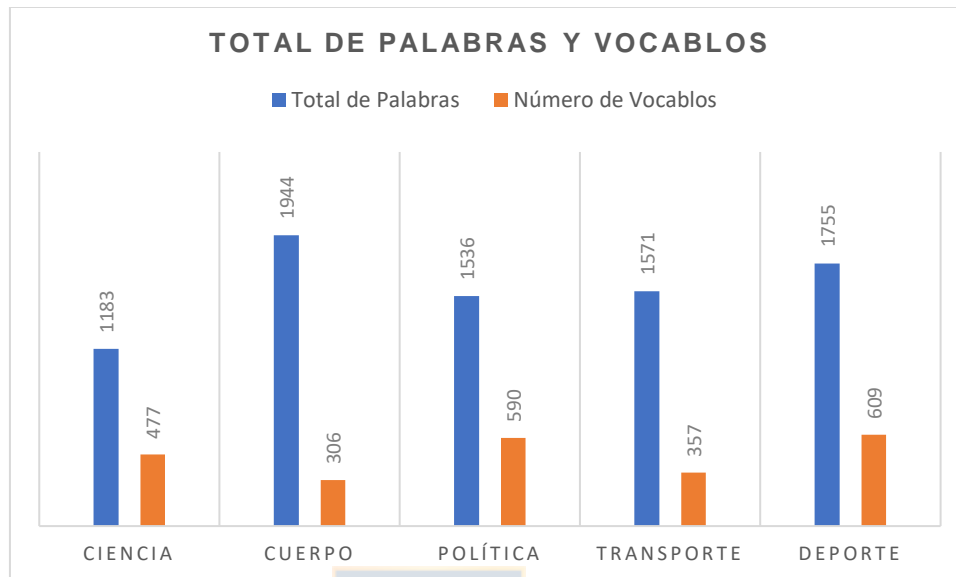


Gráfico 6. Total de palabras y vocablos obtenidos en cada centro de interés.

Como se muestra en el Gráfico 6, en el centro de interés ciencia y tecnología, la muestra reportó un total de 1183 palabras, de ellas 477 corresponden a vocablos, es decir a palabras que no fueron repetidas dentro del corpus de palabras de este centro. En el centro de interés partes del cuerpo, se presentó el mayor número de respuestas por parte de la muestra, con 1944 palabras; sin embargo, fue también el centro de interés con menos vocablos evocados, cuyo número fue de 306 conceptos diferentes. El centro de interés política quedó compuesto por 1536 palabras y 590 vocablos, medios de transporte por 1571 palabras y 357 vocablos y, finalmente, el centro de interés deporte obtuvo un total de 1755 palabras y 609 vocablos, el más alto de los cinco campos estudiados.

ii) *PROMEDIO DE RESPUESTAS*

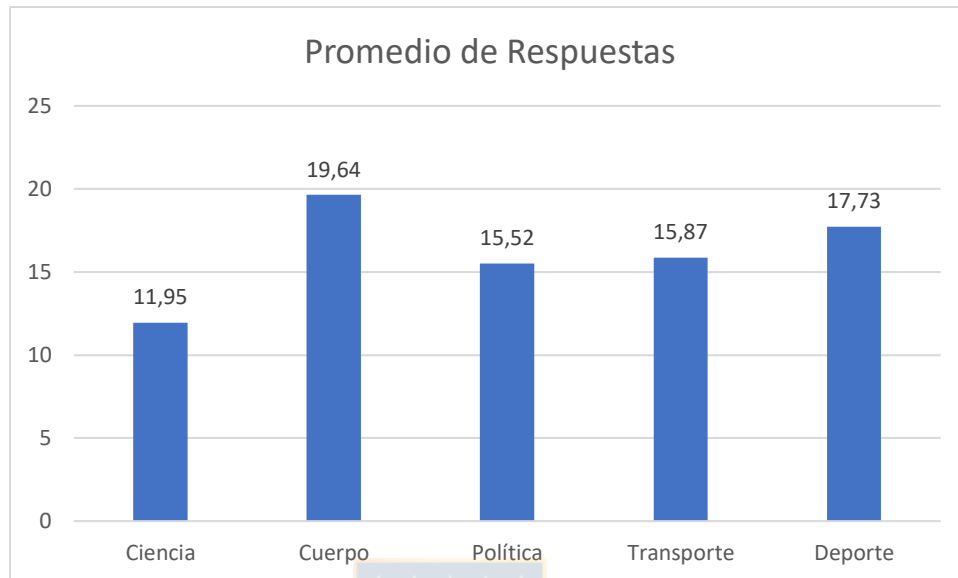


Gráfico 7. Gráfico según promedio de respuestas

El Gráfico 7 expone el promedio de respuestas obtenido en cada centro de interés. Al igual que en el total de palabras, *partes del cuerpo* fue el eje que obtuvo el valor más alto entre los cinco centros, con un promedio de 19,6 palabras por sujeto; seguido por *deporte*, con 17,7 palabras en promedio; a continuación, *medios de transporte*, con 15,8 palabras en promedio por sujeto; luego, en cuarto lugar, *política* con 15,5 palabras y, por último, el centro de *interés ciencia y tecnología*, con 11,9 palabras en promedio por sujeto.

Siguiendo a Echeverría et al (1987), los centros de interés *cuerpo humano* y *medios de transporte* pertenecen a un grupo denominado "de alto grado de

coherencia semántica"¹⁷. En ese sentido sería posible explicar por qué el centro *partes del cuerpo* obtuvo los puntajes más altos en los tres indicadores: promedio de respuesta, número de vocablos y número de palabras.

iii) *ÍNDICE DE COHESIÓN*

Tabla 21. Orden según Índice de cohesión.

Centro de Interés	Índice de Cohesión
Cuerpo	0,064
Transporte	0,044
Deporte	0,029
Política	0,026
Ciencia	0,025

Al ordenar los centros de interés según su Índice de cohesión, la revisión propuesta por Echeverría et al (1987), sobre los grupos de alto grado de coherencia mencionados con anterioridad, se replican con los resultados obtenidos en esta investigación. Los índices más altos corresponden a los centros *partes del cuerpo* y *medios de transporte*, como se puede apreciar en la Tabla 21.

¹⁷ En su estudio Echeverría identifica siete centros de interés de este tipo, junto con los señalados están los centros *Ropa, muebles, herramientas y materiales de construcción, alimentos y animales*.

iv) *FRECUENCIA DE PALABRAS*

La Tabla 22 presenta las cinco palabras con mayor frecuencia¹⁸ en cada centro de interés. En ésta se puede apreciar, también, cómo varía el índice de disponibilidad de cada palabra y la categoría gramatical (o clase) a la que corresponde, las cuales en su mayoría son nombres (sustantivos).

Tabla 22. Listado de las cinco palabras más frecuentes.

Centro de Interés 01 (Ciencia y tecnología)				Centro de Interés 02 (Partes del cuerpo)			
Vocablo	IDL	Frec.	% Aparición	Vocablo	IDL	Frec.	% Aparición
computador	0,29	34	34,34	pierna	0,54	69	69,70
internet	0,26	33	33,33	ojo	0,54	71	71,72
avance	0,25	33	33,33	brazo	0,54	66	66,67
investigación	0,20	27	27,27	cabeza	0,52	58	58,59
biología	0,19	24	24,24	mano	0,49	64	64,65
Centro de Interés 03 (Política)				Centro de Interés 04 (Medios de transporte)			
Vocablo	IDL	Frec.	% Aparición	Vocablo	IDL	Frec.	% Aparición
corrupción	0,27	35	35,35	automóvil	0,75	84	84,85
senador	0,25	38	38,38	bus	0,62	74	74,75
diputado	0,24	35	35,35	bicicleta	0,61	82	82,83
gobierno	0,23	32	32,32	avión	0,60	81	81,82
presidente	0,22	31	31,31	tren	0,45	61	61,62
Centro de Interés 05 (Deporte)							
Vocablo	IDL	Frec.	% Aparición				
fútbol	0,72	78	78,79				
tenis	0,52	64	64,65				
básquetbol	0,49	59	59,60				
natación	0,31	45	45,45				
atletismo	0,25	35	35,35				

¹⁸ Ver Anexo 04 con listado completo de palabras, su frecuencias de aparición y el Índice de disponibilidad léxica

4.1.3.2. ANÁLISIS SEGÚN VARIABLES ESTUDIADAS

A continuación, se expondrán los resultados descriptivos de las variables predictoras estudiadas según centro de interés. Este análisis resulta significativo debido a que dentro de los objetivos de esta investigación se encuentra el establecer comparaciones entre los grupos que participaron, así como el tipo de establecimiento educacional del cual egresó cada sujeto, y sexo.

i) VARIABLE SEXO

En la muestra participaron 98 personas¹⁹, 39 de ellas son de sexo masculino y 59 de sexo femenino. Al analizar la variable según el número de palabras recogidas (Gráfico 8), los resultados indican que las mujeres produjeron mayor número de palabras que los hombres, un total de 4.959 palabras; obteniendo un 69,9% del total; mientras que los hombres, con 3.039 palabras conforman el 38,1% del total de palabras. La diferencia, con relación a los datos generales, es que en esta perspectiva el centro de interés *deporte* concentra, entre ambos, el mayor número de palabras (404 palabras las mujeres y 367 palabras los hombres), seguidos por los centros *política* y *ciencia y tecnología*. Los dos últimos corresponden a los centros de interés *medios de transporte* y *partes del cuerpo*. La incongruencia entre los resultados se debe a que al separar por grupos existen palabras que se repiten entre los centros de ambos grupos y otras palabras, que evoca un grupo y otro no. Esta diferencia se verá reflejada a

¹⁹ Ver Tabla 1, con el resumen según grupo.

lo largo de todo el análisis según las variables independientes debido a la explicación recién entregada.

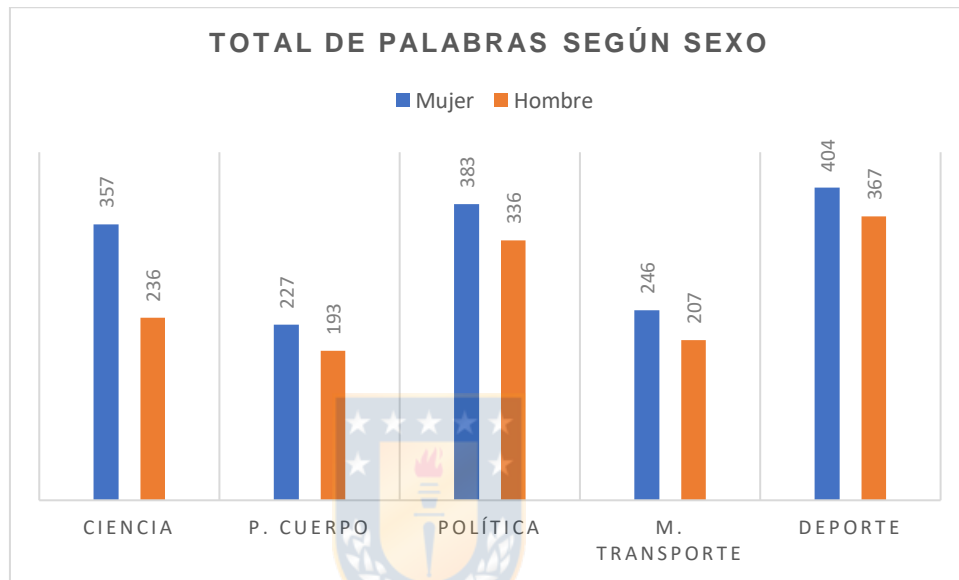


Gráfico 8. Total de palabras según sexo.

Si bien el resultado indica que las mujeres tienen una mayor producción léxica, hay que observar que la muestra en su mayoría estuvo conformada por sujetos de sexo femenino, por lo tanto una afirmación como aquella puede constituir una equivocación estadística.

Otro indicador significativo y que entrega un panorama más preciso es el Total de vocablos, entendidos estos últimos como las palabras sin repetir dentro del corpus evocado.

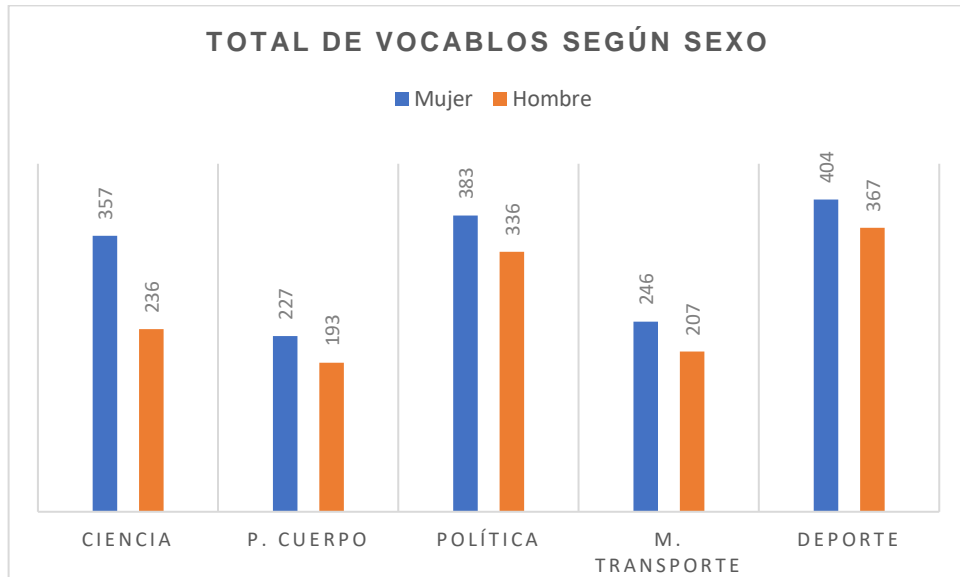


Gráfico 9. Total de vocablos según sexo

En el Gráfico 9 los indicadores muestran que existe un resultado que, si bien favorece nuevamente a las mujeres, con un 54,7% del total de vocablos con 1.617²⁰ evocados; no es significativamente mayor que el de los hombres, quienes obtuvieron un 45,3% del total de vocablos, evocando este grupo en total 1.339.

Tal como se ha señalado en investigaciones previas, si bien existe una diferencia entre ambos grupos, ésta no es significativa. Lo anterior queda explícito con el estadístico U de Man Whitney, prueba no paramétrica que permite

²⁰ Cabe aclarar que la disminución de vocablos, con relación a lo indicado en la primera parte del análisis de Disponibilidad léxica, responde a que en dicho punto se suma el total de vocablos producido por ambos grupos, sin embargo, al separar por la variable sexo existen vocablos que fueron producidos por un grupo y por otro no, y otros vocablos que se repiten entre ambos grupos. El análisis tomó en cuenta la suma de la producción de vocablos de ambos grupos, para no distorsionar la visión de los datos.

comparar grupos cuya principal característica es que la distribución de sus medias no es normal o no representan una curva de normalidad.

Los resultados señalan que se debe aceptar la hipótesis nula para cada comparación de centros de interés, es decir que la distribución es la misma para ambos grupos, menos para la hipótesis con el centro de interés partes del cuerpo, donde sí las diferencias son significativas: 0,045. Sin embargo, esa significación se aproxima mucho a la línea de error aceptada en las ciencias sociales (0,05), por lo que, para efectos de esta investigación, y al tratarse de una prueba no paramétrica, y de muestras que cuentan con una variación numérica en el sexo de sus participantes, será considerada como no significativa, aceptando la hipótesis nula presente en la Tabla 23.



Tabla 23. Contraste de hipótesis.

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de Total personal de vocablos ciencia es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,689	Conserve la hipótesis nula.
2	La distribución de Total personal de vocablos partes del cuerpo es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,045	Rechace la hipótesis nula.
3	La distribución de Total personal de vocablos política es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,361	Conserve la hipótesis nula.
4	La distribución de Total personal de vocablos medios de transporte es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,738	Conserve la hipótesis nula.
5	La distribución de Total personal de vocablos deporte es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,221	Conserve la hipótesis nula.
6	La distribución de Total personal de vocablos según sujeto es la misma entre las categorías de SEXO.	Prueba U de Mann-Whitney para muestras independientes	,181	Conserve la hipótesis nula.
Se muestran significaciones asintóticas. El nivel de significación es ,05.				

El motivo por el cual no se considerará al centro de interés *partes del cuerpo* como una variable dependiente con diferencias significativas entre hombres y mujeres, queda mejor explicado en el Gráfico 10 el cual muestra la alta probabilidad de que la significancia de 0.45 se deba a que el número de mujeres es mayor que el de los hombres.

Prueba U de Mann-Whitney para muestras independientes

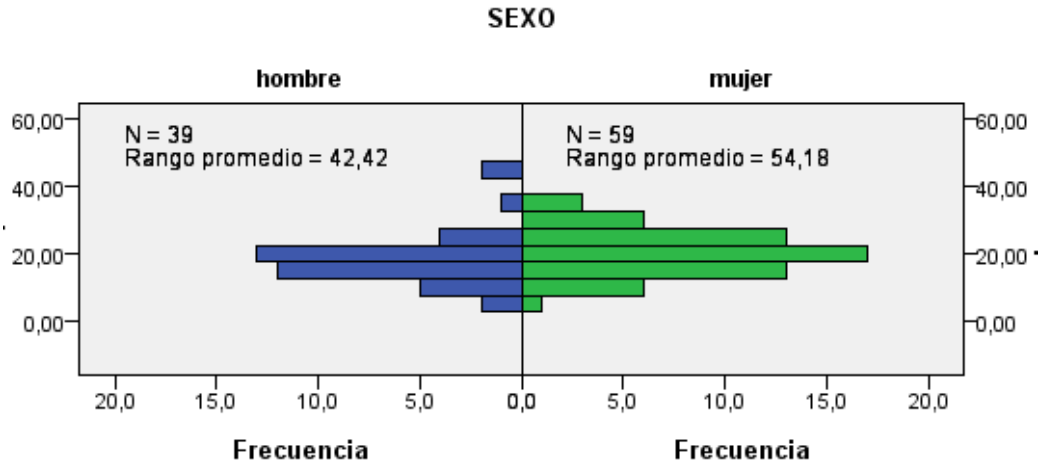


Gráfico 10. Prueba U de Mann Whitman para muestras independientes.

Finalmente, una mirada que complementa lo expuesto por las pruebas estadísticas, son las que se obtienen a partir del software Dispogen II (Gráfico 11), el cual realiza una comparación de grupos, y los muestra en perspectiva según su número de respuestas.

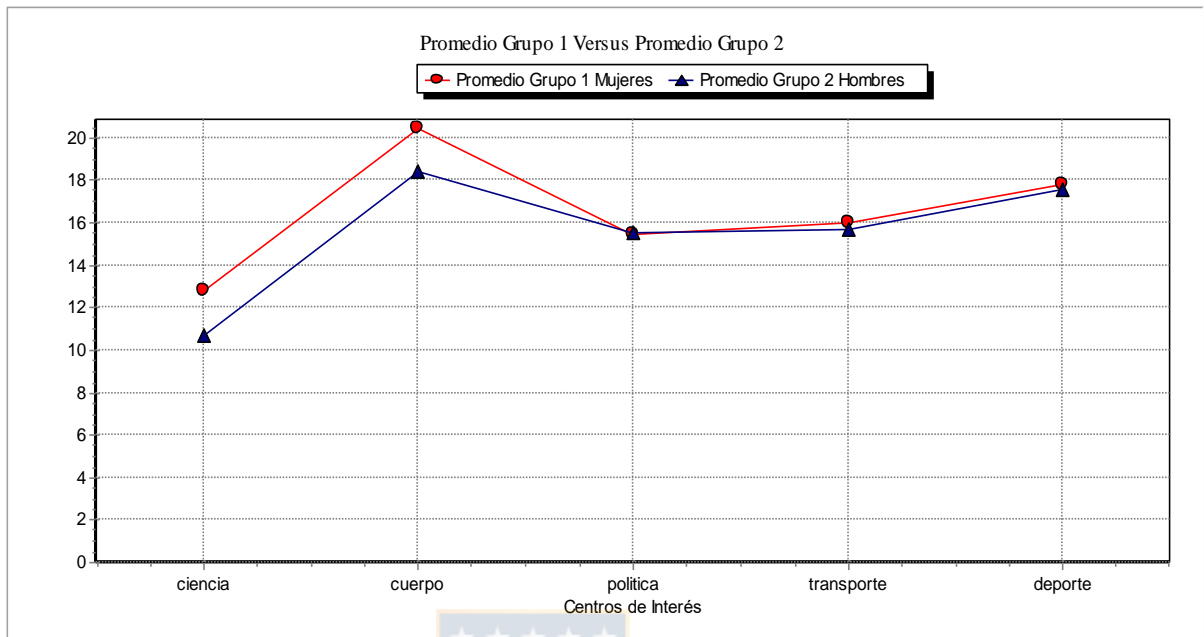


Gráfico 11. Comparación de Promedios de grupo según centro de interés.

El último parámetro revisado, en esta parte del análisis, es el **Índice de Cohesión** (Tabla 24). Se ha hablado en puntos anteriores del alto índice de cohesión semántica de ciertos centros de interés. Ya en la exposición de los resultados globales se comprobó que, pese a ser un estudio completamente distinto y con diferentes propósitos, los centros de interés *partes del cuerpo* y *medios de transporte* son temáticas con un alto grado de coherencia semántica. Estos resultados serán nuevamente corroborados cuando la variable discriminante sea sexo.

Tabla 24. Índice de Cohesión.

	Mujer	Hombre
Ciencia y tecnología	0,0358	0,0453
P. cuerpo	0,0901	0,0953
Política	0,0404	0,0463
M. transporte	0,0649	0,0759
Deporte	0,0441	0,0479

Los resultados muestran también una tendencia a la correlación de puntajes promedio entre hombres y mujeres, es decir, varían de la misma forma y de manera positiva, aunque poseen datos cuantitativos distintos, con $p=0.9$, lo que se considera una correlación casi perfecta.

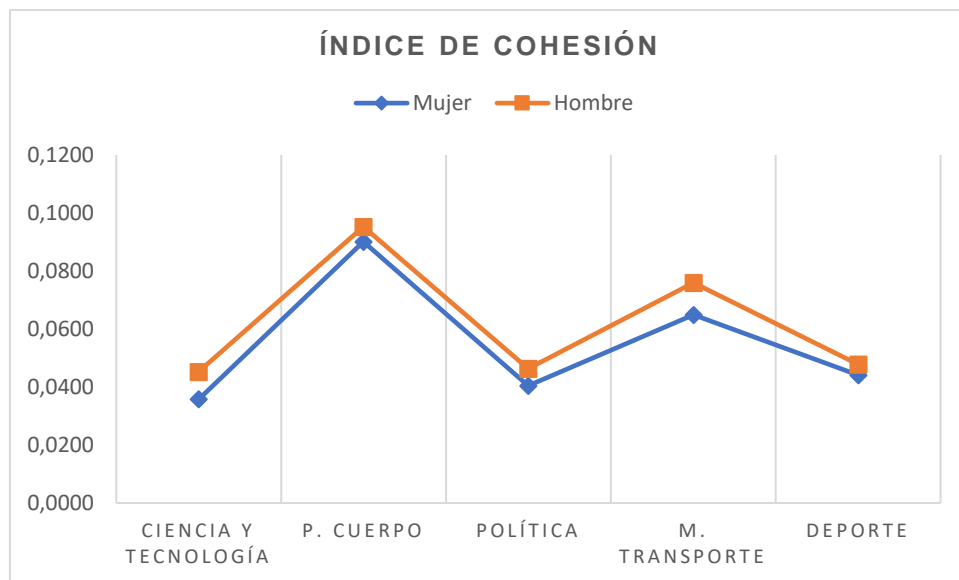


Gráfico 12. Índice de cohesión.

ii) *VARIABLE GRUPO*

El efecto de la disponibilidad léxica según la segunda variable independiente en estudio corresponde a los *grupos* con los cuales estuvo compuesta la muestra, a saber, estudiantes de primer año de periodismo, estudiantes de cuarto año de periodismo y profesionales periodistas. Al igual que en la variable sexo, no debe extrañar que los datos no sean congruentes con el total de palabras expuestas en los resultados globales, esto debido a que, al dividir los grupos, la cantidad de vocablos cambia según lo que expuso cada grupo como conjunto.

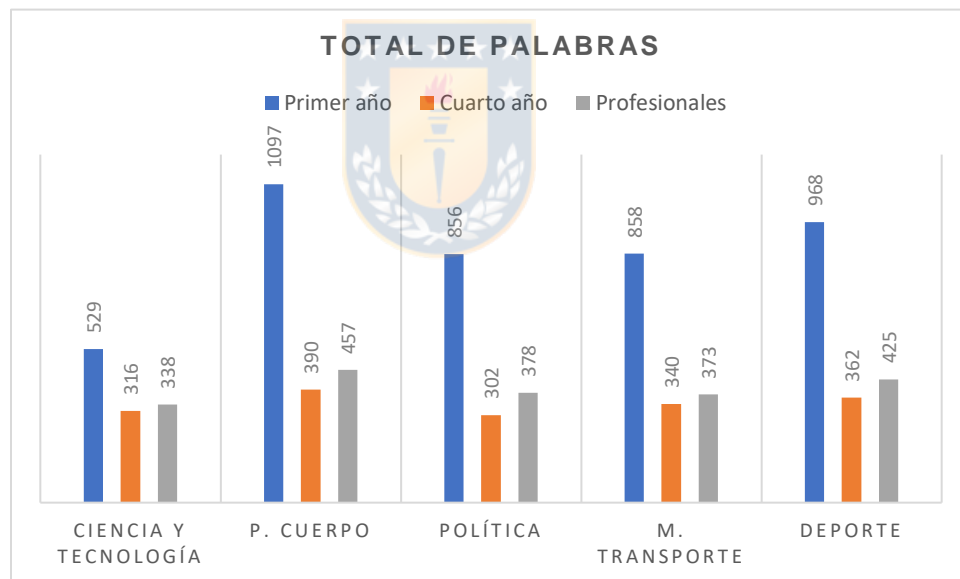


Gráfico 13. Total de palabras según grupo.

El Gráfico 13 indica el total de palabras que cada grupo evocó en cada uno de los centros de interés. Los estudiantes de primer año claramente presentaron el mayor número de palabras en todos los centros de interés, con 4038 en total; luego los profesionales, con 1971 palabras; y, finalmente, los estudiantes de

cuarto año, quienes escribieron 1710 palabras en total. Estos resultados no deben extrañar si se toma en consideración que la muestra estuvo compuesta por 58 estudiantes de primer año, 20 de cuarto año y 20 profesionales. En ese sentido, es interesante destacar que la diferencia entre los estudiantes de cuarto año y los profesionales es de aproximadamente 260 palabras. Aquí nuevamente el centro de interés con el mayor número de respuestas es partes del cuerpo, tendencia que ha ocurrido en los puntos de Resultados globales y en Variable Sexo.

Para analizar estadísticamente los resultados entre el total de palabras y el grupo que los produjo se efectuó la prueba no paramétrica Kruskal Wallis, que examina y compara la distribución de k muestras distintas. A simple vista, se podría señalar que, entre los grupos, principalmente el G1 (primer año) y los otros dos, la diferencias en la producción de palabras es significativa, sin embargo, al aplicar la prueba comparando los cinco centros de interés, se demostró que no lo son.

Gráfico Kruskal Wallis en cada Centro de interés

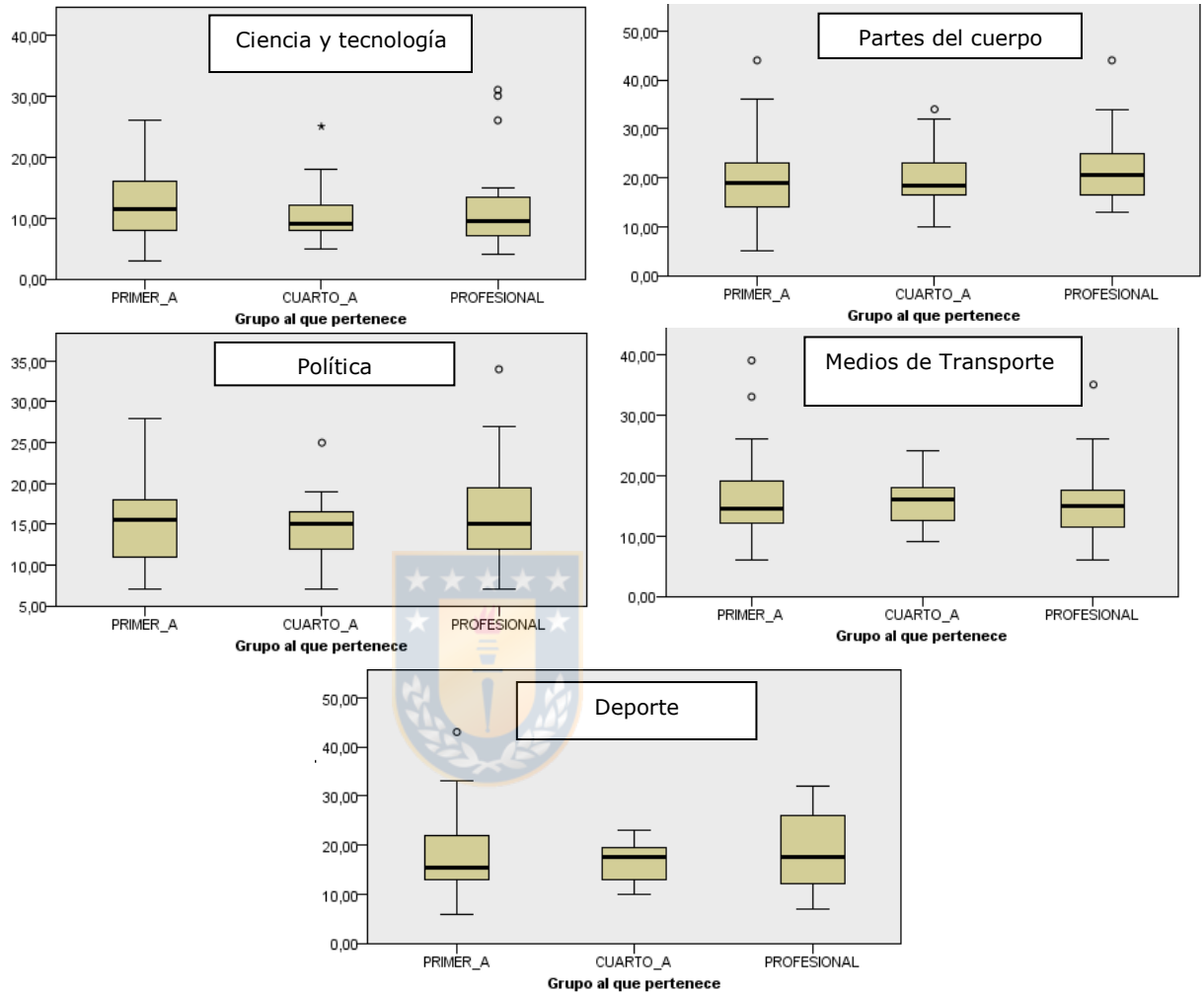


Figura 13. Estadístico Kruskal Wallis en distribución de vocablos según grupo.

Lo que nos indican los gráficos en la Figura 13 es que la distribución de las respuestas entre los grupos no presenta variaciones significativas estadísticamente, esto pese a la magnitud de la diferencia entre lo que reportó cada grupo.

Al analizar desde la perspectiva del **promedio de palabras** emitidas por los sujetos, el análisis cambia. Si bien, son los estudiantes de primer año quienes entregaron el mayor número de palabras, sus promedios no son los más altos e, incluso, son menores que los de sus pares de cuarto año. El estudio de los promedios siempre debe abordarse con precaución, debido a que son muy susceptibles al total de los casos. Sin embargo, para este estudio, que no sólo entrega la mirada a un aspecto de la disponibilidad léxica, sino que pone el énfasis en diversos indicadores para intentar contemplar una fotografía más general de lo que ocurre en los diferentes grupos de la muestra, el promedio es un aspecto que viene a complementar dicho panorama y no se puede dejar de enunciar sus resultados.

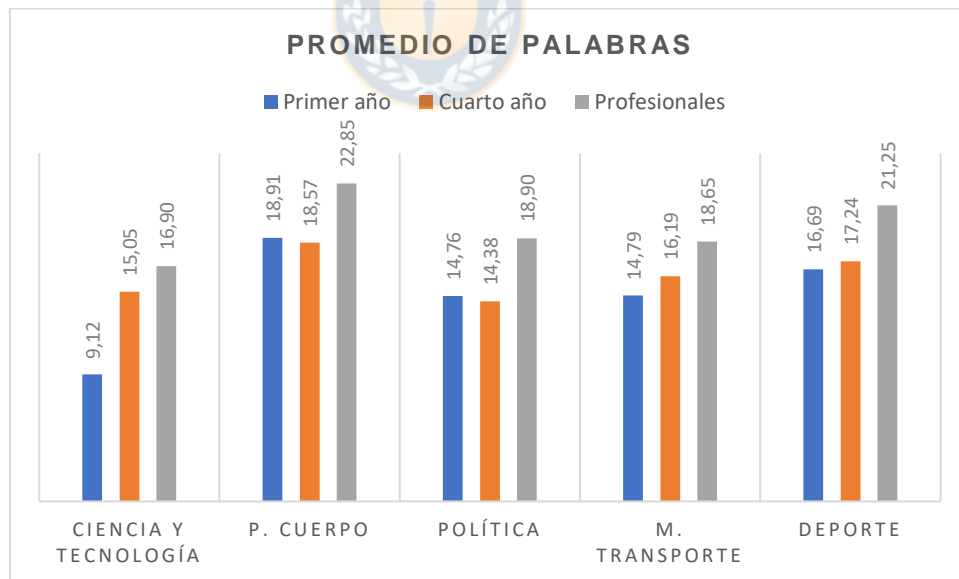


Gráfico 14. Promedio de palabras según grupo.

Atendiendo a las limitantes propias de enfocar los estudios en los promedios, el Gráfico 14 cambia por completo la visión del Gráfico 13 (que presenta el total de palabras); es más, bajo esta concepción de resultados, son los profesionales quienes reportan los indicadores más altos, seguidos por los estudiantes de cuarto año y, por último, los estudiantes de primer año. Sin embargo, y como ya se ha establecido con la prueba estadística presente en la Figura 13, estas diferencias no tienen un grado de significación.

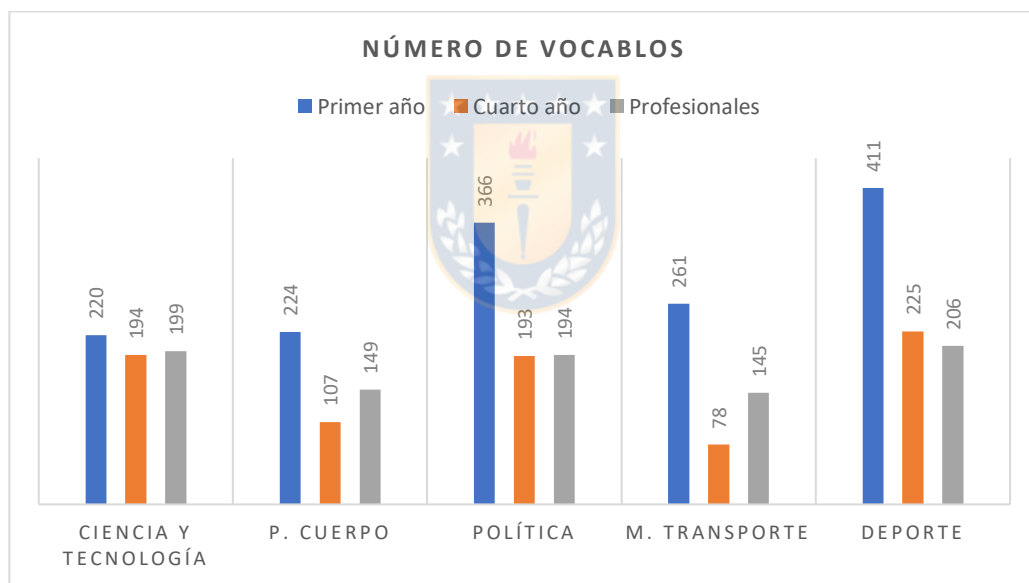


Gráfico 15. Número de vocablos según grupo.

El Gráfico 15 expone la cantidad de **vocablos** (número de palabras distintas) emitidos por cada grupo. Bajo esta mirada son los estudiantes de primer año quienes, nuevamente, expresaron el mayor número reportando en total 1482 palabras distintas en los cinco centros de interés, con el mayor número en *deporte*

y la cantidad más baja en *ciencia y tecnología*. Luego, el grupo de profesionales, con 892 vocablos en total, presentó el mayor número de palabras distintas en el centro de interés *deporte*²¹ y el menor número de vocablos lo reportó en el centro de interés *medios de transporte*. Los estudiantes de cuarto año emitieron un total de 797 vocablos en los cinco centros de interés, su mejor puntaje lo obtuvieron en el campo *deporte* –al igual que los otros dos grupos–, y su centro de interés con menos vocablos fue, al igual que los profesionales, *medios de transporte*.

Finalmente, el Índice de Cohesión, que implica el grado de coincidencia en las respuestas, permite complementar la mirada obtenida con el indicador anterior, el número de palabras distintas (o vocablos, como se les denominó). Este índice permite acceder al grado de homogeneidad en el uso del léxico.

En la variable *sexo* este índice mostró que entre hombres y mujeres se presenta una correlación casi perfecta, siendo el más alto índice el del centro de interés *partes del cuerpo*.

²¹ Cabe mencionar que en este centro G3 reportó menos vocablos que G1 y G2.

Tabla 25. Índice de cohesión en los centros de interés según grupo.

	Primer año (G1)	Cuarto año (G2)	Profesionales (G3)
Ciencia y tecnología	0,0415	0,0776	0,0849
P. cuerpo	0,0844	0,1736	0,1534
Política	0,0403	0,0745	0,0974
M. transporte	0,0567	0,2076	0,1286
Deporte	0,0406	0,0766	0,1032

En la Tabla 25 se presentan el índice de cohesión de cada centro de interés, ordenado según la variable predictora grupo. En un sencillo ordenamiento de mayor a menor, se puede señalar que el centro de interés medios de transporte, en el G2, es el que logra mayor cohesión, con 0,2 puntos; luego, bajo el mismo grupo, el centro de interés partes del cuerpo (0,17); posteriormente, se ubican los centros de interés de los profesionales en partes del cuerpo, medios de transporte y deporte, ambos superando el índice de 0,1 puntos.

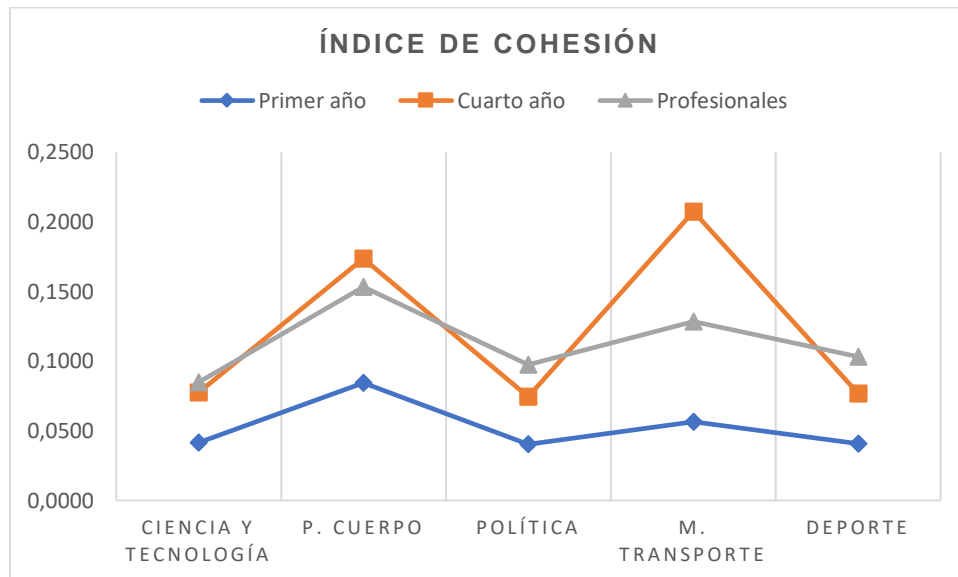


Gráfico 16. Índice de cohesión en centros de interés según grupo

El Gráfico 16 muestra la relación en la que varían los índices de cohesión de cada centro de interés, ésta es una visión que complementa lo expuesto en la Tabla 25. El gráfico nuevamente aventura una posible correlación en los puntajes de los tres grupos, la cual se confirma al realizar una Correlación de Pearson entre los grupos. Entre G1 y G2 existe un $p=0,7$; mientras que la correlación entre G1 y G3 es de $p=0,9$; finalmente, entre G2 y G3, la correlación alcanza un $p=0,8$.

iii) VARIABLE TIPO DE ESTABLECIMIENTO EDUCACIONAL

Un análisis muy propio de los estudios de disponibilidad léxica es el que se efectúa con un enfoque socioeconómico o sociocultural. Si bien ambos enfoques se construyen con diversos indicadores, que pueden incluir ingreso per cápita del grupo familiar, tipo de vivienda, camas en la vivienda y otros, para los fines de

esta investigación, se definió utilizar como indicador sólo la dependencia administrativa del establecimiento educacional del cual egresó cada sujeto de la muestra. Si bien esto pudiera parecer entregar una mirada sesgada del efecto de esta variable, se decidió así debido a los alcances descriptivos de esta investigación.

Tabla 26. Distribución de los sujetos según tipo de establecimiento educacional de egreso.

	Dependencia establecimiento educacional			Total
	privada	municipal	subvencionada	
Total	28	26	44	98

Al revisar la distribución de los sujetos según el establecimiento educacional del cual egresaron (Tabla 26), nuevamente puede observarse que la muestra está desbalanceada: un 45% de la muestra corresponde a egresados de un establecimiento educacional subvencionado, mientras que un 28,5% egresó de una institución educacional privada, y el 26,5% de una municipal. Tomando en consideración dicha situación se procedió a revisar los indicadores.

El total de palabras evocadas por los integrantes de cada tipo de establecimiento educacional fue de 2291 palabras para los provenientes de colegios privados, 2103 palabras en los egresados de establecimientos subvencionados y 3595 de establecimientos municipalizados o públicos.

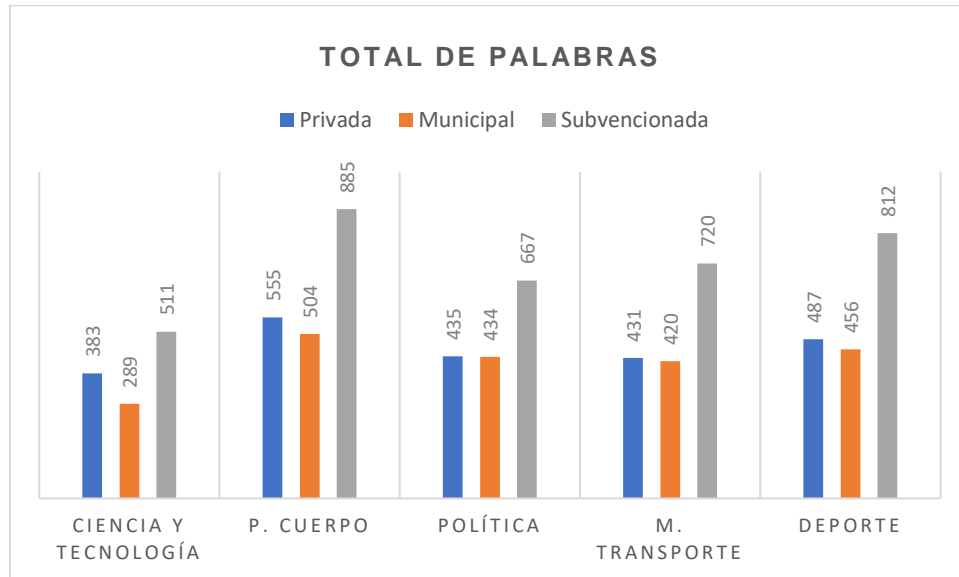


Gráfico 17. Total de palabras en cada centro de interés según tipo de establecimiento educacional.

Considerando la disparidad que presenta la muestra, no es de extrañar que, en este indicador, los sujetos procedentes de instituciones municipalizadas sean quienes evocan el mayor número de palabras y que los provenientes de los otros dos revelen un número similar entre sí. Por otro lado, el centro de interés partes del cuerpo es nuevamente el que concentra el mayor número de palabras, siguiendo la tendencia mostrada anteriormente; lo sigue el centro de interés deporte; por último, el centro con menor cantidad de palabras es el de ciencia y tecnología. En el Gráfico 17 se presentan los totales absolutos en cada categoría.

Bajo una mirada estadística²², los resultados no presentan diferencias significativas en la distribución de sus medias. Los resultados obtenidos se contradicen con la bibliografía estudiada, la cual indica que el aspecto sociocultural o socioeconómico es una variable influyente en la disponibilidad léxica, por este motivo es que tomamos estos resultados con cuidado de no adelantar, ni aseverar que esta variable no tiene preponderancia en la disponibilidad léxica. La comparación de varianzas se presenta en la Figura 14, que agrupa los gráficos comparativos en cada indicador del total de palabras según sujetos, agrupados según el tipo de establecimiento educacional del cual egresaron.



²² Prueba no paramétrica Kruskal Wallis para muestras independientes, la cual fue corroborada con una Prueba T de muestras independientes, las cuales no arrojaron diferencias significativas.

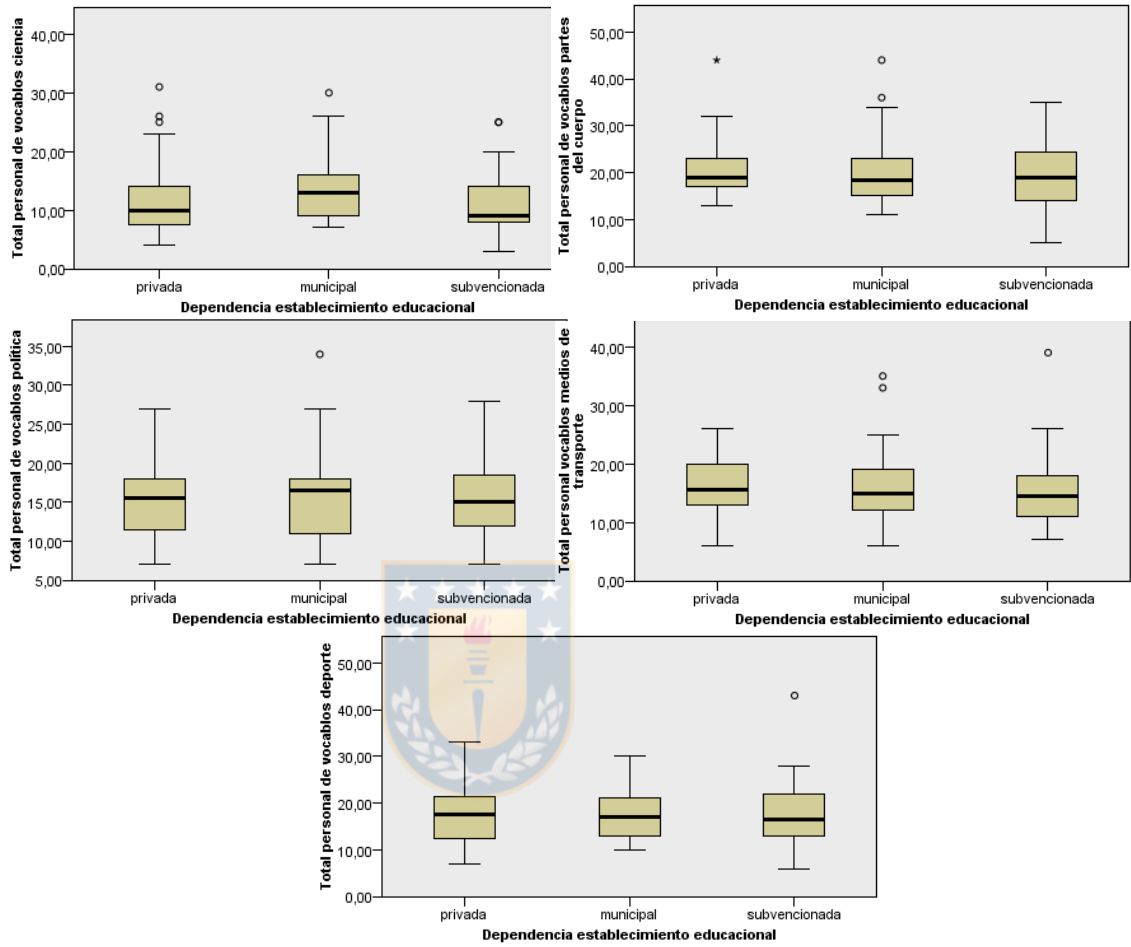


Figura 14. Prueba de varianzas entre centros de interés según dependencia educativa.

El indicador sobre promedio de palabras, en el **Gráfico 18**, al igual que en las anteriores variables, presenta un acercamiento entre los resultados de cada grupo estudiado, acomodando los valores a porcentajes.

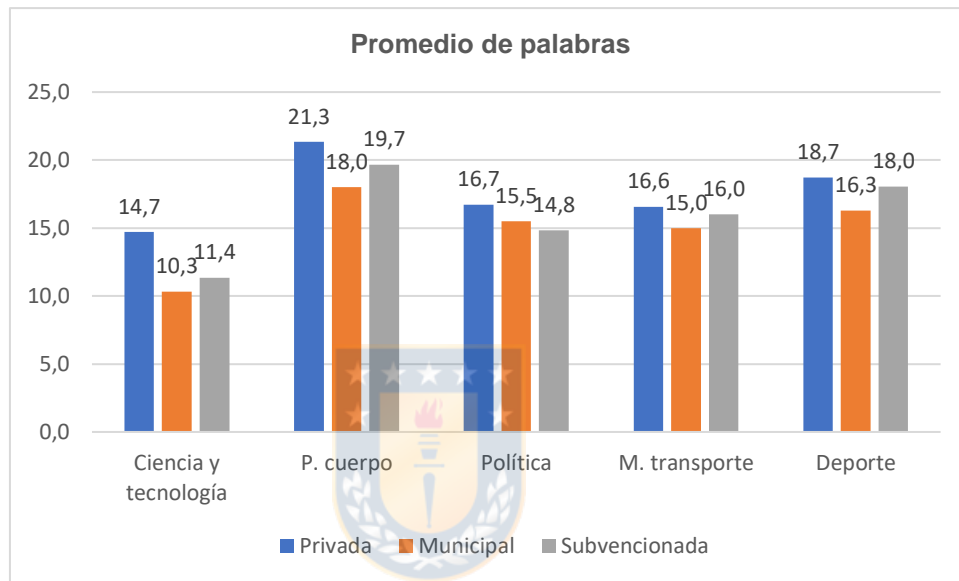


Gráfico 18. Promedio de palabras según dependencia educacional.

Queda explicitado que, a nivel de promedio, las diferencias no son de gran magnitud: estos resultados se corroboran con lo señalado con la prueba estadística, la cual no considera que la distribución de palabras tenga diferencias significativas al tener la variable *tipo de establecimiento educacional* como predictora, aunque el grupo compuesto por sujetos egresados de establecimientos particulares presentan mejores promedios que establecimientos municipales y los subvencionados.

El Índice de cohesión global lo abordaremos desde los datos obtenidos de manera automática por el *software* Dispogen II. El índice de cohesión que obtuvo cada muestra indica que los mejores puntajes son obtenidos por los egresados de establecimientos educacionales privados, donde el mayor índice fue en el centro de interés *partes del cuerpo*, seguido por *medios de transporte*. El centro de interés *ciencia y tecnología*, indica un índice de cohesión muy similar entre los establecimientos privados y los municipales, pero los establecimientos subvencionados son los que obtuvieron el índice más bajo. La situación cambia en el centro de interés *política*, donde obtuvieron mayor índice los colegios municipalizados, en segundo los privados y en tercer lugar los subvencionados. Situación que queda mejor presentada en el gráfico de líneas (Gráfico 19), el cual muestra la variación entre los centros de interés de cada muestra.

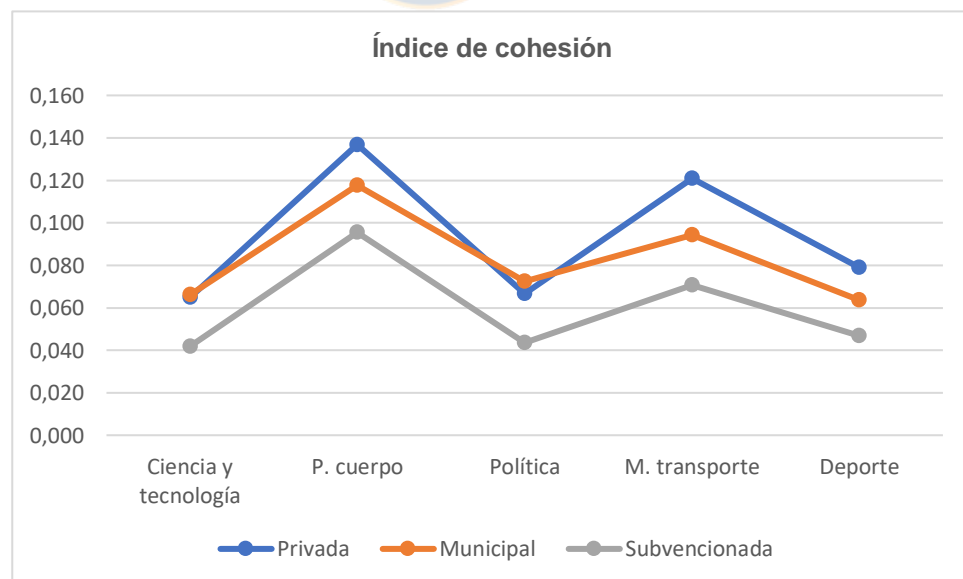


Gráfico 19. Índice de cohesión según establecimiento educacional.

Debido a la preponderancia de la variable Índice de Cohesión en los estudios de disponibilidad léxica, y en particular en esta investigación –que busca, como se ha indicado, analizar y describir la competencia léxica de estudiantes de periodismo de una determinada casa de estudios superiores–, se ha decidido realizar un análisis parcial, extrayendo sólo los resultados de los estudiantes, sin que la muestra de profesionales afecte a estos últimos. Debido a que sólo participaron 20 sujetos de cuarto año (G2) se ha decidido considerar únicamente la variable tipo de establecimiento educacional para efectuar las comparaciones. Por otra parte, se trabajará con el Índice de Disponibilidad Léxica Individual, con el que es posible medir la participación de cada individuo en el resultado global de la muestra. Esto quiere decir que, si un sujeto evoca vocablos de mayor índice de disponibilidad léxica, su disponibilidad léxica –valga la redundancia– será más alta que la de otro. Esta idea fue recogida por Echeverría (2002), del trabajo propuesto por López Chávez y Strassburger (1991).

En este caso, la prueba de Kruskal Wallis, indica diferencias significativas para el promedio de disponibilidad léxica entre estas tres muestras, con una significancia de $\alpha=0,14$. La muestra tuvo un $n=78$; en la cual, 18 estudiantes provenían de establecimientos privados; 22 de municipalizados y 38 de educación subvencionada.

Prueba de Kruskal-Wallis para muestras independientes

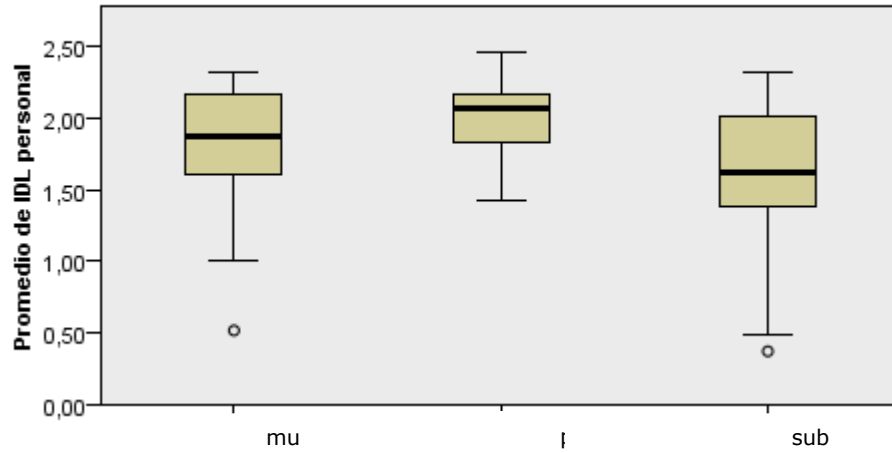


Gráfico 20. Diferencias en el promedio de disponibilidad léxica.



4.1.3. REDES DE ASOCIACIÓN LÉXICA

4.1.3.1. CIENCIA Y TECNOLOGÍA

Al analizar el centro de interés *ciencia y tecnología*, con los tres grupos que compusieron la muestra, tanto el software Gephi como Dispografo reportaron un total de 484 nodos y 956 aristas, demostrando, con este último el total de relaciones que existen entre los vocablos dentro de la red.

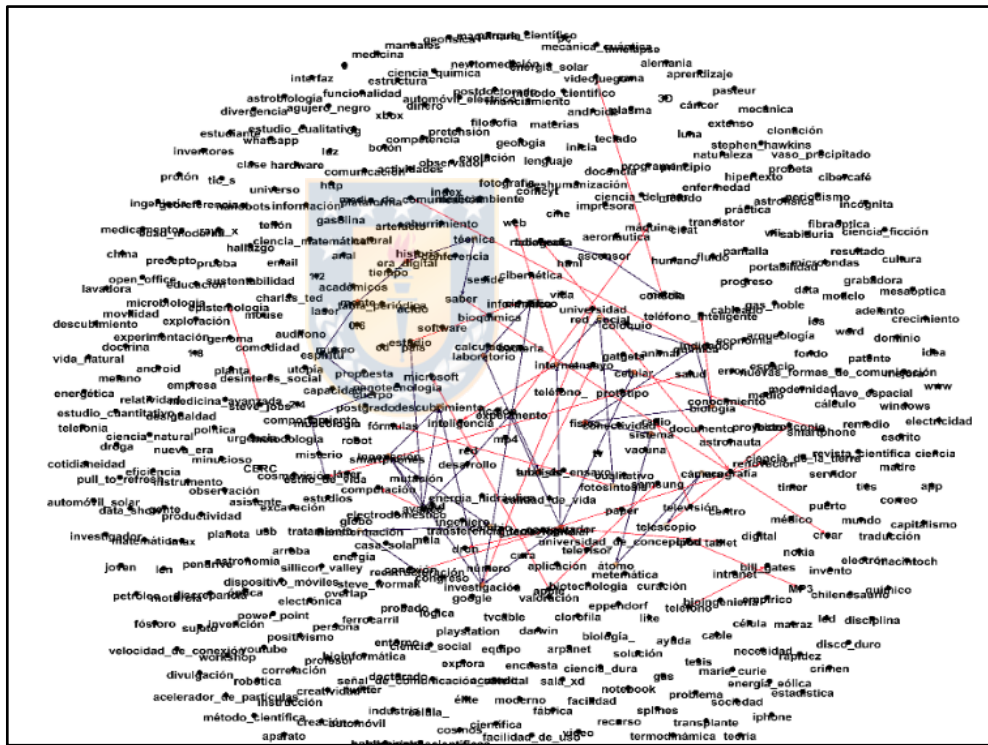


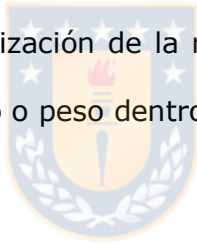
Figura 15. Red de ciencia y tecnología sin filtro.

La Figura 15 expone todos los nodos que fueron evocados por los sujetos. En este caso la imagen describe lo que en el análisis de la disponibilidad léxica muestra el índice de cohesión del centro de interés ciencia y tecnología, un campo al que pueden asociar un alto número de conceptos. Es lo que se denomina un

centro de interés con una baja relación semántica, pero que, sin embargo, en este caso puede evocar un número significativo de conceptos.

Al efectuar una poda de los nodos, dejando sólo aquellos con más de una arista, la imagen de asociatividad que entrega Gephi visualiza los vocablos que se relacionan con otro nodo. Dentro de las utilidades de este programa, se encuentra la posibilidad de aplicar diferentes algoritmos de distribución que facilitan el ordenamiento de los nodos y sus relaciones, levantando una nueva mirada a la relación entre palabras. Uno de los más utilizados es el algoritmo ForceAtlas II, que sirve especialmente para grafos de pequeña y mediana escala.

Para una mejor visualización de la red, se efectuó una poda de aquellos nodos con un rango de grado o peso dentro de la red menor a 7.



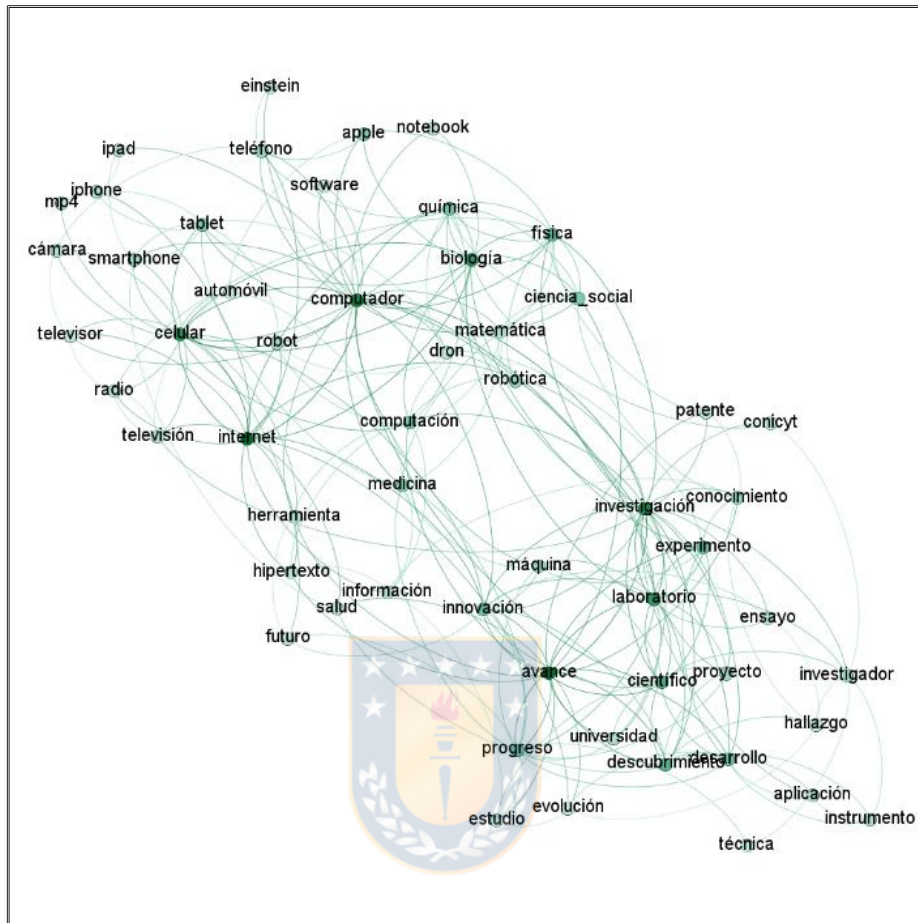


Figura 16. Red de asociación ciencia y tecnología con rango de grado 7.

La Figura 16 muestra cómo, luego de aplicar la herramienta que filtra según el *rango de grado* de un nodo, la red se redujo a 56 nodos (es decir, el 11% del total que compone la red) y 228 aristas (un 23% del total). Con esta poda se aprecia, además, que se distribuyen los nodos según su peso (o frecuencia) dentro de la red y cómo se asocian con otros nodos. Este porcentaje reducido del total es otro indicador de que se trata de una red de baja cohesión semántica.

4.1.3.1.1. CIENCIA Y TECNOLOGÍA SEGÚN GRUPO

Uno de los objetivos de este trabajo es describir cómo los estudiantes organizan sus redes de asociación semántica, para ello, se utilizará el programa Dispogen, que permite la selección de variables y se filtrará de la red completa (Figura 15) a los profesionales y a los estudiantes de cuarto año, para realizar un análisis posterior en cada uno.

i) ESTUDIANTES DE PRIMER AÑO (G1)

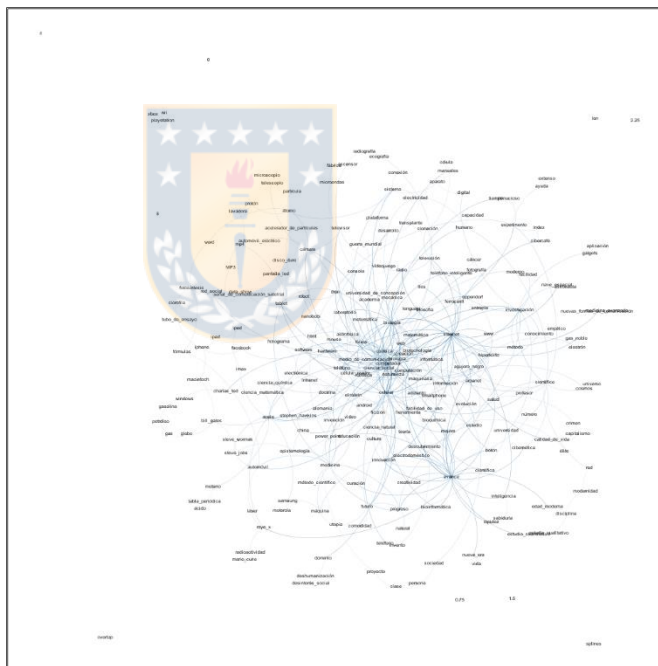


Figura 17. Red Ciencia y tecnología de primer año.

El gráfico dirigido de la red semántica en ciencia y tecnología de los estudiantes de primer año reportó 229 nodos y 420 aristas. Al igual que en la red

general, se efectuó un ordenamiento según el algoritmo *ForceAtlasII*²³. Al utilizar esta herramienta se logra obtener una red con *clusters* más tensos, por lo tanto, mejor identificables.

Para proveer una visualización más significativa de la red, se ordenarán los tamaños de los nodos según su medida de *centralidad interina*²⁴, la cual indica para cada nodo qué tan a menudo aparece en el camino más corto entre dos nodos aleatorios de la red. Mientras más grande es el nodo, mayor es su grado de centralidad y mayor influencia tiene debido a su función como conector dentro de la red.

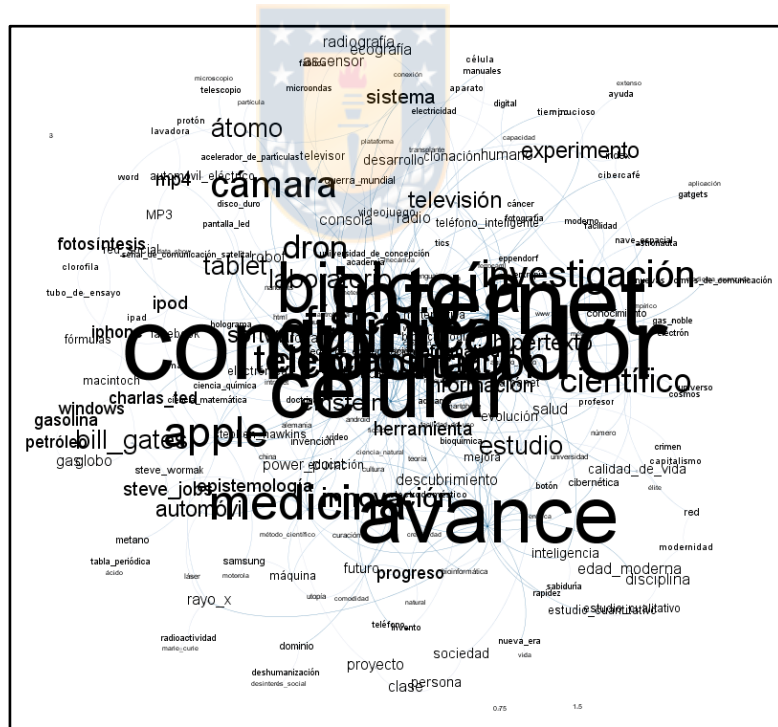


Figura 18. Grado de centralidad en ciencia y tecnología.

²³ Los parámetros utilizados para obtener la imagen de la Figura 14, corresponden a Escalado: 100; Gravedad: 20; evitando el solapamiento de nodos.

²⁴ En inglés *Betweenness centrality*.

La Figura 18 representa la centralidad interina de algunas palabras: computador, avance, celular, biología, medicina e internet; son algunos de los nodos con mayor grado de centralidad dentro de la red y representan, por lo tanto, los vocablos más utilizados y significativos dentro de la red semántica de los estudiantes de primer año en el centro de interés *ciencia y tecnología*.

Gephi permite, además, utilizar estadísticos especializados en el estudio de grafos. Así se utilizó el algoritmo *Modularidad* que permite detectar comunidades dentro de una red, mediante un cálculo automático que pondera un algoritmo de distribución aleatoria y el peso de las aristas. En el caso de los estudiantes de primer año, la Modularidad identificó 19 comunidades.

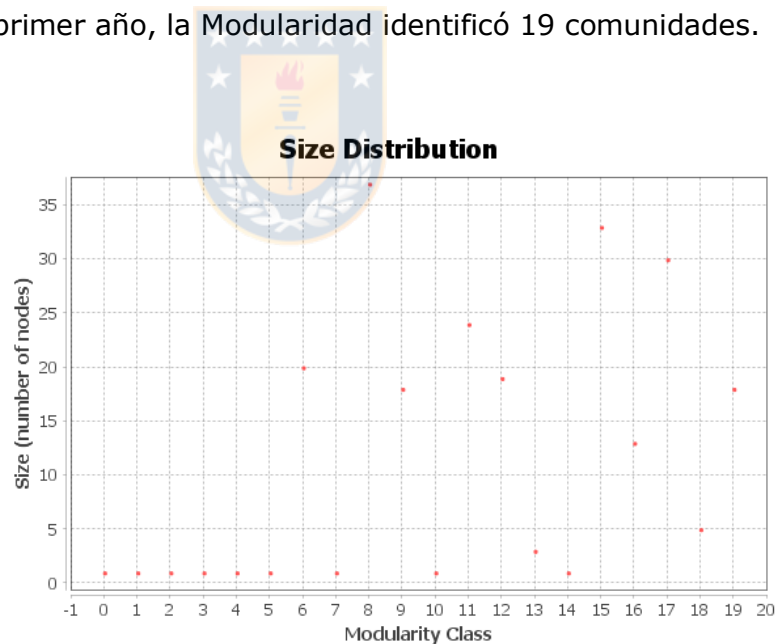


Gráfico 21. Distribución de las comunidades según modularidad.

El primer dato que llama la atención del Gráfico 21 es el alto número de comunidades de tamaño pequeño, es decir que se representan por los nodos de menos peso. Debido a la gran cantidad de nodos presentes se efectuaron dos filtros a la red. El primero fue de carácter topológico, en el cual se seleccionó el filtro rango de grado, que permite la visualización de nodos más significativos de la red; en este caso se eligió el grado 7, que corresponde a 28 nodos y 87 aristas. El segundo filtro fue grado de centralidad, el cual ya había sido calculado por el estadístico de Gephi para la red completa; en este caso se seleccionaron los nodos cuyos grados de centralidad fueran los más cercanos al de mayor puntaje y que, además aún se configuraran en una red de asociación, es decir, que no quedarán nodos inconexos.

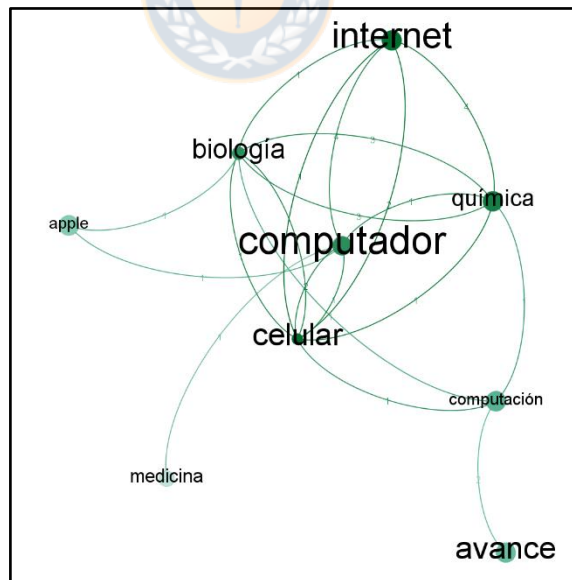


Figura 19. Resultado de la red con nodos y aristas significativas.

La Figura 19 representa los nodos con mayor grado de centralidad dentro de la red, así como las aristas más significativas. Además, muestra los vocablos que unen, centralizadamente, los significados dentro de esta red. La Modularidad, en este caso, encontró tres comunidades de nodos.

1. Comunidad "0": biología – química - apple – física
2. Comunidad "1": avance – computación
3. Comunidad "2": computador - internet – celular – medicina

ii) ESTUDIANTES DE CUARTO AÑO (G2)

La red de los estudiantes de cuarto año está compuesta por 201 nodos y 265 aristas, las cuales fueron reportadas mediante la prueba de disponibilidad léxica aplicada a 20 estudiantes. Al igual que en los estudiantes de primer año, se aplicó el algoritmo de distribución *ForceAtlasII*²⁵. Así también se ordenó según el *Grado de centralidad* de los nodos.

²⁵ Utilizando los mismos parámetros del grupo de estudiantes de primer año, G1.

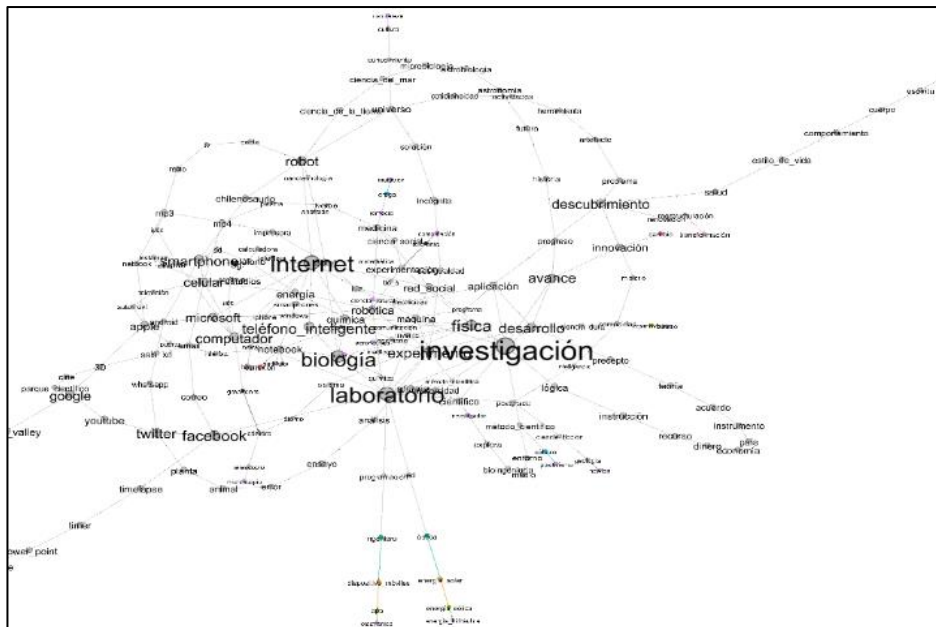


Figura 20. Red Ciencia y tecnología 4 año.

La Figura 20 es la representación de la red de los estudiantes de cuarto año (G2). Al igual que en el G1, en el centro se encuentran los vocablos con mayor grado de centralidad y, por lo tanto, mejor conectados con el resto de los nodos, a diferencia de los vocablos que se encuentran en la periferia, los cuales tienen una menor capacidad de unión dentro del campo semántico.

Por otra parte, el coeficiente de *Modularidad* de la red identificó 22 comunidades. El Gráfico 22 expone la distribución según talla de las diferentes comunidades de palabras. Si bien la distribución tiene similitud a la presentada en G1, difieren en la talla. Mientras que la primera se caracterizaba por una alta distribución con una baja talla, la de G2 presenta comunidades de mayor tamaño.

En este sentido, se puede deducir que la red semántica de los estudiantes de cuarto año logra presentar vocablos con mayor grado de centralidad y conexión. Este grupo cuenta, además, con la capacidad de relacionar mayor cantidad de vocablos que sus pares de primer año (G1).

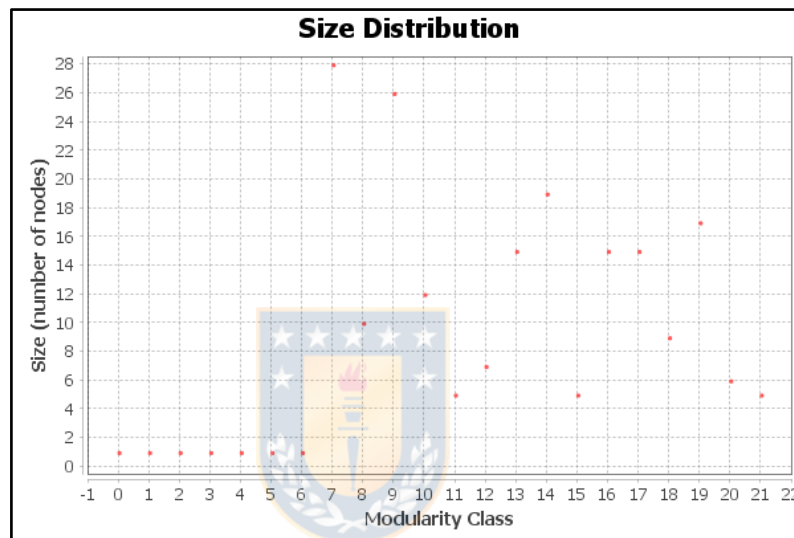


Gráfico 22. Modularidad de ciencia y tecnología G2.

Finalmente, con los filtros topológicos *rango de grado* y *grado de centralidad*, se obtendrá el núcleo de vocablos con los que el G2 representa la red de *ciencia y tecnología*. La Figura 21 representa la red de G2 con los nodos de mayor grado de centralidad y con un rango de grado²⁶ de =5 (el rango mayor fue =12). En esta red los nodos claves, según su centralidad fueron *internet*,

²⁶ El rango de grado implica la cantidad de aristas que salen y que recibe un determinado nodo.

laboratorio e investigación con un grado =12, los dos primeros y de =11 el tercero.

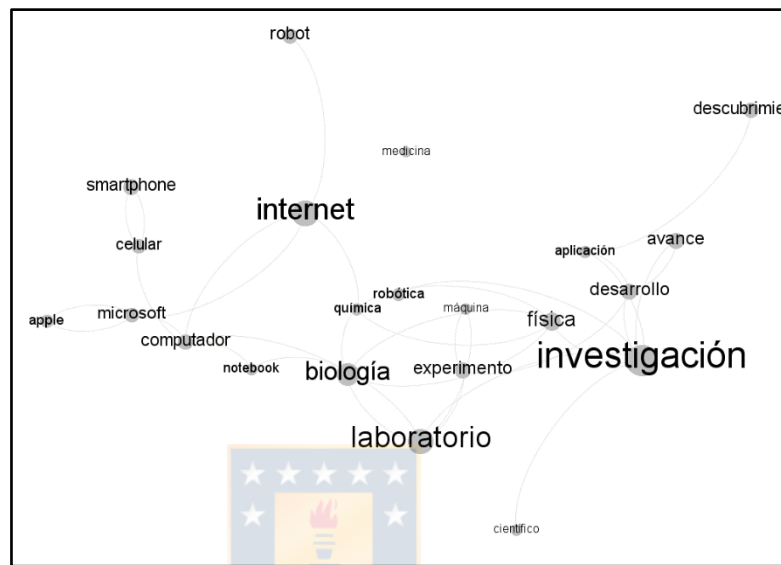


Figura 21. Red filtrada de G2.

La *Modularidad*, bajo estos parámetros, arrojó seis comunidades, aunque de ellas tres solo contaban con un nodo. La explicación para esto es que al subir el rango de grado quedaron fuera de la red vocablos que podían ser parte de una comunidad, pero que no compartían el mismo peso de significancia dentro de la red. Las otras tres comunidades están compuestas por un número de nodos, donde el grado menor fue =5.

1. Comunidad "7": laboratorio – biología – física – desarrollo – avance – científico – máquina -investigación.
2. Comunidad "8": internet – smartphone – computador – celular – microsoft – apple – química – notebook

3. Comunidad "16": robótica - medicina

iii) PROFESIONALES (G3)

La red de los profesionales está conformada por 203 nodos y 302 aristas y fue elaborada a partir del léxico disponible de 20 periodistas profesionales.

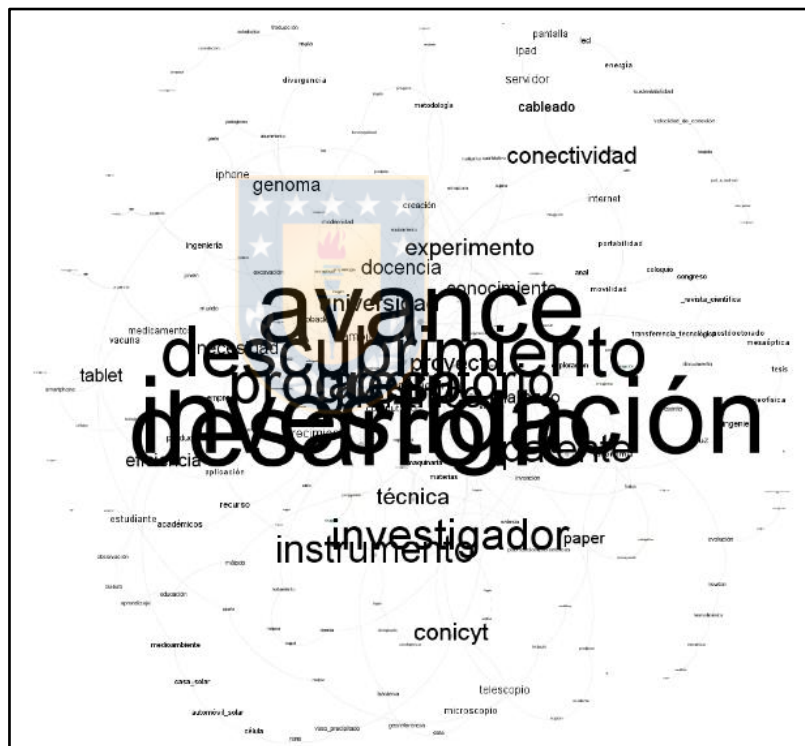


Figura 22. Red de Ciencia y tecnología de profesionales.

Al igual que lo realizado en las redes de G1 y G2 para distribuir la red de los profesionales, se utilizó la distribución *ForceAtlasII*, con los parámetros Escalado 100 y Gravedad 20. Además, la red se ordenó según la apariencia de *Betweenness centrality*, por lo ya explicado en los grupos anteriores.

Por otra parte, el estadístico *Moduaridad* identificó 20 comunidades de palabras. Siete de esas comunidades corresponden a una talla pequeña (entre uno y dos nodos); dos comunidades de tres y seis nodos, y el resto de las comunidades con más de 10 nodos, tal como lo indica el Gráfico 23.

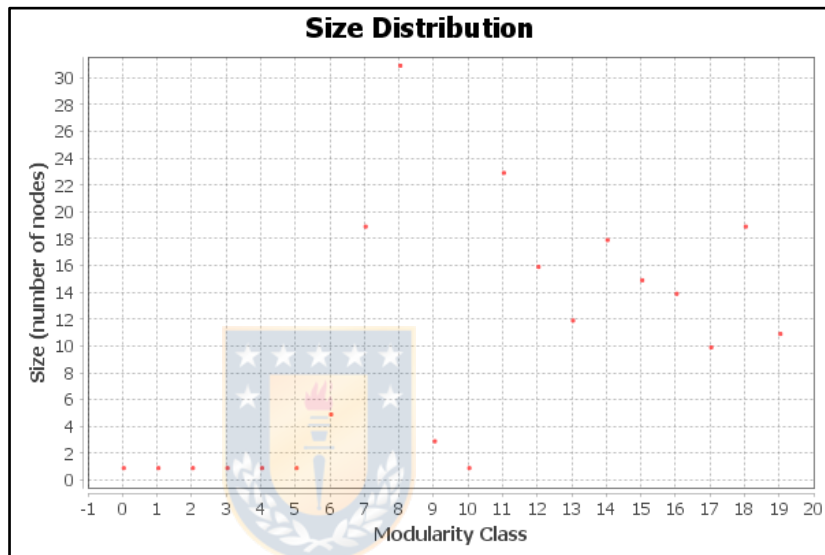


Gráfico 23. Comunidades en ciencia y tecnología G3.

Con el **Gráfico 23** encontramos la primera distinción sustantiva entre los tres grupos. Mientras G1 se distingue por un alto número de comunidades (19 en total), estas presentaban una talla o *size* -número de nodos- baja, 11 de ellos tenían una talla entre 1 y 5 nodos, el resto se ubicó sobre una talla de 6 nodos por comunidad. El G2, estudiantes de cuarto año, evidenció 22 comunidades en total, de ellas siete son nodos únicos y el resto se distribuyó desde la talla de 5 nodos y más. En el caso de los profesionales, G3, si bien de las 20 comunidades, ocho tienen una talla pequeña, se debe tomar en cuenta que el total de nodos

(203) y aristas (302), superaron el número emitido por G2 que presentó 201 nodos y 265 aristas.

Mientras G1, grupo compuesto por 58 sujetos, presentó 220 nodos y 420 aristas, sus comunidades son visiblemente mas pequeñas que las comunidades detectadas en la red de ciencia y tecnología de los profesionales.



Figura 23. Red de G3 filtrada, en ciencia y tecnología.

La Figura 23 presenta la red de asociación de G3 filtrada con los mismos parámetros de los grupos anteriores. Con los estadísticos *grado de centralidad* y *rango de grado* se observa que los vocablos clave de esta red, según su centralidad y tamaño de nodo, se relacionan semánticamente en ciencia y tecnología de una forma más concreta que la de los otros grupos (G1 y G2).

En términos de modularidad, luego de los filtros aplicados, Gephi identificó seis comunidades:

1. Comunidad "7": investigación – patente – técnica.
2. Comunidad "8": instrumento – conicyt – investigador
3. Comunidad "11": proyecto – laboratorio – descubrimiento – desarrollo – progreso – avance – científico
4. Comunidad "18": experimento – conectividad

En un plano cualitativo se puede señalar que, mientras para los estudiantes de primer año las palabras clave en *ciencia* y *tecnología* corresponden a sustantivos que indican marcas (Microsoft, Apple), una rama científica (química, biología), o nombrar objetos vinculados a las tecnologías hápticas, como computador y smartphone; para los profesionales, las claves centrales con las que vinculan la temática de ciencia y tecnología tienen que ver con el uso, la aplicación e incluso con conceptos abstractos, que utilizados en contexto implican el conjunto estudiado. Es claro, también, que para los profesionales este centro de interés está más enfocado a la práctica científica, al evocar con gran peso vocablos como investigador, investigación, laboratorio, y sus resultados como avance, progreso, desarrollo, patente.

El grupo de estudiantes de cuarto año (G2) se encuentra en una especie de "medio camino", entre lo que reporta G1 y lo que hace por su parte G3; como se puede comprobar con las comunidades 7 y 8 del dicho grupo.

4.1.3.2. POLÍTICA

4.1.3.2.1. POLÍTICA SEGÚN GRUPO

La red de asociación semántica de *política* está compuesta por 607 nodos y 1218 aristas. Al igual que en la red ciencia y tecnología participaron tres grupos: G1, con 58 sujetos de primer año; G2, con 20 sujetos de cuarto año; y G3, compuesto por 20 periodistas profesionales. La Figura 24 presenta la red compuesta por los tres grupos; en esta figura se priorizó exponer la forma de la red por sobre las etiquetas de los nodos, con el objetivo de dimensionar la estructura que presenta una forma centralizada, pero que a la vez, muestra un gran número de nodos que se encuentran en la periferia asumiendo una forma *satelital* del centro de esta red.

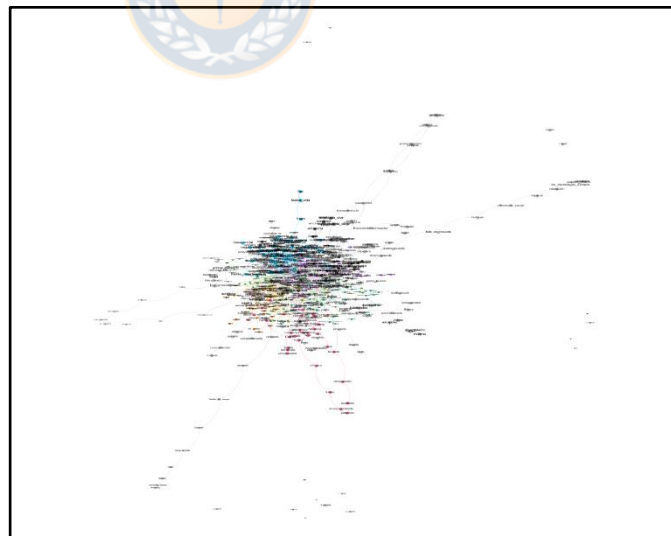


Figura 24. Red de política con tres grupos incluidos, al 100%.

i) ESTUDIANTES DE PRIMER AÑO (G1)

El G1 reportó un total de 377 nodos y 714 aristas. Para el análisis de este grupo se efectuaron las mismas mediciones que en el centro de interés *ciencia y tecnología*: nuevamente con el algoritmo de distribución *ForceAtlasII* y los parámetros: Escalado=100 y Gravedad=20.

Como la intención, al igual que en *ciencia y tecnología*, fue encontrar los vocablos claves dentro de la red y las comunidades de vocablos, se utilizaron los estadísticos aportados por el *software* Gephi: Modularidad, Componentes Conexos y Grado Medio, que permiten extraer de la red la información señalada, mediante las herramientas *Betweenness Centrality* y *Modularity Class*.

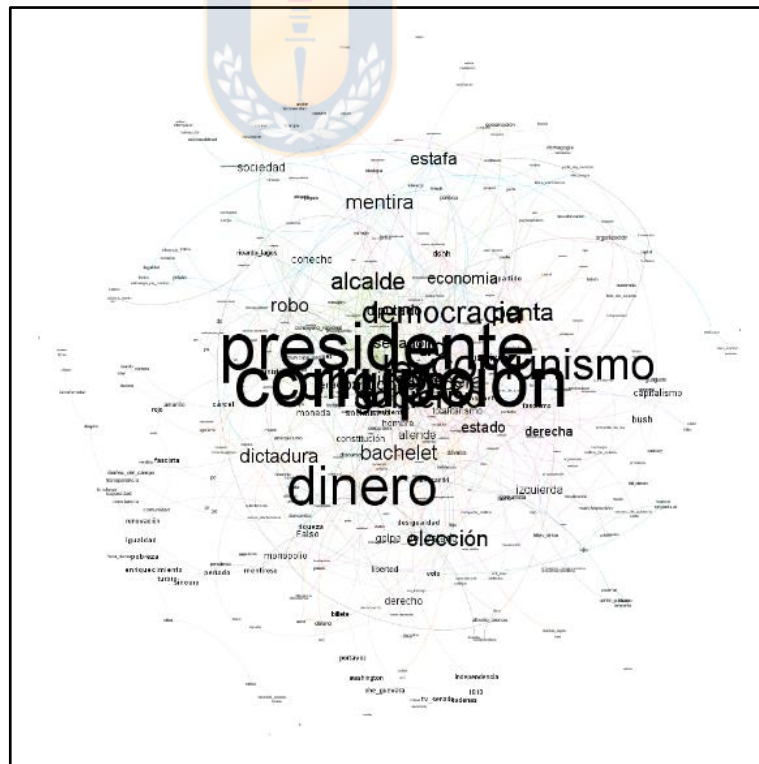


Figura 25. Red se asociación en política, G1.

Con la Figura 25 aparecen los primeros conceptos significativos para este grupo en torno a *política*. La figura muestra la red y sus vocablos clave según el análisis de centralidad interina. Los conceptos considerados como centrales por este grupo son *corrupción*, *presidente*, *dinero*, *partido_político* y *gobierno*. Más adelante se examinarán las comunidades que se desprenden de esta red. Por lo pronto, llama la atención que G1 considere a vocablos con carga emocional negativa como nodos centrales para definir esta temática.

Lo que podría indicar esta red de asociación es que G1 organiza su lexicón mental sobre *política* basado en características negativas, sin embargo, esta sería una apreciación adelantada y solo basada sobre los datos más llamativos de la red.

El estadístico de modularidad levantó 26 comunidades, 11 de ellas con una talla menor a 5. La distribución del resto de las comunidades se puede visualizar en el Gráfico 24. comunidades en política G1.

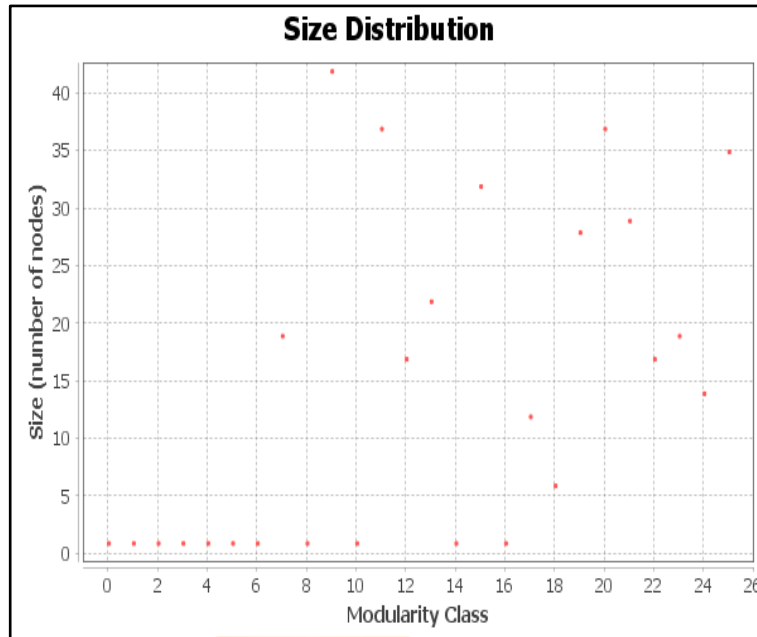


Gráfico 24. comunidades en política G1.

De las 26 comunidades detectadas por el software, se visualizaron aquellas que con el filtro topológico *rango de grado 6* y con los atributos de *centralidad interina*²⁷ que deja visibles aquellos nodos con un valor superior a 0,03. Lo anterior se puede ver en la Figura 26.

²⁷ La centralidad interina o *Betweenness centrality* tiene la opción de normalizar sus grados centralidad entre los parámetros 0 y 1, donde 0 es el nodo más alejado del centro y 1 es el que se en el punto central de la red.

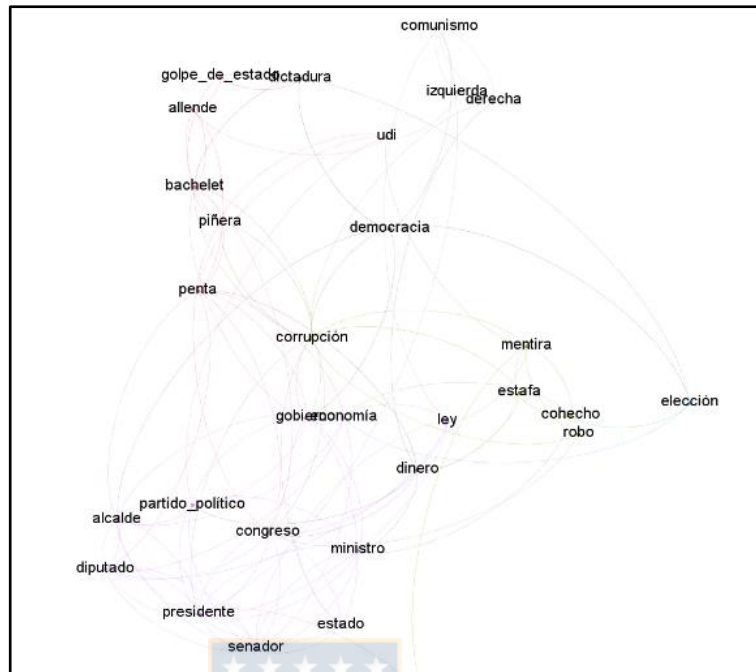


Figura 26. Red Política con filtro, G1.

Las comunidades obtenidas con este filtro son nueve y aquellos que tienen más de un nodo bajo estos parámetros se aglutinan de la siguiente manera:

1. Comunidad "11": corrupción – mentira – robo – estafa – cohecho
2. Comunidad "12": dinero – democracia – dictadura
3. Comunidad "16": penta – piñera – bachelet – allende – golpe_de_estado
4. Comunidad "17": comunismo – derecha
5. Comunidad "18": presidente – congreso – ley – gobierno – alcalde – ministro – partido_político – economía – senador – estado – diputado

ii) ESTUDIANTES DE CUARTO AÑO (G2)

En total, la red de asociación de política en G2 obtuvo 194 nodos y 259 aristas. Morfológicamente, bajo los mismos parámetros estadísticos y de ordenamiento utilizado en las redes anteriores, la de G2 se asemeja, a menor escala, de la red que representa a los tres grupos, un centro con una elevada cantidad de nodos y, a la vez, un grupo de nodos y aristas ubicados satelitalmente lo muestra la Figura 27.

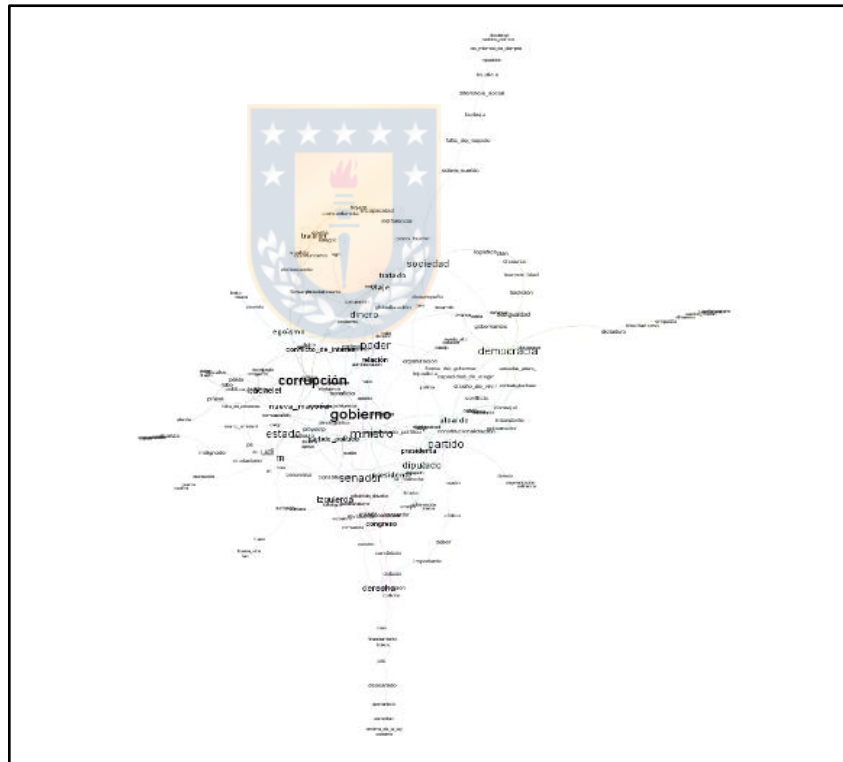


Figura 27. Red asociación semántica política, G2.

En términos léxicos, la red de G2 presentó similitudes en los vocablos con mayor índice de *centralidad interina*, los nodos más representativos de este índice

fueron los mismos cuyo grado²⁸ fueron los mayores: *corrupción, gobierno y ministro*.

El estadístico de *Modularidad*, en el Gráfico 25, identificó 21 comunidades de nodos, nueve de ellas de talla menor a $=2$ y el resto con una talla que partía desde los 10 nodos, es decir, la mayoría de las comunidades de política en G2 cuentan con un alto número de vocablos.

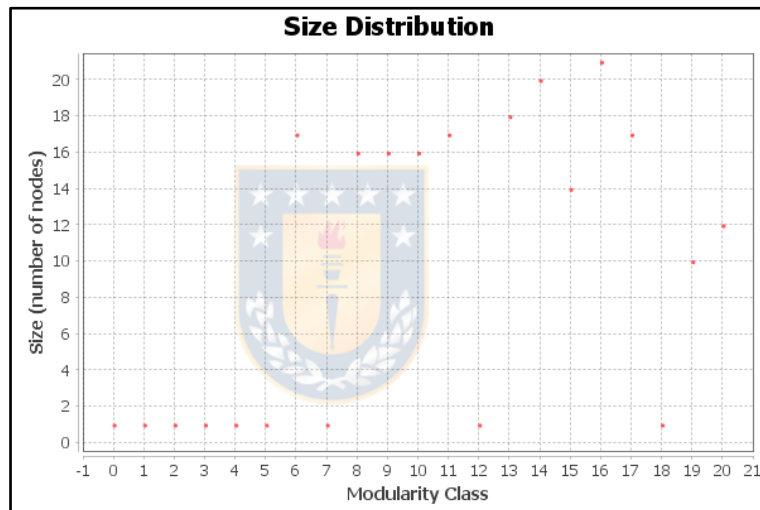


Gráfico 25. Comunidades en política G2.

Con Gephi se utilizó el filtro topológico *rango de grado = 4* y el atributo de visualización *grado de centralidad interina > 0.05*, el resultado visual de la red es el que se presenta en la Figura 28. Dicha figura muestra, además, mediante los colores las comunidades que quedan presentes luego de los filtros y la distribución

²⁸ Aquellos nodos con más aristas, y se mide sumando tanto las que reciben como las que salen de él.

más cercana al centro de la red. Lo anterior implica que para G2 *corrupción* y *gobierno* son los vocablos más centralizados en su lexicón mental.

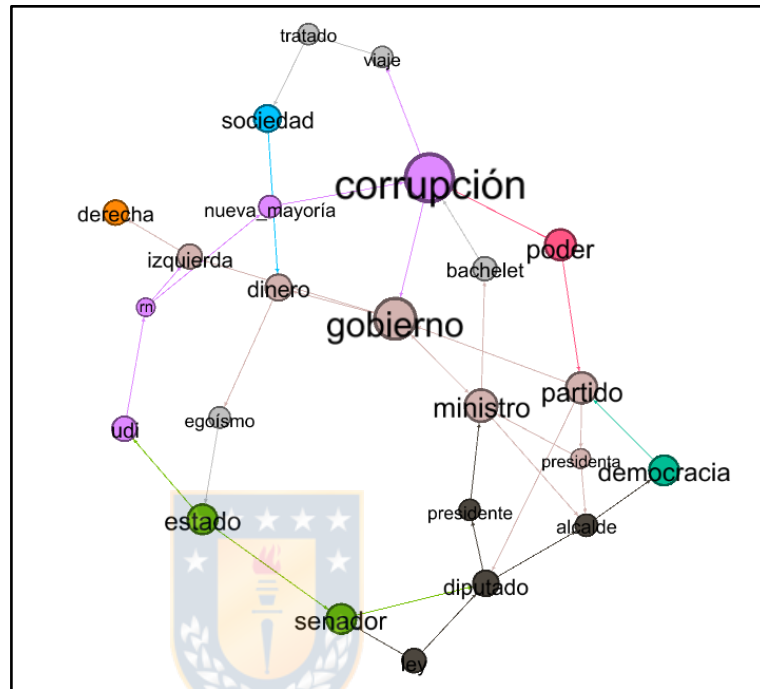


Figura 28. Política G2, con filtros.

A su vez, las comunidades desprendidas de estos vocablos son:

1. Comunidad "9": gobierno – ministro – partido – dinero – izquierda – presidenta
2. Comunidad "6": ley – alcalde – presidente – democracia
3. Comunidad "14": estado - senador
4. Comunidad "16": corrupción – rn – udi - nueva_mayoría
5. Comunidad "19": egoísmo – tratado - viaje

iii) PROFESIONALES (G3)

La red de asociación semántica de los profesionales está compuesta por 194 nodos y 314 aristas. Morfológicamente, con los mismos parámetros utilizados en las redes anteriores, es similar a la red de G2 y a la red total (de los tres grupos: G1, G2 y G3), con un centro que aglutina a los nodos más centralizados y de mayor grado.

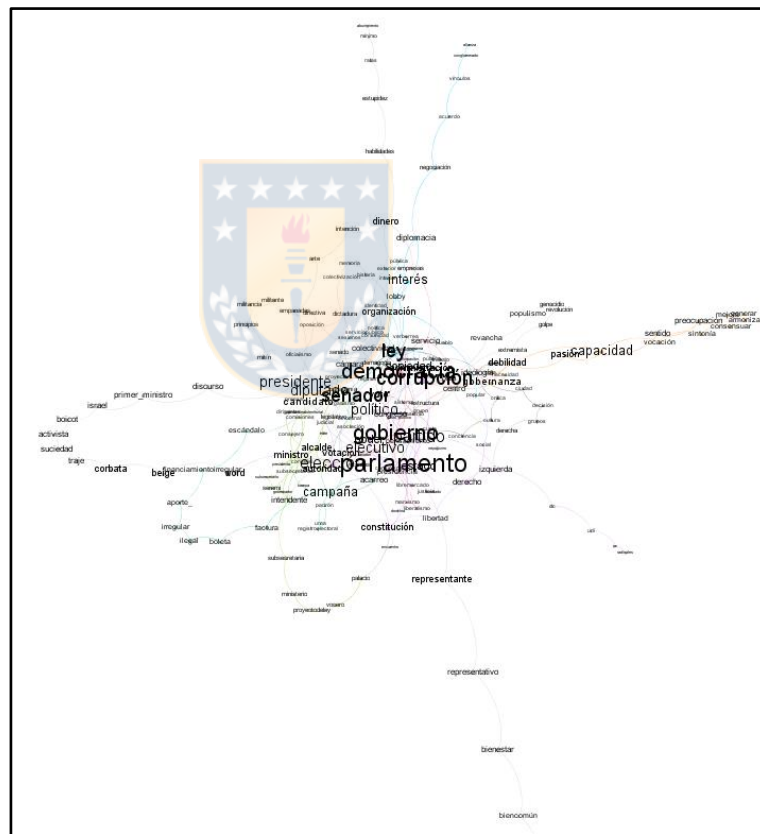


Figura 29. Red en tema Política, G3.

En términos léxicos los nodos destacados por su *grado* y por su *centralidad interina* en la red de G3 difieren de los otros grupos en un elemento fundamental:

mientras el lexicón mental central de G1 y G2 se caracteriza por sus nodos cuyo significado implican una valoración, que es negativa en ambos casos (*corrupción*), la red de los profesionales tiene como vocablos centrales nombres que no implican valoración como *parlamento*, *democracia*, *gobierno*. El vocablo *corrupción*, cuya centralidad y grado fue de gran preponderancia en los grupos anteriores, en la red de G3 también estuvo presente con una centralidad =0,14; sin embargo, su grado promedio fue =9.

Con el estadístico modularity class, Gephi identificó 23 comunidades de diversas tallas en esta red, como se observa en el Gráfico 26. De éstas, nueve corresponden a microcomunidades, cuyas tallas no superan los dos nodos; el resto parte con comunidades de talla =6 nodos, y la mayor cuenta con una talla =26, la que reúne la mayor cantidad de vocablos.

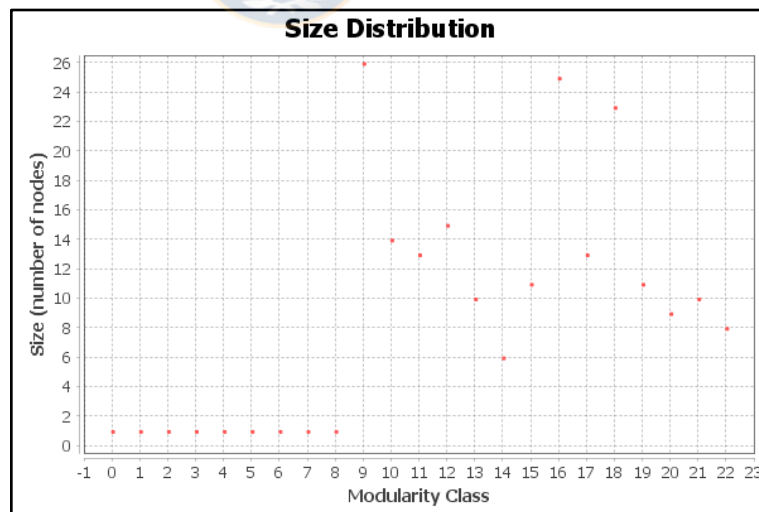


Gráfico 26. Distribución de las comunidades de política en G3.

Una vez identificadas las comunidades, se procedió al análisis de los nodos más relevantes de la red. Para ello se efectuó un filtrado con dos parámetros: el filtro topológico *rango de grado* >6 y el filtro de atributos *betweenness centrality* $> 0,05$. Lo anterior se puede ver en la *Figura 30*.

Con dichos parámetros de visualización, la red de G3 en política quedó compuesta por los nodos de mayor grado y más cercanos al centro de la organización del lexicón de este grupo. Los parámetros dejaron a la vista siete comunidades; de éstas, dos quedaron sólo con un nodo visible. Para G3, los nodos con mayor grado fueron *democracia*, *gobierno* y *ley*. Los vocablos con un mayor grado de *centralidad interina* fueron *parlamento*, *democracia* y *gobierno*.



Figura 30. Red filtrada de política, G3.

Las comunidades que conforman esta especie de *núcleo* de la red semántica en política son:

1. Comunidad "9": estado – elección – parlamento – partido – corrupción – poder – ejecutivo.
2. Comunidad "10": gobernanza – capacidad
3. Comunidad "11": gobierno – democracia
4. Comunidad "18": ley – interés
5. Comunidad "19": diputado – senador – político – candidato – presidente

4.1.3.3. DEPORTE SEGÚN GRUPO

La red de asociación de la temática *deporte* está compuesta por 609 nodos y 1430 aristas, morfológicamente²⁹ es similar a la red *política*, aunque con menos nodos satelitales. Presenta, además un alto número de comunidades de vocablos, que en la Figura 31 están representados por distintos colores. Para la elaboración de esta red se trabajó con el léxico disponible de los mismos grupos anteriores, esta vez con el centro de interés *deporte*. Cabe destacar que en este centro de interés no se encontraron diferencias significativas entre los grupos al analizar los indicadores propios de los estudios de disponibilidad léxica: Índice de disponibilidad léxica individual, Índice de cohesión, ni Promedio de palabras.

²⁹ Siguiendo los mismos parámetros de distribución utilizado en las redes anteriores.

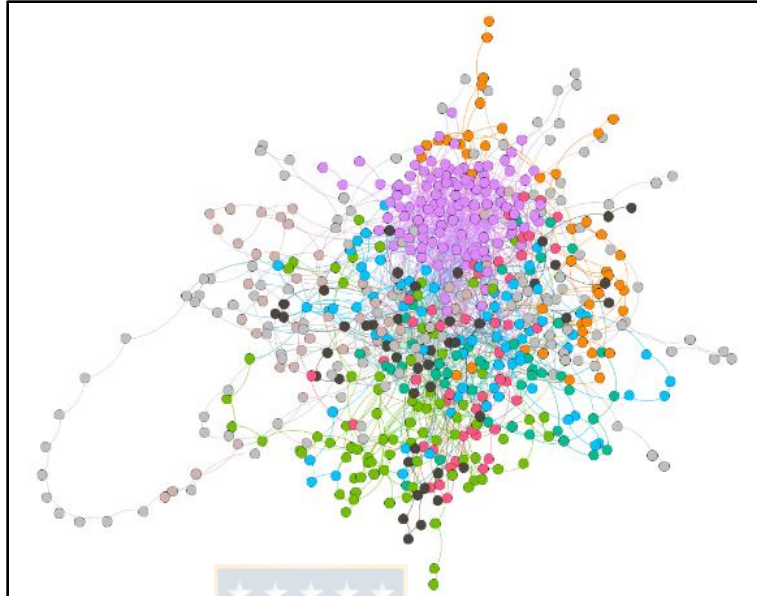


Figura 31. Red total de Deporte.

i) ESTUDIANTES DE PRIMER AÑO (G1)

El G1 presentó un total de 411 nodos y 815 aristas. Morfológicamente, siguiendo los mismos parámetros de distribución de las redes anteriores, en la Figura 32 se aprecia una red con un alto contenido de nodos en el centro. Además, presenta dos especie de colas en las cuales se unen nodos de baja centralidad interina y bajo rango de peso. Una vista más detallada indica que dichos nodos corresponden a nombres de deportistas destacados especialmente en fútbol.

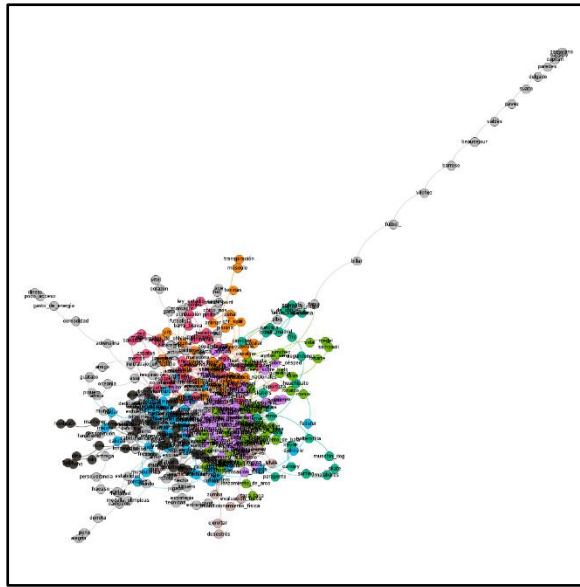


Figura 32. Red Deporte G1.



En términos de comunidades, según el estadístico *Modularity class* de Gephi, la red contiene 29 comunidades. Al igual que en las redes anteriores, existe un alto porcentaje de comunidades que presentan talla pequeña, es decir, que no agrupan más de dos nodos en su comunidad, la mayor talla de comunidad, es decir la que agrupa el más numeroso, llega a reunir más de 50 nodos, según muestra el Gráfico 27.

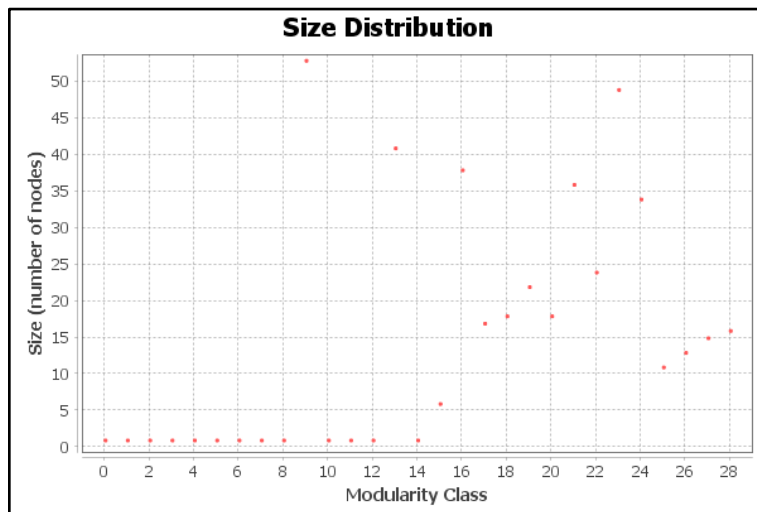


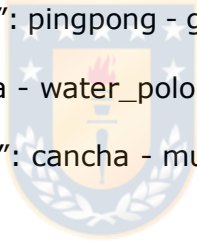
Gráfico 27. Comunidades en deporte, G1.

Una vez identificadas las comunidades, se procedió al análisis de los nodos más relevantes de la red. Para ello se efectuó nuevamente el filtrado con dos parámetros, el filtro topológico *Rango de grado* > 8 y el filtro de atributos *Betweenness centrality* > 0.05 .

La *centralidad interina*, en este caso, la entrega el vocablo *natación*, cuyo grado normalizado en una escala de 0 a 1 corresponde a 0.022; luego vienen, según este mismo parámetro, los vocablos: *fútbol*, *tenis*, *pelota* y *atletismo*; aunque no todos pertenecen a la misma comunidad. Hecho que se replica con el estadístico *grado medio*, es decir de los nodos que obtuvieron más aristas de entrada y salida.

Las comunidades de nodos que quedan visible bajo estos filtros son:

1. Comunidad "9": básquetbol - voleibol - natación - hockey - patinaje - rugby - fútbol_americano - atletismo - béisbol - bicicleta - gimnasia - estadio - ciclismo - equitación - juegos_olímpicos - polo
2. Comunidad "13": salud - vida_sana - competencia - vida
3. Comunidad "16": correr - sudor
4. Comunidad "19": danza - gimnasia_rítmica - ajedrez -karate - yoga
5. Comunidad "21": fútbol - pelota - colocolo - pasto
6. Comunidad "23": pingpong - gimnasia_artística - taekwondo - trote - golf - jabalina - water_polo - hándbol
7. Comunidad "24": cancha - mundial



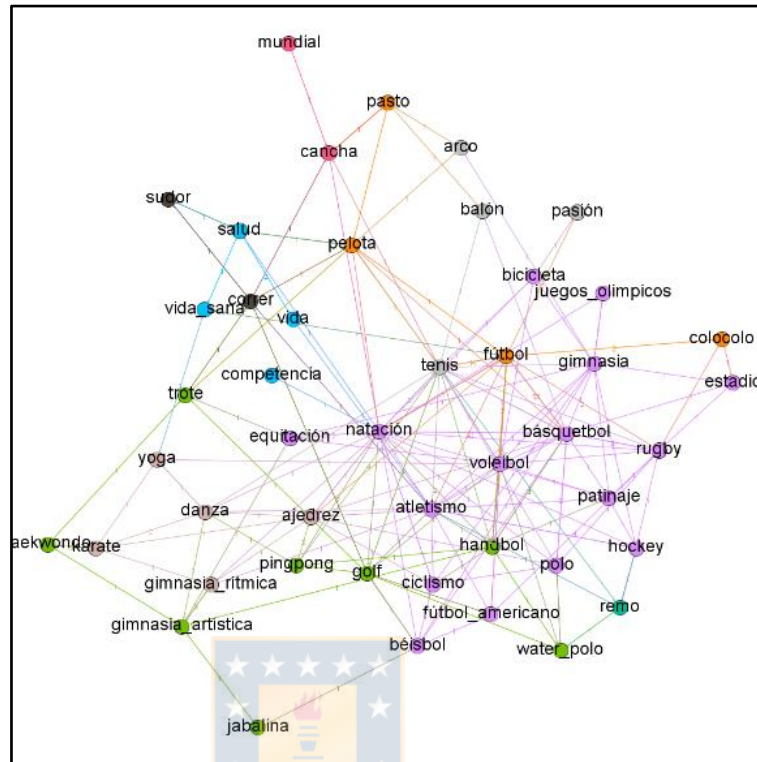


Figura 33. Red filtrada deporte, G1.

La Figura 33 representa la red con los nodos y aristas luego del filtrado realizado, los distintos colores, representan, a la vez, las comunidades de palabras presentes en dicha red. Como referencia se debe indicar que, si imaginamos la red dentro de una circunferencia, el vocablo más cercano al punto central del diámetro es aquel que tendría grado = 1. En esta red, el nodo más cercano a dicha centralidad es, como ya se indicó, el vocablo *natación* cuyo grado de *centralidad interina* fue de 0.02, muy alejado del núcleo imaginario que hemos planteado. Sin embargo, como las redes tienen la capacidad de variar según diversas circunstancias y necesidades, para el momento en que la prueba de

disponibilidad léxica fue tomada, *natación* se configuró como el vocablo que los sujetos utilizaron como clave para llegar a otros vocablos.

ii) ESTUDIANTES DE CUARTO AÑO (G2)

G2 presentó un total de 225 nodos y 316 aristas. En cuanto a su forma, siguiendo los mismo parámetros de distribución anteriores, se aprecia una red con un alto contenido de nodos en el centro. Además, presenta al igual que G1, dos especies de "colas" en las cuales se unen nodos de baja centralidad interina y bajo rango de peso.

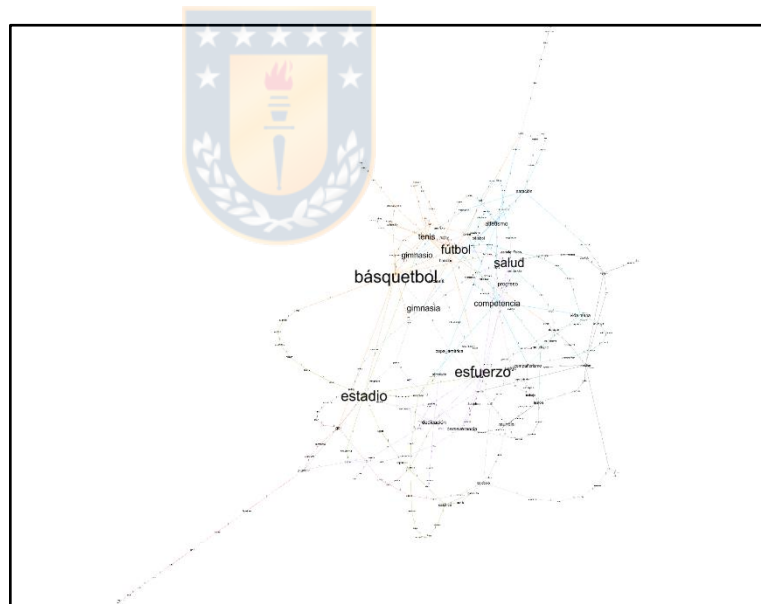


Figura 34. Red deporte G2.

Para G2 el reporte entregado por *Modularity class*, señala que existen 13 comunidades, las que a diferencia de G1 sí cuentan con una talla superior. De

hecho, el 92,4% de las comunidades presentes en esta red tienen una talla >10. Así indica en el Gráfico 28 sobre la distribución de las comunidades de nodos.

En términos de centralidad, el nodo clave para la red de G2 en deporte es el vocablo *básquetbol*. Cercano a este se encuentra el nodo *esfuerzo*, y más alejados los nodos *estadio* y *fútbol*.

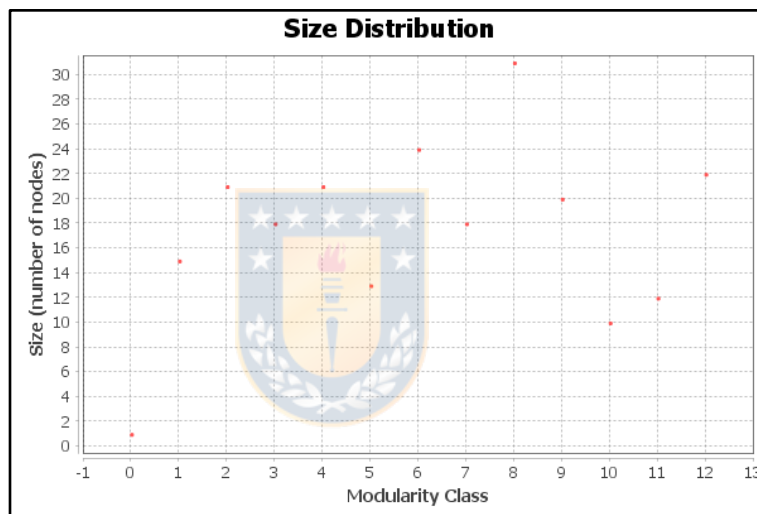


Gráfico 28. comunidades de G2 distribuidas según talla.

El siguiente paso que se ha establecido para el análisis de las redes de asociación semántica, es realizar el filtrado de los datos y así quitar de la visualización los nodos con menos preponderancia y más alejados del núcleo de la red. Para ello se utilizó el filtro topológico *rango de grado > 7* y *Betweenness centrality > 0.05*

Como ya se ha indicado, el vocablo clave -en términos de centralidad- que reporta esta red es *básquetbol*, con un grado de centralidad interina >0.23 . Las comunidades visibles bajo estos filtros, y siempre dejando comunidades desde dos nodos visibles, son siete:

1. Comunidad "1": mundial – compañerismo
2. Comunidad "3": vida_sana – saludable – estado_físico
3. Comunidad "4": fútbol – tenis – básquetbol – rugby
4. Comunidad "5": copa_américa – gimnasia
5. Comunidad "8": salud – esfuerzo – progreso
6. Comunidad "10": gimnasio – crossfit – hándbol
7. Comunidad "12": competencia – béisbol – natación

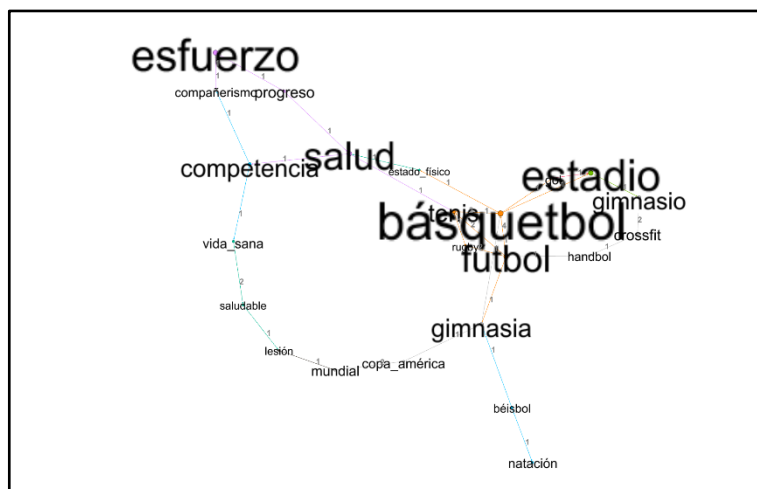


Figura 35. Red filtrada de deporte, G2.

En términos del léxico presente en la Figura 35, reflejo de la red filtrada según los parámetros descritos más arriba, se puede señalar que el vocablo clave, en términos de centralidad interina y grado del nodo, forma la comunidad "4". Dicho grupo está compuesto también por los vocablos *tenis*, *rugby* y *fútbol*, todos nombres que corresponden a disciplinas deportivas.

El siguiente nodo clave, según centralidad interina y grado, fue el vocablo *esfuerzo* que junto a *salud* y *progreso*, forman parte de la comunidad "8". Estos tres nodos en conjunto se refieren a características que los sujetos asocian a la temática deporte, esta idea continúa cuando se aumenta la visualización de comunidad "8" con los nodos que fueron filtrados de la red ya que no cumplían con las características utilizadas hasta ahora.

iii) PROFESIONALES (G3)

La red de asociación de G3 está conformada por 206 nodos y 373 aristas. Su morfología, bajo los parámetros establecidos en esta investigación, se asemeja a la forma presentada por la red de G1: un núcleo de nodos y varias órbitas satelitales con nodos que presentan, visualmente, menor tamaño que el resto (Figura 36).

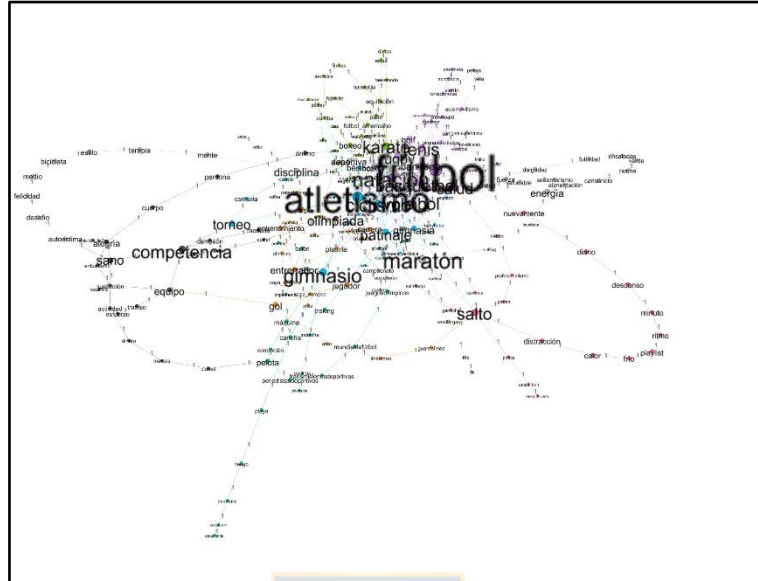


Figura 36. Red Deporte, G3.

Las características que entregan los estadísticos *Betweenness Centrality* y *Modularity* corresponden –como ya se ha mencionado– al grado de centralidad que ciertos nodos tienen dentro de la red. Dichos nodos son llamados vocablos clave, debido a que sirven como unión de paso para otros nodos presentes en la red. Por otra parte, el estadístico *Modularity* realiza un levantamiento de los grupos de comunidades en las que se organizan los nodos.

En este caso se identificaron 11 comunidades de palabras. De éstas sólo una es una comunidad aislada, con una talla baja; el resto presenta tallas por sobre 10 nodos. El detalle de la distribución queda explícito en el Gráfico 29, donde se muestra que la comunidad más grande es la “1”, que agrupa a más de 30 nodos.

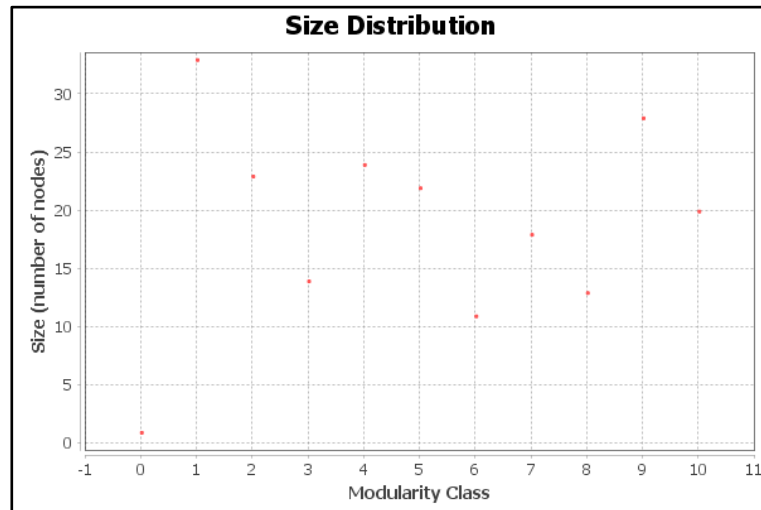


Gráfico 29. Distribución de las comunidades en red Deporte G3.



Una vez establecidas las comunidades presentes en la red, se visualizan los nodos claves, mediante el estadístico *Betweenness centrality*, que calcula la centralidad interina de los vocablos. En este caso, el nodo que se encuentra más cercano al núcleo de la red de asociación semántica del grupo compuesto por profesionales es el vocablo *fútbol*, con un grado de centralidad = 0.24.

Con estos datos filtrados y ordenados, Gephi permite la visualización deseada que en este caso se basa en la centralidad del nodo como elemento de vínculo entre sus pares. En la red se filtraron todos aquellos nodos cuyo *Betweenness Centrality* >0,05 y, a la vez, aquellos nodos cuyo *rango de grado* >5. Con dichos filtros las comunidades visibles son:

1. Comunidad "1": tenis – fútbol – natación – hándbol – golf – rugby
– básquetbol – pingpong – voleibol – salud – hockey

2. Comunidad "4": atletismo – patinaje – ciclismo – gimnasia – gimnasio – béisbol
3. Comunidad "5": carrera – atleta – triatlón
4. Comunidad "6": maratón – judo
5. Comunidad "9": fútbol_americano - boxeo - polo - karate – equitación

En términos léxicos, el núcleo de esta red, cuyo nodo clave es el vocablo *fútbol*, perteneciente a la comunidad "1", se agrupa junto a otros nodos que también implican el nombre de una disciplina deportiva, son sustantivos que, con este nivel de visualización, no presentan valoraciones.

E El siguiente nodo preponderante en la red es el vocablo atletismo, el cual pertenece a la comunidad "4" que también la forman nombres de disciplinas deportivas. Lo anterior ocurre en toda la red visualizada, la cual se caracteriza por ser una red que enuncia nombres y que no entrega valoraciones sobre el

concepto del centro deporte, algo que sí se visualizó en las redes de G1 y G2. enuncia nombres y que no entrega valoraciones sobre el concepto.

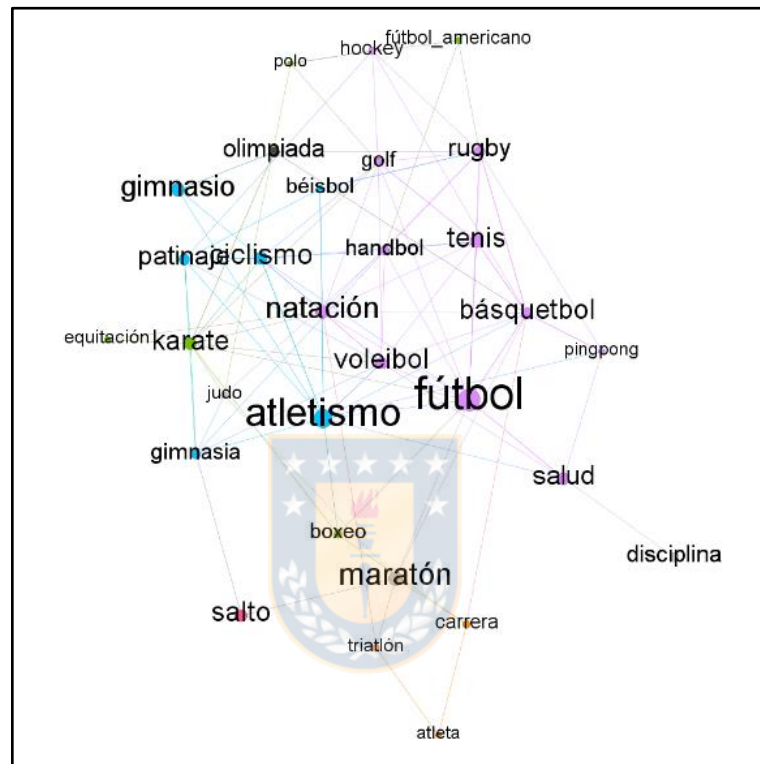


Figura 37. Red filtrada en Deporte, G3.

4.2. ANÁLISIS GENERAL SEGÚN OBJETIVOS ESPECÍFICOS

La finalidad de los análisis descriptivos es definir, clasificar o categorizar fenómenos de interés, ver cómo son y cómo se manifiestan, intentando especificar propiedades, características y perfiles de personas, grupos, comunidades, etc. (Herrera, Martínez y Amengual, 2011). A continuación, respondiendo a los objetivos específicos de este trabajo, se realizará un análisis descriptivo de los resultados obtenidos.

1. Objetivo específico 1. Medir la disponibilidad léxica de la muestra en los centros de interés específicos: ciencia y tecnología, deporte y política.

Como se ha señalado en capítulos anteriores, la disponibilidad léxica se mide a través de diversos indicadores, tales como el Índice de cohesión (IC); Número de palabras distintas (NPD); Promedio de respuesta por sujetos (XR) y el Índice de disponibilidad léxica individual (IDLI). Los datos expuestos en el punto 4.1.3.1 indican que, al analizar los grupos que participaron del estudio como un todo, los índices de disponibilidad léxica que éste presenta en los tres centros de interés son significativos, incluso si se comparan con investigaciones en las cuales los participantes superaron en número a este trabajo (Echeverría et al., 1987; Ferreira y Echeverría, 2010; Hernández, 2007).

En términos de Número de palabras distintas (NPD), también nombradas durante este estudio como vocablos, G1, G2 y G3 emitieron más de 450 palabras distintas en ciencia y tecnología, política y deporte (ver Gráfico 6), superando incluso lo reportado en centros de interés tradicionales como partes del cuerpo y medios de transporte. La riqueza, en términos de NPD de los centros de interés, deja a deporte como la temática con un mayor NPD (609), luego política con 590 NPD y, en tercer lugar ciencia y tecnología con 477 NPD.

Si bien los resultados cambian al ser observados bajo el índice de Promedio de respuesta (XR), en el cual el centro de interés partes del cuerpo es el mayor con un promedio de 19,64 palabras por sujeto, los tres centros de interés propuestos para este estudio fluctúan entre las 11,95 palabras en ciencia y

tecnología y las 17,73 en deporte, mientras que política alcanza un promedio de 15,52 por sujeto.

Finalmente, el Índice de cohesión con la muestra de sujetos en general, indica que el centro de interés más cohesionado es deporte (IC=0,029); luego política (IC=0,026) y finalmente ciencia y tecnología (IC=0,025).

Bajo este análisis, no se encontraron diferencias significativas entre los centros de interés.

2. Objetivo específico 2. Comparar la disponibilidad léxica entre estudiantes y profesionales.

El análisis se realizó desde los indicadores de disponibilidad léxica. En términos de NPD según centro de interés, los estudiantes de primer año (G1) presentaron en total 1482 palabras distintas: NPD=220 en ciencia y tecnología, NPD=366 en política y NPD= 411 en deporte. Por su parte, los estudiantes de cuarto año (G2) obtuvieron en total 797 vocablos con un NPD=194 en ciencia y tecnología, NPD=193 en política y NPD=225 en deporte. Mientras que los profesionales (G3) evocaron 892 vocablos en total, obteniendo en ciencia y tecnología un NPD=199, en política un NPD=194 y en deporte NPD=206.

En el indicador Promedio de respuesta por sujetos (XR), G1 pese a ser el grupo con mayor cantidad de sujetos, obtuvo un XR=9,17 en ciencia y tecnología, XR=14,75 en política y un XR=16,69 en deporte. El G2 obtuvo promedios de palabras similares a G1: XR=15,05 en ciencia y tecnología, XR=14,38 en política y XR= 17,24 en deporte. Finalmente, los profesionales (G3) obtuvieron el mayor

promedio de los tres grupos en todos los centros de interés estudiados: $XR=16,9$ en ciencia y tecnología, $XR=18,9$ en política y $XR=21,5$ en deporte.

En tanto el Índice de cohesión (IC) alcanzado por cada grupo en los centros de interés, que se puede observar en detalle en la Tabla 25, deja claro que la mayor coherencia entre sus resultados lo obtiene G3 (profesionales), con un $IC=0,084$ en ciencia y tecnología, $IC=0,097$ en política y un $IC=0,13$ en deporte.

3. Objetivo específico 3. Evaluar la riqueza léxica de los textos producidos por la muestra, para compararla entre los grupos de estudiantes y profesionales.

La riqueza léxica que estudiantes y profesionales presentan en sus textos fue largamente abordada en el punto 4.1.2. Se efectuaron mediciones basadas en los indicadores diversidad léxica y densidad léxica a nivel de sexo de los participantes, donde no hubo diferencias significativas, las cuales fueron comprobadas por pruebas estadísticas no paramétricas. Donde sí hubo diferencias significativas en los índices fue al utilizar la variable grupo como predictora. Las diferencias, no atribuibles al azar o a la coincidencia, indican que en términos de densidad léxica los estudiantes y los profesionales presentan diferencias significativas; también ocurre al analizar los indicadores de diversidad léxica total, diversidad léxica aleatoria y diversidad léxica del lead (ver Tabla 16).

Otro indicador que es sensible en este estudio es la variable dependencia administrativa del establecimiento educacional, especialmente en los indicadores densidad léxica y diversidad léxica total. En esta variable quedó comprobado que, para esta muestra, quienes egresaron de una institución privada cuentan con una

diversidad y densidad léxica mayor de la que presentan los egresados de enseñanza municipalizada y subvencionada.

4. Objetivo específico 4. Identificar las redes semánticas de la muestra y realizar comparaciones entre los grupos de sujetos y los centros de interés.

Los resultados relacionados con las redes semántica se encuentran presentes en el punto 4.1.3 ordenados según temática y grupo. En este tipo de estudio se pudo comprobar que las redes de asociación de los estudiantes de primer año se refieren, en gran medida, a vocablos que tienden a nombrar elementos y nombres propios que asocian a cada temática, Además, presenta una tendencia valorativa de los temas (e.g. deporte-sano; política- corrupción) que queda evidenciada en las comunidades identificadas en la investigación.

Estos resultados cambian al comparar el léxico que asocian los profesionales en cada temática, quienes optan por un vocabulario con mayor tecnicismo y menor valoración de los campos estudiados. Las redes de los estudiantes de cuarto año, en tanto, muestran elementos de ambos grupos, tanto valorativos o emocionales, como un vocabulario más concreto similar al de G3.

CAPÍTULO 5: CONCLUSIONES

Para comenzar este capítulo se recordará, en primer lugar, el objetivo de esta investigación, el cual fue obtener una descripción de la competencia léxica de estudiantes de primer año de la carrera de periodismo, en el discurso especializado escrito –entendido como una noticia informativa–, mediante una aproximación léxica basada en los estudios de disponibilidad léxica, riqueza léxica y asociación semántica.

En un principio la idea pareció un tanto sencilla y bastante numérica. A través del léxico se podría lograr potenciar un modelo o aproximación que permitiera a una carrera de periodismo establecer la competencia léxica de los estudiantes que reciben, y desde ahí poder implementar mejoras y puntos de control en la evolución de la habilidad escrita de sus futuros profesionales. Es por ello que se decidió considerar a estudiantes de primer año y de un curso superior –cuarto año–, que según su currículo universitario ya hubiese cursado asignaturas relativas a la escritura, como son, en el caso de la carrera de Periodismo de la Universidad de Concepción, las pertenecientes a la línea de *redacción periodística*, electivos especializados en cultura, deporte o ciencia; además de ya haber tenido un primer acercamiento a lo que es un medio de comunicación real –tras haber cursado tres asignaturas semestrales que los obligan a trabajar fuera de la carrera: Pasantías I, Pasantías II y Prepráctica–, con sus límites de tiempo, precisión en los datos y otros objetivos de dichos cursos universitarios.

Al mismo tiempo se pensó en obtener un parámetro, en otras palabras, establecer cuál era el nivel que debe alcanzar un estudiante o qué tan alejado de éste se encuentran. Para ello se solicitó a un grupo de periodistas que facilitaran sus textos periodísticos y participaran de la misma prueba de disponibilidad léxica que los estudiantes.

Los resultados, presentados en el Capítulo 4, permitieron no sólo analizar a la riqueza léxica de los estudiantes, su léxico disponible y cómo está organizado a nivel de grupo su lexicón mental, también posibilitaron establecer e indagar la existencia de diferencias entre grupos en las variables sexo, tipo de establecimiento educacional del cual egresó y nivel universitario (primer año, cuarto año y profesional).

En el plano de la riqueza léxica se puede describir al grupo de primer año, G1, como un grupo capaz de elaborar textos con un alto nivel de diversidad léxica, especialmente cuando la medición se realiza con el sistema de diversidad léxica aleatoria –explicada en el Capítulo 3: Metodología–, índice bajo el cual se acercan, en promedio, a un TTR =0,6 en los tres ámbitos temáticos abordados: ciencia y tecnología, política y deporte. Por otra parte, este grupo presenta una alta diversidad léxica en el lead, que sobrepasa el TTR =0,65, lo que implicaría una entrada de la noticia con una importante fuente informativa, ya que se podría suponer que en esas 200 primeras palabras, los estudiantes son capaces de organizar y exponer discursivamente una serie significativa de hechos.

El problema surge cuando el índice de riqueza léxica es abordado también por la densidad léxica de un texto. Se ha definido como densidad léxica la relación que existe entre el total de palabras de contenido semántico y la extensión de un texto. Al remover todas las palabras gramaticales (artículos, conjunciones, y verbos auxiliares, entre otras categorías) en el texto sólo quedan vocablos con significado, los que contienen la información relevante de un hecho. Este indicador es en $G1 = 0,3$; es decir, por cada 10 palabras que hay en un texto sólo tres son palabras de contenido que aportan, en este caso, a la función informativa del texto noticioso.

Bajo estos mismos indicadores, el G2, compuesto por estudiantes de cuarto año, demuestra una riqueza léxica más elevada en sus textos, tanto en la diversidad léxica aleatoria y en la diversidad léxica en el *lead*, de las noticias que elaboran. Por otra parte, la densidad léxica es $=0,53$. Si lo visualizamos en términos porcentuales, este grupo tiene aproximadamente un 20% mejor reporte en sus noticias escritas debido a que su utilización de palabras con contenido semántico se ve más equilibrada con relación a las palabras gramaticales.

Dentro de los objetivos específicos de este trabajo, se propuso comparar la riqueza léxica de los estudiantes con la de profesionales. Al respecto, los índices que arroja G3, compuesto por periodistas, no parecían muy distanciados de los obtenidos por los estudiantes; incluso en la diversidad léxica del *lead* fue menor que lo obtenido por los estudiantes, de cuarto año. Los profesionales marcan una diferencia significativa en el índice de densidad léxica, donde $G1 = 0.33$, $G2 = 0.43$, y $G3 = 0.5$.

Cuando abordamos la competencia de estudiantes de periodismo, en términos de riqueza léxica en textos de un discurso especializado, se pueden caracterizar como estudiantes cuya competencia léxica les permite elaborar noticias informativas con un nivel de diversidad léxica medio-alto, con buenos promedios de diversidad especialmente en la elaboración de lead noticiosos, situación que se mantiene entre los estudiantes de cuarto año. La diferencia radica en la utilización de palabras con contenido que aporten informativamente al texto que elaboran. Algo en lo que concuerda Zuluaga (2006) al señalar que los textos periodísticos tienen una configuración lingüística especial, que debe adecuarse para transmitir la mayor cantidad posible de información en textos de extensión reducida, imposición que los convierte precisamente en prototipos de los textos informativamente densos.

Según los resultados, los textos elaborados por estudiantes, especialmente por los de primer año, no logran cumplir con este requisito. Sin embargo, se puede inferir que, en el transcurso de su formación profesional, el alumno de periodismo mejora esta capacidad hasta llegar a la que un periodista profesional debe ser capaz de demostrar.

Todo lo anterior implicaría que, si bien los estudiantes son capaces de abordar un hecho noticioso con un vocabulario con tanta variedad como los profesionales, fallan en el momento de evaluar sus noticias desde un punto de vista informativo. Este no es un hecho menor, una noticia está constituida por la

exposición de hechos concretos de la realidad, los cuales son -obligatoriamente- manifestados mediante palabras nocionales o con contenido semántico, y no por una cantidad considerable de palabras distintas sin contenido

El siguiente plano desde el cual se propuso abordar la competencia fue desde los estudios de disponibilidad léxica. En este ámbito se observaron interesantes factores que ayudan a complementar lo hallado en riqueza léxica. En primer lugar, se confirma que la variable sexo, si bien presenta diferencias, no es significativa en la evocación de conceptos para estudiar el léxico disponible, por lo tanto la condición del género en este caso no implica una variable que discrimine a la hora de evocar palabras. También se explicitó que las temáticas - centros de interés- ciencia y tecnología, política y deporte son grupos de bajo grado de coherencia semántica, debido a sus bajos resultados obtenidos en los tres grupos; el índice de cohesión (IC) en ciencia y tecnología fue =0,025, en política fue =0,026, mientras que en deporte fue =0,029. Un posible explicación a estos resultados sería que al ser categorías con un carácter más abstracto que el hecho de nombrar solo partes del cuerpo, en el lexicón de los estudiantes no existe una unidad definas de palabras, o rasgos asociados a estas temáticas - según la teoría de Marconi, que dé cuenta del mencionado centro de interés.

Debido a las diferencias en la cantidad de sujetos de cada grupo - especialmente en G1 que contó con 58 estudiantes, mientras que en G2 y G3 participaron 20 sujetos en cada uno-, los resultados a primera vista parecían abrumadores. G1 siempre fue el grupo que proporcionó el mayor número de palabras en los cinco centros de interés estudiados, seguido muy de lejos por G3.

La comprensión de los datos se observó en perspectiva al analizar desde indicadores como promedio de palabras por grupo, donde pese a ser G1 el más numeroso, las diferencias con los otros dos no alcanzaban a ser significativas; incluso fue el grupo cuyo promedio de palabras fue el menor en los cinco centros de interés. Así también, G1 presentó sistemáticamente los IC más bajos en los centros de interés, mientras que los profesionales (G3) los más altos; esto implica que el léxico disponible de los estudiantes si bien es amplio y diverso –al igual que lo hallado en riqueza léxica–, no contiene un núcleo que logre cohesionar el significado semántico de las temáticas estudiadas. Lo anterior se puede concluir incluso considerando que, de los cinco centros de interés estudiados, tres corresponden a centros de bajo grado de coherencia semántica (Echeverría et al, 1987; Echeverría et al. 2008; Ferreira y Echeverría, 2010).

El tercer componente para la aproximación a la competencia léxica es el análisis de redes de asociación semántica. Mediante el uso de grafos se pudo acceder a la forma en la que los grupos organizarían su lexicón mental, para ello se siguió la línea de análisis llamada *Pathways meaning text network*, en la cual se busca identificar aquellos conceptos clave que sirven como una unión dentro de la red. A la vez, aprovechando las utilidades del *software* empleado –Gephi–, se obtuvo un listado de comunidades de palabras dentro de la red. Esta aproximación permitió un análisis cualitativo del léxico en cada grupo.

En ciencia y tecnología, G1 identificó como vocablo clave computador, y luego otro tipo de palabras que corresponden a nominalización de marcas tecnológicas y disciplinas científicas. Para G2, el concepto clave fue investigación,

seguido por otros conceptos asociados, por una parte, a la propia práctica de la investigación (*laboratorio, experimento, descubrimiento*), y por otra –al igual que en G1–, a la nominalización de marcas que asocian a dicho tema y tecnologías de uso común.

Finalmente, los profesionales tuvieron como nodo clave dentro de su red el vocablo investigación, y asociaron su red semántica a los conceptos como uso, aplicación e incluso ideas abstractas, que al utilizarlas en contexto implican la idea de ciencia y tecnología. Para G3 este centro de interés está enfocado a la práctica científica, al evocar con gran peso vocablos como investigador, investigación, laboratorio, y sus resultados como avance, progreso, desarrollo, patente. Mientras los estudiantes de primer año utilizaron como vocablos clave sustantivos que indican marcas (*microsoft, apple*), una rama científica (*química, biología*), o nombrar objetos vinculados a las tecnologías hápticas, como computador, celular y smartphone.

En el caso de la red de deporte, G1 utilizó como vocablo clave natación, nombre que corresponde a una disciplina deportiva; luego, se encuentran algunas asociaciones que corresponden a una especie de estímulo-respuesta, por ejemplo, se asocian estrechamente correr y sudar. Sin embargo, esta red tiene la particularidad de ser más homogénea para los tres grupos: si en G1 el vocablo clave fue natación, para G2 fue básquetbol y para G3 fútbol. Así también las vecindades de cada grupo mostraron similitudes de asociación.

Probablemente la red con diferencias más visibles fue política, una red que tanto para G1 como para G2 está constituida por valoraciones negativas hacia la política como práctica, donde el concepto corrupción fue el vocablo clave dentro de la red. La diferencia entre ambas redes es que mientras G1 acentúa la concepción de política bajo términos valorativos que en general son negativos, G2 muestra un balance al constituir su red semántica con vocablos asociados a ideas como presidente, senado, partido_político, entre otros. La red de G3 fue la menos valorativa con vocablos clave como democracia, gobierno y ley, los cuales son conceptos que aparecen constantemente en las noticias diarias y que por lo tanto están en el uso cotidiano de los profesionales.

Durante la investigación se obtuvieron logros inesperados. El más significativo fue la realización de un script que pudiera medir de manera automática –y en cada texto– los indicadores de diversidad y riqueza léxica los cuales, tenían la ventaja de realizar diferentes mediciones a la vez: diversidad léxica total, que realiza el cálculo con la extensión completa del texto; diversidad aleatoria, cálculo que toma una cantidad X de tokens X veces de manera aleatoria en el texto y, finalmente, el cálculo de la diversidad léxica del lead, medida que se efectúa realizando una estimación de la longitud de un lead –que para este caso fue de 200 palabras– sobre cual se mide la relación de palabras presentes.

Junto con lo anterior se logró adecuar elementos propios de la lingüística computacional para que el script pudiera ser sensible a la densidad léxica. Para ello se elaboró un extenso listado de stop words, que permitiera eliminar del texto las palabras que se consideran vacías, gramaticales o nocionales, así como otras

palabras que, en el contexto del discurso periodístico, son consideradas poco precisas, vagas o que algunos manuales de redacción periodísticas han indicado como vicios en el periodismo. Como resultado de esta preparación previa se realizó un análisis que en vez de trabajar con textos etiquetados, la cual se considera la forma tradicional de analizar la riqueza léxica, lo que hace es limpiar de los textos todas aquellas palabras no deseadas, dejando sólo las que constituyen palabras de contenido semántico dentro de la noticia.

Este tipo de depuración se efectúa para llevar adelante investigaciones relacionadas con lingüística computacional como las basados en análisis semántico latente y otras que buscan investigar sobre las relaciones semánticas dentro de un corpus.



5.1. LIMITACIONES DE LA INVESTIGACIÓN

Se identifica como primera limitación el bajo número de sujetos que participan en las muestras de profesionales y de estudiantes de cuarto año. Con una muestra de 98 sujetos, donde 58 corresponden a un grupo y 20 en cada uno de los restantes, la posibilidad de cometer errores estadísticos del Tipo II (no rechazar una hipótesis nula cuando es falsa) tiene una alta probabilidad de ocurrencia, es por lo que se ha decidido que en aquellos puntajes muy cercanos a un $\alpha=0,05$ de significancia se ha optado por aceptar la hipótesis nula. Éste fue el caso ocurrido con la variable sexo en el estudio de disponibilidad léxica, al analizar el resultado del indicador total de palabras para el centro de interés

partes del cuerpo: el resultado fue $\alpha=0,045$, el cual si bien indicaba rechazar la hipótesis nula, la decisión tomada fue aceptarla por lo antes explicado.

En el ámbito de la riqueza léxica, si bien se contó con un corpus que en número superó las 250 noticias informativas, el análisis basado en el número de palabras distintas y en el número de palabras de contenido siempre dependerá de la extensión que presenten los textos. Bajo este precepto la generación de nuevas palabras dentro de un texto declina luego de alcanzar una curva máxima, la cual está influenciada por la extensión de los textos.

Un texto corto, con menos de 200 palabras, presentará índices de riqueza léxica mayores que un texto de mil palabras. Para intentar cubrir esta debilidad es que se pensó en metodologías de medición que logran cierto control de la variabilidad del TTR como indicador de riqueza: diversidad léxica total, que incluye el texto en toda su extensión; diversidad léxica aleatoria, el cual toma tres muestras de X palabras dentro del texto, calcula su diversidad y entrega el promedio, y finalmente la diversidad del lead, donde se analiza la diversidad léxica de la parte con mayor carga informativa de una noticia.

Otra limitante que se debe considerar al realizar una aproximación como la propuesta, es el factor de actualidad. Durante las fechas en la que fueron realizadas las pruebas de disponibilidad léxica, ocurrieron algunos hechos que pudieron influir en la configuración u ordenamiento de las redes de asociación semántica. Entre estos, hubo conflictos políticos mediáticos en el país, salieron a

la luz pública temas como el Caso Penta, luego el Caso Caval y en el plano deportivo se realizó la Copa América 2015 (fútbol) en Chile.

Se estima que el grupo de profesionales podría ser especialmente sensible a estos factores debido a que debe estar al tanto de ellos. Lo mismo debería ocurrir con los estudiantes, sobre todo con el grupo de cuarto año, el cual ya ha estado en contacto con el medio profesional de los medios de comunicación mediante asignaturas como pasantías y una prepráctica, las cuales se realizan fuera de la universidad.

5.2. PROYECCIONES DE LA INVESTIGACIÓN

Como ya se indicó, uno de los resultados adicionales de la tesis fue el desarrollo del script para medir la riqueza léxica bajo diversos indicadores. Con esta aplicación se facilitó cuantitativa y cualitativamente el análisis de los textos, tanto en cálculos matemáticos como en ahorro de tiempo en el procesamiento de las noticias. Una línea que se proyecta con este resultado no esperado de esta tesis es el perfeccionamiento de la herramienta tanto en sus funciones –las cuales podrían incluir el análisis de densidad léxica en parte de la estructura de un texto–, como en las de visualización a través del diseño de una interfaz amigable con el usuario.

Por otra parte, se considera que la aproximación a la competencia léxica de un estudiante es una herramienta que puede significar un elemento diferenciador en el desarrollo de las habilidades oral y escrita de los futuros profesionales; lo anterior no sólo en el caso de los periodistas ya que, mediante

una adaptación de temáticas y de tipo discursivo, se puede proyectar su utilización en otras disciplinas.

Al establecer la variable tipo de establecimiento educacional se pudo confirmar que éste es un elemento con un gran peso predictivo en la riqueza léxica. Sin embargo, ampliando el estudio a otras variables, por ejemplo de tipo sociocultural y de rendimiento académico previo, se podrían identificar indicadores que permitan estimar el posible nivel de la competencia léxica de los estudiantes al ingresar a la educación superior.

Otra posible línea de desarrollo es trabajar en la elaboración de un modelo en el cual se logre unir estas tres aproximaciones –riqueza léxica, disponibilidad léxica y redes de asociación semántica–, utilizando los datos aportados en estos tres estudios, a través de un ensamblaje que pondere cada uno de estos resultados y permita acceder a un análisis más completo de la competencia léxica. Una posibilidad es el desarrollo de estudios de regresión múltiple u otro tipo de metodología que aborde a estos subestudios –y posiblemente otros como índice de madurez sintáctica o sofisticación léxica– integrándolos en un solo modelo predictivo de competencia léxica.

REFERENCIAS BIBLIOGRÁFICAS

- Aitchison, J. (1994) *Words in the mind: an introduction to mental lexicon*. 2º Edición, Wiley-Blackwell.
- Arnaud, P. (1985) The lexical richness in write in production and the validity of vocabulary test. *Practice and Problems in Language Testing*. Papers from the International Symposium on Language Testing. Recuperado de <https://eric.ed.gov/?id=ED275164>
- Ávila, R. (1994). El lenguaje de la radio y la televisión: Primeras noticias. En Alegría Alonso et al. (Eds.) *II Encuentro de lingüistas y filólogos de España y México*. Ediciones Universidad de Salamanca, 1994, pp. 102-117.
- Ávila, A. y Villena, J. (Eds.) (2010). *Variación social del léxico disponible en la ciudad de Málaga*. Málaga: Editorial Sarriá.
- Baralo, M. (2005). La competencia léxica en el Marco Común europeo de referencia. *Revista Carabela*, nº58: 27-48.
- Baralo, M. (2007). *Adquisición de palabras: redes semánticas y léxicas*, en *Foro de español internacional: Aprender y enseñar léxico*, Munich, Alemania. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/munich_2006-2007/04_baralo.pdf
- Bartol, J. (2006). La Disponibilidad Léxica, *Revista Española de Lingüística (RSEL)* N°36.
- Berton, M. (2014). *La riqueza léxica en la producción escrita de estudiantes suecos de ELE*. (tesis doctoral) Universidad de Estocolmo, Suecia.
- Bulté, B. Housen, A. (2012) Defining and operationalising L2 complexity. En A. Housen, F. Kuiken y I. Vedder (eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins
- Cabré, T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártica/Em-puréis.
- Cabré, T. (1998a). El Discurso Especializado y la Variación Funcional determinada por la Temática. *Caplletra* 25: 173-194.

- Cabré, T. (1998b). Las Fuentes Terminológicas para la Traducción. En Nistal, P. y Bravo Gonzalo, J.Mª. coords. 1998: *La Traducción: Orientaciones Lingüísticas y Culturales*. Universidad de Valladolid: SAE: 27-59.
- Casado, M. (1990) Notas sobre el léxico periodístico de hoy. En *El lenguaje en los medios de comunicación*, (pp. 49-71). Asociación de la Prensa de Zaragoza. Recuperado de: <http://dadun.unav.edu/handle/10171/18474>
- Computational Linguistics & Psycholinguistics Research Center. (2017). Clips. *Stop words*. Recuperado de: <https://www.clips.uantwerpen.be/pages/stop-words>.
- Consejo de Europa. 2001. Marco Europeo Común de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación, en *RLA*, vol 31.
- Cushing Weigle, S. (2002) *Assessing Writing*. Cambridge University Press, UK. 7ma Edición.
- Daller, H., van Hout, R., and Treffers-Daller, J. (2003) Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24/2 197-222.
- Diccionario Tecnológico. (2017). Usando.info. *Stop words*. Recuperado el 20/01/2017 desde: <http://usando.info/diccionario/s#Stopwords>.
- Echeverría, M. (1982). *Hacia una evaluación de la competencia lingüística en lengua materna*. Ponencia realizada en el I Congreso internacional sobre el español de América, San Juan de Puerto Rico, entre el 4 y 8 de octubre de 1982. Recuperado en noviembre de 2014 desde: http://www.lecturayvida.fahce.unlp.edu.ar/numeros/a4n4/04_04_Echeverria.pdf
- Echeverría, M.; Herrera, Mº O.; Moreno, P.; Pradenas, F. (1987). Disponibilidad Léxica en Educación Media. *RLA. Revista de Lingüística Teórica y Aplicada*, 25, pag 25-111.
- Echeverría, M. (1993). Estructura de un perfil de competencia léxica, en *RLA. Revista de Lingüística Teórica y Aplicada*, vol.31, pag 55-66.
- Echeverría, M. (1997). Noticias y deportes en el español público de Chile en el I Congreso Internacional de la Lengua Española, Zacatecas, México. Recuperado de. <http://congresosdelalengua.es/zacatecas/plenarias/television/echeverr.htm#n1>
- Echeverría, M. (2001). Estructura y funcionales de un software de vocabulario disponible, en *RLA. Revista de Lingüística Teórica y Aplicada*, vol.39, pag 87-100.
- Echeverría, M. (2002). Programas computacionales para el español como lengua materna. *Revista signos*, 35(51-52), 163-193. <https://dx.doi.org/10.4067/S0718-09342002005100012>

- Echeverría, M.; Vargas, R.; Urzua, P.; Ferreira, R. (2008). DispoGrafo: una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 81-91. <https://dx.doi.org/10.4067/S0718-48832008000100005>
- Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139-155.
- Ferreira, R.; Echeverría, M. (2010). Redes semánticas en el léxico disponible de inglés L1 e inglés LE, en *Revista Onomázein* nº21-Vol1, 133-153.
- Ferreira, A., Salcedo, P., y Del Valle, M. (2014). Estudio de disponibilidad léxica en el ámbito de las matemáticas. *Estudios Filológicos* [online]. nov. 2014, no.54, p.69- 84. Disponible en: <http://mingaonline.uach.cl/scielo.php?script=sci_arttext&pid=S0071-17132014000200004&lng=es&nrm=iso>. ISSN 0071-1713.
- Fontcuberta, M. (1987). *La noticia*. Ed. Paidós Espasa Libros. Barcelona, España.
- Fusté-Herrmann, B; Silliman, E.; Bahr, R.; Fasnacht, K.; Federico, J. (2006). Mental State Verb Production in the Oral Narratives of English -and Spanish- Speaking Preadolescents: An Exploratory Study of Lexical Diversity and Depth. *Learning Disabilities Research & Practice*. 21. 44 - 60. 10.1111/j.1540-5826.2006.00206.
- Gómez, J. (2002). La competencia léxica en el currículo de español para fines específicos. II Congreso Internacional de Español para Fines Específicos, Ámsterdam, Holanda. Recuperado en noviembre de 2013 de http://cvc.cervantes.es/ensenanza/biblioteca_ele/ciefe/indice2.htm
- Gregory-Signes, C., Clavel-Arroitía, B. (2015) Analyzing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Science*, 198 546-556.
- Herrmann, B; Silliman, E.; Bahr, R.; Fasnacht, K.; Federico, J. (2006). Mental State Verb Production in the Oral Narratives of English -and Spanish- Speaking Preadolescents: An Exploratory Study of Lexical Diversity and Depth. *Learning Disabilities Research & Practice*. 21. 44 - 60. 10.1111/j.1540-5826.2006.00206. x.
- Herrera, H., Martínez, R. y Amengual, M. (2011). *Estadística aplicada a la investigación lingüística*. EOS Universitaria. Madrid.

- Hernández, R.; Fernández, C.; y Baptista, P. (2003). Metodología de la investigación. 3º Edición, México D.F. McGraw-Hill Interamericana.
- Hernández, N. (2004). *El léxico disponible de los estudiantes conguenses*. 1º Edición, Madrid, España: Ediciones Universidad de Salamanca.
- Hernández, N.; Izura, C.; Ellis, A. (2006). Cognitive aspects of lexical availability, *European Journal of Cognitive Psychology*. N°18(5), pág. 730-755.
- Hernández, S. (2016) *Evaluación automática de la estructura semántica de pirámide invertida en noticias escritas* (tesis doctoral) Universidad de Concepción, Concepción, Chile.
- Housen, A., Kuiken, F. y Vedder, I. (eds.) (2012) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins.
- Hymes, D., Gómez Bernal, J. (1996). Acerca de la competencia comunicativa. *Forma y Función*, 0(9), 13-37. Recuperado de <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/17051/17909>
- Jarvis, S. (2002) Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing*, 19: 1-15.
- Jarvis, S. (2013) Capturing the Diversity in Lexical Diversity. *Language Learning Research Club*. doi.org/10.1111/j.1467-9922.2012.00739
- Jiménez-Catalán, R. (2002) El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas. *Atlantis* vol XXIV (2) 149-162.
- Johansson, V. (2008) Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers* [Lund: Lund University, Department of Linguistics and Phonetics], 53, 61-79. Recuperado el 10/06/2015 desde: <http://journals.lub.lu.se/index.php/LWPL/article/view/2273/1848>
- Lahuerta, J.; Pujol, M. (1996) El lexicón mental y la enseñanza del vocabulario La enseñanza del léxico como lengua extranjera : homenaje a Anton e Inge Bemmerlein. En Segoviano, C. Madrid: Iberoamericana.
- Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *The Canadian Modern Language Review*, 59, 567-587.
- Laufer, B. & P. Nation. (1995). A vocabulary size and use. Lexical richness in L2 written production. *Applied Linguistics* 16, 307- 322.

- Laufer, B. & P. Nation. (1999). A vocabulary size test of controlled productive ability. *Language Testing* 16, 33- 51.
- Leñero, V., Marín, C. (1986). *Manual de Periodismo*, 12 Edición, México D.F. Editorial Grijalbo.
- López Chávez, J. y Strassburger, C. (1991). Un modelo más para el cálculo de disponibilidad léxica individual. En H. López Morales (ed.), *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico, p. 91-112.
- López-Mezquita, T. (2007). *La evaluación de la competencia léxica: Test de vocabulario. Su fiabilidad y validez*. Editado por CIDE, Centro de investigación y documentación educativa y Ministerio de Educación y Ciencia. España.
- López Morales, H. (1995). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología*, 35 (1), Pág. 245-259. Recuperado el 10/01/2014 desde: <http://www.nuevosfoliosbioetica.uchile.cl/index.php/BDF/article/view/19231/20354>
- López Morales, H. (2002) *Los índices de 'Riqueza léxica' y la enseñanza de lenguas*. En: Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español I2-le. Vol 1. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/5419218.pdf>.
- Martín Vivaldi, G. (1993). *Géneros Periodísticos*. 5º Edición, Madrid, España: Editorial Paraninfo.
- Martínez Albertos, J.M. (1989). *El Lenguaje periodístico. Estudios sobre el mensaje y la producción de textos*. Madrid, España: Editorial Paraninfo.
- Marconi, D. (2000) *La competencia léxica*. Madrid: Machado Libros
- Martín Vivaldi, G. (1993) *Géneros periodísticos*. Madrid: Paraninfo.
- Martínez Albertos, J.M. (1989) *El Lenguaje periodístico. Estudios sobre el mensaje y la producción de textos*. Madrid, España: Editorial Paraninfo.
- Martínez Albertos, J.M. (1991) *Curso general de redacción periodística*. Madrid, España: International Thomson Editores, Spain, Editorial Paraninfo. 5º Edición
- Malvern, D.; Richards, B.; Chipere, N.; Durán, P (2004) *Lexical Diversity and Language Development: Quantification and Assessment*. Palgrave Macmillan UK
- Matsuda, K., Sadowsky, S., y Sabaj, O. (2011) Índice de Palabras de Contenido (IPC) y Distribución Porcentual de Legomena (DPL) en artículos de investigación en español. *Revista Signos*. 45(78) 70-82.

- McCarthy, P. M. y Jarvis, S. (2007) Vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459-488.
- McCarthy, P.; Jarvis, S. (2010) MTL, Vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42 (2), 381-392.
- Müller, Ch. (1973). *Estadística lingüística*. Madrid, Editorial Gredos, S.A.
- Núñez, M., del Moral, C (2010) Competencia léxica y competencia comunicativa: bases para el diseño de programas didácticos en la educación escolar. *Lenguaje y Textos*, 32, 91-97.
- Obegi, N. 2005. La riqueza léxica como parámetro de distinción en los textos del inglés antiguo, en *Interlingüística*, Nº. 16.
- Paredes, F. (2012). Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica. *Revista Nebrija de Lingüística Aplicada* 11 (6), 78-100.
- Parodi, G. (Ed.) (2005a). *Discurso especializado e instituciones formadoras*. Valparaíso: Ediciones Universitarias de Valparaíso.
- Parodi, G. (2005b). La comprensión del discurso especializado escrito en ámbitos técnico-profesionales: ¿Aprendiendo a partir del texto? *Revista Signos*, 38(58), 221-267.
- Parodi, G. (2006). Discurso especializado y lengua escrita: Foco y variación. *Estudios Filológicos*, 41, 165-204.
- Parodi, G. (2007). El discurso especializado escrito en el ámbito universitario y profesional: Constitución de un corpus de estudio. *Revista Signos*, 40(63), 147-178. <https://dx.doi.org/10.4067/S0718-09342007000100008>.
- Pérez Salazar, C.; Olza, I. (2014). *Del discurso de los medios de comunicación a la lingüística del discurso*. Berlin: Frank & Timme.
- Pelegrín, B.; Cánovas, L.; y Fernández, P. (2009) *Algoritmos en grafos y redes*. Primera Edición, Barcelona. Promociones y Publicaciones Universitarias S.A.
- Read, J. (2011) *Assessing Vocabulary* Cambridge University Press, UK. 9a Edición.
- Real Academia Española, RAE. 2014. Diccionario de la lengua española. 23a edición. Madrid, España: Espasa.
- Reyes, J. (2010) Riqueza léxica de textos redactados por alumnos del bachillerato de las Palmas Gran Canaria. *Anuario de Lingüística Hispánica*. XXIII-XXIV, 147-163.

- Romero, M (2003) Léxico periodístico. Usos culturales en las páginas de prensa *Revista de Investigación Lingüística* nº2 vol VI, 107-124.
- Salcedo, P., del Valle, M. (2016) Disponibilidad Léxica Matemática en Estudiantes de Enseñanza Media de Concepción, Chile. *Atenas, Revista Científica Pedagógica* 4 (21) ISSN: 1682-2749.
- Salcedo, P., Ferreira, A., Barrientos, F. (2013). A Bayesian Model for Lexical Availability of Chilean High School Students in Mathematics". *Natural and Artificial Models in Computation and Biology*. 245-253.
- Sánchez-Saus Laserna, (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. Tesis Doctoral en Filología Universidad de Cádiz, España. Disponible en: <http://hdl.handle.net/10498/15862>.
- Saralegui, C., Taberner, C. (2008) "Aportación al Proyecto Panhispánico de Léxico disponible: Navarra" *Actas del XXXVII Simposio internacional de la Sociedad Española de lingüística (SEL)*, Pamplona, Universidad de Navarra. 745-761. Publicación electrónica en: <http://www.unav.es/linguis/simposiosel/actas>
- Šišková, Z. (2012) Lexical richness in EFL students' narratives. *Language Studies Working Papers*. Vol. 4 26-36.
- Torruella, J., Capsada, R. (2013) Lexical statistics and typological structures: a measure of lexical richness. *Procedia Social and Behavioral Sciences* 95 447-454
- Urrutia, M. (2003). Redes semánticas en línea: Una tarea de acceso léxico a partir de un estudio experimental. *RLA*, Vol 41, 119-141.
- Urzúa, P; Sáez, K; y Echeverría, M. (2006). Disponibilidad léxica matemática: Análisis cuantitativo y cualitativo. *RLA. Revista de lingüística teórica y aplicada*, 44(2), 59-76.
- Valencia, A., Echeverría M. (1999). Disponibilidad léxica en estudiantes chilenos. Santiago de Chile y Concepción: Universidad de Chile y Universidad de Concepción
- Vermeer, A. (2000) Coming to grips with lexical richness in spontaneous speech data. *Language Testing* 17, 65-83.
- Wimmer, G., Altmann, G. (1999). Review Article: On Vocabulary Richness, *Journal of Quantitative Linguistics*, 6, 1-9.
- Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31, 236-259.

Yu-feng, Z.; Chao, H. (2010) Research of Text Classification Model Based on Latent Semantic Analysis and Improved HS-SVM. 2nd International Workshop on Intelligent Systems and Applications. DOI: 10.1109/IWISA.2010.5473702 Recuperado de <https://ieeexplore.ieee.org/document/#>

Zuloaga, M. B. (2006). Naturaleza y función de los encapsuladores en los textos informativamente densos (la noticia periodística). *Cuadernos de Filología Italiana*, 13, 73-95. https://doi.org/10.5209/rev_CFIT.2006.v13.17827



Anexos

Anexo 01: Listado de *Stop words* utilizadas para el análisis de densidad léxica.

a	esos	estoy	habías	hubieses	otros
al	esta	estuve	habida	hubimos	para
algo	está	estuviera	habidas	hubiste	pero
algunas	estaba	estuvierais	habido	hubisteis	poco
algunos	estabais	estuviéramos	habidos	hubo	por
ante	estábamos	estuvieran	habiendo	la	porque
antes	estaban	estuvieras	habrá	las	pues
así	estabas	estuvieron	habrán	le	que
bajo	estad	estuviese	habrás	les	qué
bien	estada	estuvieseis	habré	lo	quien
cabe	estadas	estuviésemos	habréis	los	quienes
como	estado	estuviesen	habremos	luego	se
con	estados	estuvieses	habría	mas	sea
conque	estados	estuvimos	habríaís	más	seáis
contra	estáis	estuviste	habríamos	me	seamos
cual	estamos	estuvisteis	habrían	mi	sean
cuando	están	estuvo	habrías	mí	seas
de	estando	fue	hacia	mía	según
del	estar	fuera	han	mías	ser

desde	estará	fuerais	has	mío	será
durante	estarán	fuéramos	hasta	míos	serán
e	estarás	fueran	hay	mis	serás
él	estaré	fueras	haya	mucho	seré
ella	estaréis	fueron	hayáis	muchos	seréis
ellas	estaremos	fuese	hayamos	muy	seremos
ellos	estaría	fueseis	hayan	nada	sería
embargo	estaríais	fuésemos	hayas	ni	seríais
en	estaríamos	fuesen	he	no	seríamos
entre	estarían	fueses	hemos	nos	serían
era	estarías	fui	hube	nosotras	serías
erais	estás	fuimos	hubiera	nosotros	sí
éramos	estas	fuiste	hubierais	nuestra	sidó
eran	esté	fuisteis	hubiéramos	nuestras	siendo
eras	este	ha	hubieran	nuestro	sin
eres	estéis	habéis	hubieras	nuestros	sino
es	estemos	haber	hubieron	o	so
esa	estén	había	hubiese	os	sobre
esas	estés	habíais	hubieseis	otra	sois
ese	esto	habíamos	hubiésemos	otras	somos
eso	estos	habían	hubiesen	otro	son
soy	tendréis	tengo	todo	tuviésemos	unos
su	tendremos	tenía	todos	tuviesen	vosotras

sus	tendría	teníais	tras	tuvieses	vosotros
suya	tendríais	teníamos	tu	tuvimos	vuestra
suyas	tendríamos	tenían	tú	tuviste	vuestras
suyo	tendrían	tenías	tus	tuvisteis	vuestro
suyos	tendrían	tenían	tuve	tuvo	vuestros
tal	tened	tenidas	tuviera	tuya	y
también	tenéis	tenido	tuvierais	tuyas	ya
tanto	tenemos	tenidos	tuviéramos	tuyo	yo
te	tenga	teniendo	tuvieran	tuyos	el
tendrá	tengáis	ti	tuvieras	un	
tendrán	tengamos	tiene	tuvieron	una	
tendrás	tengan	tienen	tuviese	unas	
tendré	tengas	tienen	tuvieseis	uno	

Anexo 02: Pruebas de disponibilidad léxica.

Prueba A: Periodistas



UNIVERSIDAD DE CONCEPCION
FACULTAD DE HUMANIDADES Y ARTE
PROGRAMA DE DOCTORADO EN LINGÜÍSTICA

La siguiente encuesta se enmarca en el proyecto de tesis del Doctorado en Lingüística de la Universidad de Concepción. Trabajo realizado por la alumna tesista Karina Fuentes Riffo, denominado "Competencia léxica de estudiantes de periodismo en el discurso profesional especializado".

Por favor complete los siguientes datos:

1.-Nombre:

2.- Sexo:

a) Hombre _____ b) Mujer _____

3.- Edad:

4.- Lugar de trabajo:

5.- Universidad donde cursó la carrera de Periodismo:

6.- Año de egreso:

7.- Su etapa de enseñanza media fue cursada en instituciones (marque con una "X"):

a) Municipal _____

b) Particular subvencionada _____

c) Particular _____

8.- Comuna de residencia actual:



Prueba B: Estudiantes.



UNIVERSIDAD DE CONCEPCION
FACULTAD DE HUMANIDADES Y ARTE
PROGRAMA DE DOCTORADO EN LINGÜÍSTICA

La siguiente encuesta se enmarca en el proyecto de tesis del Doctorado en Lingüística de la Universidad de Concepción. Trabajo realizado por la alumna tesista Karina Fuentes Riffo, denominado "Competencia léxica de estudiantes de periodismo en el discurso profesional especializado".

Por favor complete los siguientes datos

1.-Nombre

2.- Sexo:

a) Hombre _____ b) Mujer _____

3.- Edad:

4.- Año de ingreso a la Universidad:

5.- Su etapa de enseñanza media fue cursada en instituciones (marque con una "X"):

a) Municipal _____

b) Particular subvencionada _____

c) Particular _____

6.- Comuna de procedencia:

***Cada centro de interés utilizó una hoja distinta**

Ciencia y Tecnología:

Partes del Cuerpo:

Política:

Medios de Transporte:



Deporte:

Anexo 03: Listados de resultados de riqueza y densidad léxica.

I. Corpus estudiantes de primer año.

Textos Ciencia	Total de palabras	Palabras únicas	Palabras Noci o.	Densidad léxica	Div léxica	Diversidad léxica muestra	< -- ◻
078_1a_c.txt	250	164	127	0,51	0,66	0,69	0,71
069_1a_c.txt	279	151	116	0,42	0,54	0,61	0,61
099_1a_c.txt	297	159	120	0,40	0,54	0,59	0,59
060_1a_c.txt	306	190	140	0,46	0,62	0,67	0,70
068_1a_c.txt	306	204	162	0,53	0,67	0,69	0,75
076_1a_c.txt	345	203	161	0,47	0,59	0,66	0,68
018_1a_c.txt	378	201	157	0,42	0,53	0,61	0,61
047_1a_c.txt	391	219	163	0,42	0,56	0,65	0,65
051_1a_c.txt	402	227	175	0,44	0,56	0,67	0,73
098_1a_c.txt	432	217	175	0,41	0,50	0,60	0,60
004_1a_c.txt	441	245	191	0,43	0,56	0,69	0,63
048_1a_c.txt	474	224	178	0,38	0,47	0,59	0,61
052_1a_c.txt	601	297	242	0,40	0,49	0,65	0,59
104_1a_c.txt	626	230	180	0,29	0,37	0,48	0,52
101_1a_c.txt	696	321	265	0,38	0,46	0,62	0,62
109_1a_c.txt	707	323	272	0,38	0,46	0,62	0,61
079_1a_c.txt	721	333	276	0,38	0,46	0,63	0,58
105_1a_c.txt	721	333	276	0,38	0,46	0,61	0,58
006_1a_c.txt	727	323	265	0,36	0,44	0,58	0,64
100_1a_c.txt	727	323	265	0,36	0,44	0,62	0,64
097_1a_c.txt	738	347	281	0,38	0,47	0,65	0,67
089_1a_c.txt	740	315	256	0,35	0,43	0,62	0,65
103_1a_c.txt	742	375	310	0,42	0,51	0,66	0,71
106_1a_c.txt	746	297	243	0,33	0,40	0,59	0,65
107_1a_c.txt	768	340	283	0,37	0,44	0,62	0,65
096_1a_c.txt	786	342	284	0,36	0,44	0,64	0,71
026_1a_c.txt	789	363	294	0,37	0,46	0,63	0,65
111_1a_c.txt	793	364	298	0,38	0,46	0,63	0,61
094_1a_c.txt	837	366	297	0,35	0,44	0,64	0,67
091_1a_c.txt	855	408	327	0,38	0,48	0,67	0,71
102_1a_c.txt	855	415	346	0,40	0,49	0,65	0,71

092_1a_c.txt	861	434	370	0,43	0,50	0,64	0,72
053_1a_c.txt	873	392	322	0,37	0,45	0,66	0,67
114_1a_c.txt	881	394	328	0,37	0,45	0,64	0,67
075_1a_c.txt	885	399	332	0,38	0,45	0,66	0,66
108_1a_c.txt	889	386	314	0,35	0,43	0,63	0,60
113_1a_c.txt	899	389	321	0,36	0,43	0,61	0,65
110_1a_c.txt	917	415	333	0,36	0,45	0,65	0,67
112_1a_c.txt	941	422	340	0,36	0,45	0,64	0,63
029_1a_c.txt	1026	425	356	0,35	0,41	0,60	0,59
095_1a_c.txt	1132	493	413	0,36	0,44	0,62	0,71
074_1a_c.txt	1290	504	438	0,34	0,39	0,61	0,68
028_1a_c.txt	1363	589	514	0,38	0,43	0,63	0,59
031_1a_c.txt	1442	540	471	0,33	0,37	0,62	0,64
033_1a_c.txt	1511	607	524	0,35	0,40	0,63	0,67
032_1a_c.txt	1545	634	554	0,36	0,41	0,63	0,61
030_1a_c.txt	1546	600	508	0,33	0,39	0,63	0,67
050_1a_c.txt	1590	657	580	0,36	0,41	0,63	0,58
035_1a_c.txt	1667	593	519	0,31	0,36	0,62	0,65
034_1a_c.txt	2369	911	818	0,35	0,38	0,65	0,64
Texto Política	Total de palabras	Palabras únicas	Palabras conocionales	Densidad léxica	Diversidad léxica	Diversidad léxica muestra	Diversidad léxica lead
001_1a_p.txt	311	167	134	0,43	0,54	0,60	0,60
005_1a_p.txt	295	165	124	0,42	0,56	0,61	0,67
008_1a_p.txt	431	204	153	0,35	0,47	0,57	0,55
009_1a_p.txt	312	167	134	0,43	0,54	0,60	0,60
010_1a_p.txt	314	185	141	0,45	0,59	0,66	0,62
011_1a_p.txt	457	246	195	0,43	0,54	0,63	0,69
012_1a_p.txt	940	431	355	0,38	0,46	0,65	0,56
013_1a_p.txt	469	240	185	0,39	0,51	0,64	0,65
014_1a_p.txt	416	176	145	0,35	0,42	0,55	0,53
015_1a_p.txt	362	189	143	0,40	0,52	0,62	0,62
022_1a_p.txt	429	213	172	0,40	0,50	0,60	0,63
023_1a_p.txt	340	187	149	0,44	0,55	0,63	0,57
024_1a_p.txt	325	177	143	0,44	0,54	0,61	0,67
028_1a_p.txt	290	165	133	0,46	0,57	0,61	0,61
025_1a_p.txt	445	225	172	0,39	0,51	0,61	0,63
038_1a_p.txt	344	178	136	0,40	0,52	0,60	0,58

042_1a_p.txt	417	219	164	0,39	0,53	0,64	0,62
045_1a_p.txt	347	183	150	0,43	0,53	0,60	0,58
048_1a_p.txt	418	200	155	0,37	0,48	0,59	0,57
051_1a_p.txt	339	179	139	0,41	0,53	0,58	0,61
056_1a_p.txt	424	203	157	0,37	0,48	0,59	0,59
057_1a_p.txt	485	230	175	0,36	0,47	0,60	0,65
058_1a_p.txt	464	230	181	0,39	0,50	0,61	0,58
059_1a_p.txt	485	251	195	0,40	0,52	0,62	0,63
063_1a_p.txt	276	167	134	0,49	0,61	0,62	0,69
064_1a_p.txt	441	223	164	0,37	0,51	0,63	0,57
066_1a_p.txt	491	250	189	0,38	0,51	0,64	0,57
072_1a_p.txt	449	234	175	0,39	0,52	0,60	0,64
081_1a_p.txt	350	194	141	0,40	0,55	0,63	0,65
082_1a_p.txt	428	230	178	0,42	0,54	0,64	0,68
083_1a_p.txt	406	211	160	0,39	0,52	0,61	0,62
084_1a_p.txt	303	155	120	0,40	0,51	0,58	0,56
004_1a_p.txt	884	430	363	0,41	0,49	0,66	0,69
007_1a_p.txt	580	293	232	0,40	0,51	0,64	0,65
027_1a_p.txt	701	344	286	0,41	0,49	0,63	0,67
036_1a_p.txt	861	397	318	0,37	0,46	0,63	0,57
037_1a_p.txt	630	316	246	0,39	0,50	0,61	0,61
039_1a_p.txt	750	344	281	0,37	0,46	0,63	0,62
075_1a_p.txt	833	411	339	0,41	0,49	0,68	0,61
080_1a_p.txt	801	338	270	0,34	0,42	0,62	0,55
085_1a_p.txt	673	345	274	0,41	0,51	0,64	0,63
086_1a_p.txt	788	369	295	0,37	0,47	0,62	0,64
087_1a_p.txt	871	404	343	0,39	0,46	0,65	0,61
088_1a_p.txt	808	378	319	0,39	0,47	0,68	0,65
093_1a_p.txt	906	410	349	0,39	0,45	0,65	0,57
Texto Deporte	Total de palabras	Palabras únicas	Palabras noci onales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
049_1a_d.txt	415	234	173	0,42	0,56	0,64	0,82
054_1a_d.txt	390	186	130	0,33	0,48	0,56	0,64
059_1a_d.txt	383	213	167	0,43	0,55	0,59	0,77
060_1a_d.txt	337	186	163	0,47	0,55	0,59	0,62
065_1a_d.txt	325	191	148	0,45	0,59	0,60	0,76
070_1a_d.txt	405	219	174	0,43	0,54	0,62	0,66

115_1a_d.txt	300	161	108	0,36	0,54	0,56	0,66
003_1a_d.txt	502	269	214	0,43	0,54	0,65	0,62
010_1a_d.txt	531	280	222	0,42	0,53	0,65	0,69
016_1a_d.txt	699	335	288	0,41	0,48	0,66	0,61
017_1a_d.txt	696	295	245	0,35	0,42	0,62	0,65
018_1a_d.txt	620	304	250	0,40	0,49	0,65	0,63
020_1a_d.txt	710	278	235	0,33	0,39	0,55	0,57
021_1a_d.txt	817	351	304	0,37	0,43	0,63	0,69
040_1a_d.txt	755	379	314	0,42	0,50	0,68	0,63
041_1a_d.txt	972	440	366	0,38	0,45	0,65	0,59
043_1a_d.txt	723	371	308	0,43	0,51	0,65	0,70
044_1a_d.txt	757	375	308	0,41	0,50	0,66	0,65
046_1a_d.txt	903	427	364	0,40	0,47	0,64	0,65
067_1a_d.txt	501	271	225	0,45	0,54	0,66	0,68
071_1a_d.txt	513	248	198	0,39	0,48	0,61	0,66
077_1a_d.txt	532	261	221	0,42	0,49	0,62	0,67
116_1a_d.txt	546	290	234	0,43	0,53	0,65	0,78
117_1a_d.txt	724	310	245	0,34	0,43	0,62	0,63
118_1a_d.txt	724	310	245	0,34	0,43	0,59	0,63

II. Corpus estudiantes de cuarto año

Texto Ciencia	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
005_5a_c.txt	439	253	199	0,45	0,58	0,63	0,71
009_5a_c.txt	406	227	172	0,42	0,56	0,66	0,72
013_5a_c.txt	453	254	199	0,44	0,56	0,66	0,67
014_5a_c.txt	1140	548	452	0,40	0,48	0,69	0,69
015_5a_c.txt	399	240	187	0,47	0,60	0,71	0,73
017_5a_c.txt	307	173	138	0,45	0,56	0,62	0,65
020_5a_c.txt	330	212	175	0,53	0,64	0,71	0,71
021_5a_c.txt	534	298	249	0,47	0,56	0,70	0,71
033_5a_c.txt	647	317	261	0,40	0,49	0,61	0,64
037_5a_c.txt	390	204	161	0,41	0,52	0,60	0,67
038_5a_c.txt	728	338	277	0,38	0,46	0,64	0,66

042_5a_c.txt	396	232	185	0,47	0,59	0,66	0,66
Texto Política	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
004_5a_p.txt	382	230	186	0,49	0,60	0,67	0,73
008_5a_p.txt	289	173	139	0,48	0,60	0,66	0,70
010_5a_p.txt	406	236	186	0,46	0,58	0,69	0,71
012_5a_p.txt	474	243	182	0,38	0,51	0,60	0,69
018_5a_p.txt	269	164	122	0,45	0,61	0,66	0,69
022_5a_p.txt	274	160	115	0,42	0,58	0,62	0,73
023_5a_p.txt	257	161	128	0,50	0,63	0,66	0,69
025_5a_p.txt	270	157	120	0,44	0,58	0,62	0,71
026_5a_p.txt	382	213	165	0,43	0,56	0,65	0,67
030_5a_p.txt	350	228	176	0,50	0,65	0,73	0,75
031_5a_p.txt	254	137	106	0,42	0,54	0,58	0,68
039_5a_p.txt	329	203	158	0,48	0,62	0,67	0,68
040_5a_p.txt	464	225	179	0,39	0,48	0,63	0,65
041_5a_p.txt	365	215	175	0,48	0,59	0,68	0,71
043_5a_p.txt	463	244	201	0,43	0,53	0,65	0,65
044_5a_p.txt	410	245	192	0,47	0,60	0,66	0,72
Texto Deporte	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
001_5a_d.txt	372	213	165	0,44	0,57	0,66	0,71
002_5a_d.txt	329	220	176	0,53	0,67	0,73	0,76
003_5a_d.txt	1224	613	534	0,44	0,50	0,69	0,73
007_5a_d.txt	392	241	187	0,48	0,61	0,70	0,71
011_5a_d.txt	345	216	175	0,51	0,63	0,70	0,68
016_5a_d.txt	1202	574	486	0,40	0,48	0,65	0,71
019_5a_d.txt	949	471	395	0,42	0,50	0,69	0,74
024_5a_d.txt	1807	741	639	0,35	0,41	0,66	0,71
027_5a_d.txt	405	237	191	0,47	0,59	0,65	0,73
028_5a_d.txt	810	373	308	0,38	0,46	0,69	0,68
029_5a_d.txt	302	190	141	0,47	0,63	0,67	0,66
036_5a_d.txt	514	259	199	0,39	0,50	0,62	0,65

II. Corpus noticias profesionales

Texto Ciencia	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
028_p_c.txt	303	156	121	0,40	0,51	0,56	0,57
026_p_c.txt	304	139	108	0,36	0,46	0,54	0,66
002_p_c.txt	313	168	136	0,43	0,54	0,58	0,67
001_p_c.txt	316	156	123	0,39	0,49	0,57	0,59
009_p_c.txt	316	160	119	0,38	0,51	0,58	0,55
005_p_c.txt	327	191	147	0,45	0,58	0,67	0,61
006_p_c.txt	352	201	163	0,46	0,57	0,64	0,62
017_p_c.txt	361	178	141	0,39	0,49	0,58	0,56
011_p_c.txt	362	224	180	0,50	0,62	0,68	0,73
013_p_c.txt	368	220	171	0,46	0,60	0,66	0,70
003_p_c.txt	386	210	165	0,43	0,54	0,65	0,64
010_p_c.txt	412	250	195	0,47	0,61	0,68	0,69
030_p_c.txt	425	200	159	0,37	0,47	0,58	0,51
027_p_c.txt	451	236	183	0,41	0,52	0,63	0,65
025_p_c.txt	462	271	221	0,48	0,59	0,69	0,69
004_p_c.txt	474	243	201	0,42	0,51	0,61	0,59
022_p_c.txt	484	235	190	0,39	0,49	0,59	0,65
024_p_c.txt	489	239	193	0,39	0,49	0,63	0,62
012_p_c.txt	514	283	225	0,44	0,55	0,69	0,71
016_p_c.txt	520	253	210	0,40	0,49	0,64	0,65
019_p_c.txt	524	254	206	0,39	0,48	0,62	0,53
018_p_c.txt	540	295	241	0,45	0,55	0,66	0,64
020_p_c.txt	557	290	237	0,43	0,52	0,69	0,65
008_p_c.txt	574	284	231	0,40	0,49	0,62	0,61
023_p_c.txt	578	284	227	0,39	0,49	0,64	0,61

014_p_c.txt	580	297	241	0,42	0,51	0,66	0,67
029_p_c.txt	597	261	210	0,35	0,44	0,56	0,52
015_p_c.txt	636	327	261	0,41	0,51	0,64	0,66
021_p_c.txt	701	336	270	0,39	0,48	0,64	0,63
007_p_c.txt	716	356	288	0,40	0,50	0,65	0,72
Texto Política	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
004_p_p.txt	257	121	83	0,32	0,47	0,54	0,53
029_p_p.txt	270	134	101	0,37	0,50	0,55	0,55
026_p_p.txt	287	162	128	0,45	0,56	0,61	0,63
025_p_p.txt	290	140	106	0,37	0,48	0,56	0,54
005_p_p.txt	293	175	139	0,47	0,60	0,65	0,64
022_p_p.txt	309	168	122	0,39	0,54	0,62	0,61
015_p_p.txt	311	141	98	0,32	0,45	0,54	0,50
007_p_p.txt	320	157	120	0,38	0,49	0,54	0,48
016_p_p.txt	331	167	126	0,38	0,50	0,58	0,59
027_p_p.txt	352	181	133	0,38	0,51	0,60	0,61
030_p_p.txt	352	181	133	0,38	0,51	0,60	0,61
014_p_p.txt	386	202	165	0,43	0,52	0,62	0,63
020_p_p.txt	387	199	150	0,39	0,51	0,62	0,65
002_p_p.txt	416	204	165	0,40	0,49	0,60	0,55
008_p_p.txt	419	211	159	0,38	0,50	0,63	0,64
010_p_p.txt	419	211	159	0,38	0,50	0,63	0,64
013_p_p.txt	473	229	177	0,37	0,48	0,58	0,57
017_p_p.txt	474	228	169	0,36	0,48	0,63	0,59
023_p_p.txt	488	224	172	0,35	0,46	0,60	0,60
024_p_p.txt	496	239	192	0,39	0,48	0,62	0,57

003_p_p.txt	511	248	188	0,37	0,49	0,60	0,53
021_p_p.txt	518	255	195	0,38	0,49	0,62	0,62
031_p_p.txt	527	273	217	0,41	0,52	0,64	0,61
032_p_p.txt	527	273	217	0,41	0,52	0,67	0,61
009_p_p.txt	528	267	203	0,38	0,51	0,64	0,58
001_p_p.txt	543	294	248	0,46	0,54	0,65	0,69
011_p_p.txt	564	247	178	0,32	0,44	0,61	0,61
018_p_p.txt	604	278	223	0,37	0,46	0,63	0,59
006_p_p.txt	628	316	265	0,42	0,50	0,63	0,61
012_p_p.txt	801	304	253	0,32	0,38	0,59	0,57
019_p_p.txt	828	378	314	0,38	0,46	0,69	0,62
028_p_p.txt	1766	506	432	0,24	0,29	0,58	0,67
Texto Deporte	Total de palabras	Palabras únicas	Palabras nocionales	Densidad léxica	Diversidad léxica	diversidad léxica muestra	diversidad léxica lead
023p_d.txt	268	134	106	0,40	0,50	0,56	0,55
032p_d.txt	271	165	119	0,44	0,61	0,66	0,69
005p_d.txt	273	155	114	0,42	0,57	0,62	0,61
002p_d.txt	275	149	115	0,42	0,54	0,60	0,57
016p_d.txt	279	151	114	0,41	0,54	0,60	0,63
014p_d.txt	280	155	127	0,45	0,55	0,59	0,63
022p_d.txt	292	161	116	0,40	0,55	0,62	0,56
020p_d.txt	293	161	117	0,40	0,55	0,60	0,58
007p_d.txt	296	159	122	0,41	0,54	0,59	0,64
035p_d.txt	299	166	128	0,43	0,56	0,62	0,62
037p_d.txt	301	171	138	0,46	0,57	0,62	0,60
001p_d.txt	305	180	142	0,47	0,59	0,65	0,64
024p_d.txt	327	166	139	0,43	0,51	0,59	0,53

010p_d.txt	340	207	154	0,45	0,61	0,69	0,68
006p_d.txt	354	191	144	0,41	0,54	0,64	0,60
025p_d.txt	361	180	156	0,43	0,50	0,60	0,61
026p_d.txt	366	166	138	0,38	0,45	0,57	0,60
004p_d.txt	393	207	156	0,40	0,53	0,63	0,61
011p_d.txt	397	217	169	0,43	0,55	0,63	0,62
015p_d.txt	397	214	168	0,42	0,54	0,65	0,63
021p_d.txt	398	218	165	0,41	0,55	0,66	0,61
009p_d.txt	405	226	181	0,45	0,56	0,66	0,65
028p_d.txt	417	228	188	0,45	0,55	0,64	0,61
030p_d.txt	464	248	208	0,45	0,53	0,68	0,61
029p_d.txt	484	249	217	0,45	0,51	0,63	0,55
008p_d.txt	491	260	203	0,41	0,53	0,65	0,57
013p_d.txt	492	273	220	0,45	0,55	0,70	0,62
012p_d.txt	505	284	232	0,46	0,56	0,69	0,65
031p_d.txt	549	286	232	0,42	0,52	0,67	0,65
033p_d.txt	668	336	285	0,43	0,50	0,67	0,74
003p_d.txt	708	320	270	0,38	0,45	0,66	0,65
017p_d.txt	720	323	262	0,36	0,45	0,61	0,60
027p_d.txt	736	340	297	0,40	0,46	0,68	0,67
018p_d.txt	742	349	285	0,38	0,47	0,66	0,66
034p_d.txt	915	400	322	0,35	0,44	0,65	0,70

Anexo 04: Índices de Disponibilidad Léxica

1. Ciencia y Tecnología

Vocablo	IDL	Frecuencia	% Aparición
computador	0,294	34	34,34
internet	0,263	33	33,33
avance	0,257	33	33,33
investigación	0,208	27	27,27
biología	0,188	24	24,24
química	0,174	22	22,22
celular	0,171	22	22,22
innovación	0,171	21	21,21
física	0,142	20	20,20
descubrimiento	0,133	18	18,18
laboratorio	0,127	19	19,19
computación	0,110	13	13,13
científico	0,108	16	16,16
desarrollo	0,099	15	15,15
progreso	0,090	13	13,13
experimento	0,086	14	14,14
medicina	0,082	13	13,13
teléfono	0,074	9	9,09
ciencia-social	0,071	10	10,10
conocimiento	0,064	9	9,09
Tablet	0,063	11	11,11
Apple	0,061	10	10,10
estudio	0,060	8	8,08
robótica	0,048	7	7,07
dron	0,047	6	6,06
futuro	0,044	6	6,06
salud	0,043	6	6,06
televisión	0,042	6	6,06

máquina	0,041	7	7,07
hipertexto	0,041	5	5,05
radio	0,041	6	6,06
red-social	0,040	5	5,05
herramienta	0,040	5	5,05
smartphone	0,040	7	7,07
cambio	0,039	5	5,05
aplicación	0,039	7	7,07
ensayo	0,038	5	5,05
robot	0,038	6	6,06
proyecto	0,037	6	6,06
ipad	0,037	5	5,05
cámara	0,035	6	6,06
investigador	0,035	7	7,07
automóvil	0,035	5	5,05
matemática	0,034	5	5,05
información	0,034	5	5,05
consola	0,033	4	4,04
einstein	0,033	5	5,05
informática	0,032	4	4,04
televisor	0,032	4	4,04
evolución	0,031	4	4,04
iphone	0,030	5	5,05
software	0,029	4	4,04
modernidad	0,028	3	3,03
universidad	0,028	6	6,06
sistema	0,027	4	4,04
biotecnología	0,026	3	3,03
disciplina	0,026	3	3,03
hallazgo	0,026	4	4,04
técnica	0,026	5	5,05
naturaleza	0,025	4	4,04
cultura	0,024	3	3,03

teléfono-inteligente	0,024	3	3,03
medio-de-comunicación	0,024	3	3,03
web	0,024	3	3,03
método-científico	0,024	3	3,03
invento	0,024	3	3,03
notebook	0,024	4	4,04
análisis	0,023	4	4,04
www	0,022	3	3,03
hardware	0,022	3	3,03
invención	0,022	3	3,03
patente	0,022	4	4,04
mp4	0,022	4	4,04
comicito	0,022	4	4,04
átomo	0,022	4	4,04
inteligencia	0,021	3	3,03
creación	0,021	3	3,03
docencia	0,020	3	3,03
microsoft	0,020	3	3,03
steve-jobs	0,020	3	3,03
electricidad	0,020	4	4,04
ipod	0,019	3	3,03
experimentación	0,019	3	3,03
ascensor	0,019	3	3,03
mejora	0,019	3	3,03
educación	0,019	3	3,03
epistemología	0,018	3	3,03
fotosíntesis	0,018	3	3,03
moderno	0,018	2	2,02
energía	0,018	3	3,03
instrumento	0,018	5	5,05
bill-gates	0,017	3	3,03
MP3	0,017	3	3,03
rayo-x	0,017	2	2,02

transformación	0,017	2	2,02
ingeniería	0,017	3	3,03
videojuego	0,017	2	2,02
Facebook	0,017	3	3,03
ciencia-natural	0,016	2	2,02
comodidad	0,016	3	3,03
universo	0,016	3	3,03
reestructuración	0,016	2	2,02
adelanto	0,016	4	4,04
método	0,016	2	2,02
capacidad	0,016	2	2,02
creatividad	0,016	2	2,02
astronomía	0,015	3	3,03
microscopio	0,015	4	4,04
renovación	0,015	2	2,02
teoría	0,015	3	3,03
telescopio	0,015	4	4,04
necesidad	0,015	3	3,03
luz	0,014	3	3,03
exploración	0,014	2	2,02
conexión	0,014	2	2,02
maquinaria	0,014	2	2,02
tesis	0,014	2	2,02
genoma	0,014	2	2,02
humano	0,014	2	2,02
estudiante	0,014	2	2,02
electrónica	0,014	3	3,03
china	0,013	2	2,02
Samsung	0,013	2	2,02
fórmulas	0,013	2	2,02
radiografía	0,013	2	2,02
calidad-de-vida	0,013	2	2,02
papel	0,013	3	3,03

red	0,012	2	2,02
clonación	0,012	2	2,02
arpanet	0,012	2	2,02
crecimiento	0,012	2	2,02
ecografía	0,012	2	2,02
estilo-de-vida	0,012	2	2,02
artefacto	0,011	2	2,02
conectividad	0,011	2	2,02
industria	0,011	2	2,02
comportamiento	0,011	2	2,02
mecánica	0,011	2	2,02
metodología	0,011	2	2,02
profesor	0,011	2	2,02
cura	0,010	2	2,02
vida	0,010	2	2,02
cable	0,010	2	2,02
solución	0,010	3	3,03
cuerpo	0,010	2	2,02
astrofísica	0,010	1	1,01
xbox	0,010	1	1,01
ferrocarril	0,010	1	1,01
marie-curie	0,010	1	1,01
medicina-avanzada	0,010	1	1,01
ficción	0,010	1	1,01
museo	0,010	1	1,01
observador	0,010	1	1,01
pretensión	0,010	1	1,01
mecnica-cu-ntica	0,010	1	1,01
stephen-hawkins	0,010	2	2,02
power-point	0,010	2	2,02
espíritu	0,009	2	2,02
eppendorf	0,009	1	1,01
crimen	0,009	1	1,01

academia	0,009	1	1,01
playstation	0,009	1	1,01
fotografía	0,009	1	1,01
radioactividad	0,009	1	1,01
biología	0,009	1	1,01
nuevas-formas-de-comunicación	0,009	1	1,01
utopía	0,009	1	1,01
estudio-cualitativo	0,009	1	1,01
usb	0,009	1	1,01
instagram	0,009	1	1,01
telefonía	0,009	1	1,01
smartphones	0,009	1	1,01
matriz	0,009	1	1,01
transferencia	0,009	1	1,01
relatividad	0,009	1	1,01
newton	0,009	2	2,02
android	0,009	2	2,02
ingeniero	0,009	2	2,02
recurso	0,009	2	2,02
mente	0,009	2	2,02
entropía	0,009	1	1,01
capitalismo	0,009	1	1,01
universidad-de-concepción	0,009	1	1,01
wii	0,009	1	1,01
cibercafé	0,009	1	1,01
cáncer	0,009	1	1,01
aburrimiento	0,009	1	1,01
matemática	0,009	1	1,01
estudio-cuantitativo	0,009	1	1,01
electrodoméstico	0,009	1	1,01
facilidad	0,009	1	1,01
astronauta	0,009	1	1,01
pendrive	0,009	1	1,01

nokia	0,009	1	1,01
estudios	0,009	1	1,01
teflón	0,009	1	1,01
observación	0,009	1	1,01
principio	0,009	1	1,01
darwin	0,009	1	1,01
célula	0,009	2	2,02
word	0,008	2	2,02
era-digital	0,008	2	2,02
nave-espacial	0,008	1	1,01
tics	0,008	1	1,01
digital	0,008	1	1,01
índex	0,008	1	1,01
guerra-mundial	0,008	1	1,01
láser	0,008	1	1,01
ciencia-química	0,008	1	1,01
sabiduría	0,008	1	1,01
gadgets	0,008	1	1,01
número	0,008	1	1,01
disco-duro	0,008	1	1,01
teléfono	0,008	1	1,01
comunicación	0,008	1	1,01
tics	0,008	1	1,01
transferencia-tecnológica	0,008	1	1,01
equipamiento	0,008	1	1,01
periodismo	0,008	1	1,01
windows	0,008	2	2,02
electrón	0,007	1	1,01
gente	0,007	1	1,01
aparato	0,007	1	1,01
élite	0,007	1	1,01
rapidez	0,007	1	1,01
cd	0,007	1	1,01

ciencia-matemática	0,007	1	1,01
acelerador-de-partículas	0,007	1	1,01
cibernética	0,007	1	1,01
historia	0,007	1	1,01
pantalla-led	0,007	1	1,01
botón	0,007	1	1,01
lenguaje	0,007	1	1,01
http	0,007	1	1,01
aprendizaje	0,007	1	1,01
cotidianidad	0,007	1	1,01
email	0,007	1	1,01
joven	0,007	1	1,01
postdoctorado	0,007	1	1,01
divulgación	0,007	1	1,01
descubrimiento	0,007	1	1,01
saber	0,007	2	2,02
gas-noble	0,007	1	1,01
dominio	0,007	1	1,01
doctrina	0,007	1	1,01
manojos	0,007	1	1,01
natural	0,007	1	1,01
empírico	0,007	1	1,01
html	0,007	1	1,01
mundo	0,007	1	1,01
actividades	0,007	1	1,01
correo	0,007	1	1,01
impresora	0,007	1	1,01
mesa-óptica	0,007	1	1,01
estructura	0,007	1	1,01
termodinámica	0,007	1	1,01
empresa	0,007	2	2,02
twitter	0,007	2	2,02
crear	0,006	2	2,02

google	0,006	2	2,02
sociedad	0,006	1	1,01
enfermedad	0,006	1	1,01
deshumanización	0,006	1	1,01
manuales	0,006	1	1,01
steve-wormak	0,006	1	1,01
bioquímica	0,006	1	1,01
protón	0,006	1	1,01
microondas	0,006	1	1,01
tiempo	0,006	1	1,01
edad-moderna	0,006	1	1,01
bioinformática	0,006	1	1,01
facilidad-de-uso	0,006	1	1,01
motorola	0,006	1	1,01
nueva-era	0,006	1	1,01
calculadora	0,006	1	1,01
precepto	0,006	1	1,01
chilenosaurio	0,006	1	1,01
tv	0,006	1	1,01
encuesta	0,006	1	1,01
ios	0,006	1	1,01
cualitativo	0,006	1	1,01
portabilidad	0,006	1	1,01
video	0,006	1	1,01
movilidad	0,006	1	1,01
desinterés-social	0,006	1	1,01
lavadora	0,006	1	1,01
fábrica	0,006	1	1,01
persona	0,006	1	1,01
holograma	0,006	1	1,01
intranet	0,006	1	1,01
minucioso	0,006	1	1,01
droga	0,006	1	1,01

método-científico	0,006	1	1,01
oportunidad	0,006	1	1,01
timelapse	0,006	1	1,01
aeronáutica	0,006	1	1,01
astrobiología	0,006	1	1,01
plasma	0,006	1	1,01
error	0,006	1	1,01
geofísica	0,006	1	1,01
valoración	0,006	1	1,01
probeta	0,006	1	1,01
cuantitativo	0,006	1	1,01
divergencia	0,006	1	1,01
transistor	0,006	1	1,01
postgrado	0,006	2	2,02
plataforma	0,005	1	1,01
célula- madre	0,005	1	1,01
problema	0,005	1	1,01
imax	0,005	1	1,01
alemania	0,005	1	1,01
científica	0,005	1	1,01
clase	0,005	1	1,01
animal	0,005	1	1,01
acuerdo	0,005	1	1,01
3g	0,005	1	1,01
aporte	0,005	1	1,01
nanotecnología	0,005	1	1,01
timer	0,005	1	1,01
microbiología	0,005	1	1,01
tv-cable	0,005	1	1,01
mutación	0,005	1	1,01
pasteur	0,005	1	1,01
materias	0,005	1	1,01
sujeto	0,005	1	1,01

publicaciones científicas	0,005	1	1,01
mula	0,005	1	1,01
fluido	0,005	1	1,01
ciencia	0,005	1	1,01
trasplante	0,005	1	1,01
tubo-de-ensayo	0,005	1	1,01
automóvil-eléctrico	0,005	1	1,01
novedad	0,005	1	1,01
filosofía	0,005	1	1,01
ayuda	0,005	1	1,01
androide	0,005	1	1,01
ciencia-del-mar	0,005	1	1,01
planta	0,005	1	1,01
medición	0,005	1	1,01
traducción	0,005	1	1,01
país	0,005	1	1,01
médico	0,005	1	1,01
clorofila	0,005	1	1,01
señal-de-comunicación-satelital	0,005	1	1,01
movimiento	0,005	1	1,01
extenso	0,005	1	1,01
misterio	0,005	1	1,01
explora	0,005	1	1,01
open-office	0,005	1	1,01
ciencia-de-la-tierra	0,005	1	1,01
remedio	0,005	1	1,01
urgencia	0,005	1	1,01
académico	0,005	1	1,01
inventor	0,005	1	1,01
estadística	0,005	1	1,01
cosmos	0,004	1	1,01
economía	0,004	1	1,01
data-show	0,004	1	1,01

mouse	0,004	1	1,01
programa	0,004	1	1,01
bioingeniería	0,004	1	1,01
químico	0,004	1	1,01
asistente	0,004	1	1,01
correlación	0,004	1	1,01
tratamiento	0,004	1	1,01
eficiencia	0,004	2	2,02
dinero	0,004	1	1,01
competencia	0,004	1	1,01
equipo	0,004	1	1,01
macintosh	0,004	1	1,01
medio	0,004	1	1,01
cosmovisión	0,004	1	1,01
ciencia-dura	0,004	1	1,01
like	0,004	1	1,01
financiamiento	0,004	1	1,01
fibra-óptica	0,004	1	1,01
prueba	0,004	1	1,01
fondo	0,004	1	1,01
porcentaje	0,004	1	1,01
política	0,004	1	1,01
curación	0,004	1	1,01
netbook	0,004	1	1,01
fósforo	0,004	1	1,01
parque-científico	0,004	1	1,01
entorno	0,004	1	1,01
inicia	0,004	1	1,01
programación	0,004	1	1,01
workshop	0,004	1	1,01
discrepancia	0,004	1	1,01
agujero-negro	0,003	1	1,01
globo	0,003	1	1,01

partícula	0,003	1	1,01
instrucción	0,003	1	1,01
cine	0,003	1	1,01
seside	0,003	1	1,01
incógnita	0,003	1	1,01
Optica	0,003	1	1,01
prototipo	0,003	1	1,01
cableado	0,003	1	1,01
evidencia	0,003	1	1,01
gas	0,003	1	1,01
desigualdad	0,003	1	1,01
puerto	0,003	1	1,01
lógica	0,003	1	1,01
audífono	0,003	1	1,01
3D	0,003	1	1,01
cálculo	0,003	1	1,01
cerc	0,003	1	1,01
dispositivo-móvil	0,003	1	1,01
energía-solar	0,003	1	1,01
escrito	0,003	1	1,01
servidor	0,003	1	1,01
matraz	0,003	1	1,01
vaso-precipitado	0,003	1	1,01
petróleo	0,003	1	1,01
interfaz	0,003	1	1,01
sala-xd	0,003	1	1,01
positivismo	0,003	1	1,01
aroba	0,003	1	1,01
app	0,003	1	1,01
energía-eólica	0,003	1	1,01
data	0,003	1	1,01
documento	0,003	1	1,01
propuesta	0,003	1	1,01

pipeta	0,003	1	1,01
rana	0,003	1	1,01
gasolina	0,003	1	1,01
cicat	0,003	1	1,01
energía-hidráulica	0,003	1	1,01
georreferencia	0,003	1	1,01
revista-científica	0,003	1	1,01
modelo	0,003	1	1,01
pantalla	0,003	1	1,01
pc	0,003	1	1,01
congreso	0,003	1	1,01
grabadora	0,003	1	1,01
silicon-valley	0,003	1	1,01
bacteria	0,003	1	1,01
idea	0,003	1	1,01
led	0,003	1	1,01
probado	0,003	1	1,01
vacuna	0,003	1	1,01
automóvil-solar	0,003	1	1,01
charlas-ted	0,002	1	1,01
coloquio	0,002	1	1,01
medicamento	0,002	1	1,01
casa-solar	0,002	1	1,01
medioambiente	0,002	1	1,01
youtube	0,002	1	1,01
ciencia-ficción	0,002	1	1,01
laser	0,002	1	1,01
anal	0,002	1	1,01
sustentabilidad	0,002	1	1,01
geología	0,002	1	1,01
doctorado	0,002	1	1,01
velocidad-de-conexión	0,002	1	1,01
conferencia	0,002	1	1,01

metano	0,002	1	1,01
centro	0,002	1	1,01
whatsapp	0,002	1	1,01
productividad	0,002	1	1,01
teclado	0,002	1	1,01
tabla-periódica	0,002	1	1,01
museología	0,002	1	1,01
pull-to-refresh	0,002	1	1,01
ácido	0,002	1	1,01
arqueología	0,002	1	1,01
funcionalidad	0,002	1	1,01
vida-natural	0,002	1	1,01
energética	0,002	1	1,01
excavación	0,001	1	1,01
práctica	0,001	1	1,01
espacio	0,001	1	1,01
resultado	0,001	1	1,01
luna	0,001	1	1,01
indicador	0,001	1	1,01
planeta	0,001	1	1,01

2. Política

Vocablo	IDL	Frecuencia	% Aparición
corrupción	0,27	35	35,35
senador	0,25	38	38,38
diputado	0,24	35	35,35
gobierno	0,23	32	32,32
presidente	0,22	31	31,31
partido	0,16	21	21,21
democracia	0,15	23	23,23

congreso	0,15	22	22,22
alcalde	0,14	24	24,24
ley	0,14	22	22,22
estado	0,14	22	22,22
izquierda	0,14	20	20,20
ministro	0,13	20	20,20
bachelet	0,12	17	17,17
derecha	0,12	18	18,18
poder	0,12	16	16,16
dinero	0,11	18	18,18
udi	0,11	18	18,18
penta	0,11	14	14,14
piñera	0,10	14	14,14
sociedad	0,09	13	13,13
comunismo	0,09	13	13,13
político	0,09	11	11,11
robo	0,08	10	10,10
elección	0,08	15	15,15
dávalos	0,08	9	9,09
partido-político	0,07	13	13,13
mentira	0,07	9	9,09
cámara	0,07	11	11,11
presidenta	0,07	9	9,09
parlamento	0,06	10	10,10
organización	0,06	10	10,10

senado	0,06	11	11,11
constitución	0,06	13	13,13
concejal	0,06	9	9,09
economía	0,06	9	9,09
dictadura	0,05	10	10,10
estafa	0,05	8	8,08
consejero	0,05	8	8,08
desigualdad	0,04	8	8,08
marxismo	0,04	6	6,06
ideología	0,04	7	7,07
votación	0,04	6	6,06
derecho	0,04	8	8,08
rn	0,04	8	8,08
ejecutivo	0,04	5	5,05
pinochet	0,04	5	5,05
voto	0,04	7	7,07
dc	0,04	6	6,06
cohecho	0,04	5	5,05
pueblo	0,03	5	5,05
intendente	0,03	8	8,08
municipalidad	0,03	4	4,04
gobernador	0,03	7	7,07
país	0,03	5	5,05
ps	0,03	5	5,05
ddhh	0,03	6	6,06

pensamiento	0,03	4	4,04
egoísmo	0,03	6	6,06
allende	0,03	7	7,07
discurso	0,03	5	5,05
administración	0,03	5	5,05
socialismo	0,03	6	6,06
política	0,03	4	4,04
interés	0,03	5	5,05
viaje	0,03	3	3,03
ladrón	0,03	3	3,03
legislativo	0,03	4	4,04
tratado	0,03	3	3,03
justicia	0,03	4	4,04
mentiroso	0,02	3	3,03
lobby	0,02	3	3,03
consejero-regional	0,02	4	4,04
Van-risselberghe	0,02	3	3,03
demagogia	0,02	4	4,04
centro	0,02	3	3,03
ministerio	0,02	5	5,05
candidato	0,02	4	4,04
caval	0,02	3	3,03
problema	0,02	3	3,03
empresa	0,02	3	3,03
juez	0,02	3	3,03

conflicto-de-interés	0,02	3	3,03
gobernación	0,02	3	3,03
campana	0,02	4	4,04
presidencia	0,02	3	3,03
totalitarismo	0,02	3	3,03
suciedad	0,02	3	3,03
capitalismo	0,02	4	4,04
parlamentario	0,02	3	3,03
sistema	0,02	3	3,03
fascismo	0,02	4	4,04
libertad	0,02	4	4,04
aburrimiento	0,02	2	2,02
comunidad	0,02	4	4,04
revolución	0,02	3	3,03
colectividad	0,02	3	3,03
tiranía	0,02	2	2,02
obras-pública	0,02	2	2,02
ciudadanía	0,02	5	5,05
organismo	0,02	4	4,04
público	0,02	3	3,03
institución	0,02	3	3,03
grupos	0,02	3	3,03
mafia	0,02	2	2,02
comunista	0,02	3	3,03
pc	0,02	5	5,05

la-moneda	0,02	3	3,03
trabajo	0,02	4	4,04
corrupción	0,02	2	2,02
ppd	0,02	3	3,03
dictador	0,02	2	2,02
engaño	0,02	2	2,02
participación	0,02	2	2,02
decisión	0,02	2	2,02
cárcel	0,02	3	3,03
desconfianza	0,02	3	3,03
orden	0,02	2	2,02
servicio	0,02	3	3,03
ricardo-lagos	0,02	3	3,03
capacidad	0,02	3	3,03
identidad	0,02	3	3,03
facho	0,02	2	2,02
moneda	0,02	2	2,02
acuerdo	0,01	3	3,03
abogado	0,01	2	2,02
capital	0,01	2	2,02
delegado	0,01	2	2,02
artículo	0,01	2	2,02
autoridad	0,01	4	4,04
falso	0,01	3	3,03
traición	0,01	2	2,02

oportunismo	0,01	2	2,02
burocracia	0,01	2	2,02
gobernanza	0,01	2	2,02
injusticia	0,01	2	2,02
alberto-arenas	0,01	2	2,02
alianza	0,01	3	3,03
peleas	0,01	2	2,02
cargo	0,01	2	2,02
ciencia-política	0,01	2	2,02
acarreo	0,01	2	2,02
rojo	0,01	2	2,02
transantiago	0,01	2	2,02
bien-común	0,01	2	2,02
golpe-de-estado	0,01	3	3,03
gente	0,01	2	2,02
pituto	0,01	2	2,02
délano	0,01	2	2,02
militancia	0,01	2	2,02
diplomacia	0,01	2	2,02
golpe	0,01	2	2,02
flojo	0,01	2	2,02
monopolio	0,01	2	2,02
exterior	0,01	2	2,02
judicial	0,01	2	2,02
ciudadano	0,01	2	2,02

delincuente	0,01	2	2,02
sueldo	0,01	2	2,02
camila-vallejos	0,01	2	2,02
fraude	0,01	2	2,02
fascista	0,01	2	2,02
interior	0,01	2	2,02
plutocracia	0,01	2	2,02
sufragio	0,01	2	2,02
consejal	0,01	2	2,02
farsante	0,01	2	2,02
régimen	0,01	2	2,02
hombre	0,01	2	2,02
marx	0,01	1	1,01
rivera	0,01	1	1,01
color	0,01	1	1,01
nacismo	0,01	1	1,01
injusta	0,01	1	1,01
gorda	0,01	1	1,01
cúpula	0,01	1	1,01
todo	0,01	1	1,01
eficiencia	0,01	1	1,01
bush	0,01	2	2,02
nueva-mayoría	0,01	2	2,02
proyecto	0,01	3	3,03
corbata	0,01	2	2,02

historia	0,01	2	2,02
conjunto	0,01	1	1,01
religión	0,01	1	1,01
mal-uso-del-término	0,01	1	1,01
guaguas	0,01	1	1,01
robos	0,01	1	1,01
representatividad	0,01	1	1,01
fabiola-lagos	0,01	1	1,01
corruptos	0,01	1	1,01
gobernantes	0,01	1	1,01
concertación	0,01	1	1,01
conciencia	0,01	1	1,01
estructura	0,01	1	1,01
mínimo	0,01	1	1,01
control	0,01	2	2,02
populismo	0,01	2	2,02
período	0,01	2	2,02
juzgado	0,01	1	1,01
humano	0,01	1	1,01
navarro	0,01	1	1,01
desacuerdo	0,01	1	1,01
dietas	0,01	1	1,01
liderazgo	0,01	1	1,01
hipócritas	0,01	1	1,01
falsa-conciencia	0,01	1	1,01

chile	0,01	1	1,01
falla-del-sistema	0,01	1	1,01
cultura	0,01	1	1,01
sqm	0,01	1	1,01
estrategia	0,01	1	1,01
momio	0,01	1	1,01
compagnon	0,01	1	1,01
directiva	0,01	1	1,01
ratas	0,01	1	1,01
vocero	0,01	2	2,02
asociación	0,01	2	2,02
responsabilidad	0,01	1	1,01
lemebel	0,01	1	1,01
tolerancia-0	0,01	1	1,01
desvalorización	0,01	1	1,01
beneficio	0,01	1	1,01
desorganización	0,01	1	1,01
retórico	0,01	1	1,01
weber	0,01	1	1,01
necesidad	0,01	1	1,01
longueira	0,01	1	1,01
perón	0,01	1	1,01
poder-ejecutivo	0,01	1	1,01
importante	0,01	1	1,01
conveniencia	0,01	1	1,01

indignado	0,01	1	1,01
forma-de-gobernar	0,01	1	1,01
crítica	0,01	1	1,01
militante	0,01	1	1,01
pública	0,01	1	1,01
estupidez	0,01	1	1,01
extremista	0,01	1	1,01
acción	0,01	2	2,02
interacción	0,01	1	1,01
panfletos	0,01	1	1,01
pseudo	0,01	1	1,01
tráfico	0,01	1	1,01
flojera	0,01	1	1,01
bolaño	0,01	1	1,01
hillary-clinton	0,01	1	1,01
congresistas	0,01	1	1,01
mal-uso	0,01	1	1,01
zurdo	0,01	1	1,01
deber	0,01	1	1,01
desempeño	0,01	1	1,01
falta-de-referetes	0,01	1	1,01
política-pública	0,01	1	1,01
capacidad-de-elegir	0,01	1	1,01
liberalismo	0,01	1	1,01
empresas	0,01	1	1,01

habilidades	0,01	1	1,01
anarquismo	0,01	2	2,02
nación	0,01	2	2,02
internalizar	0,01	1	1,01
concejos	0,01	1	1,01
escombro	0,01	1	1,01
influencia	0,01	1	1,01
mapuche	0,01	1	1,01
bill-clinton	0,01	1	1,01
peñailillo	0,01	1	1,01
posiciones	0,01	1	1,01
crisis	0,01	1	1,01
vocerías	0,01	1	1,01
inteligencia	0,01	1	1,01
ignoracia	0,01	1	1,01
aristóteles	0,01	1	1,01
decretos	0,01	1	1,01
inmoral	0,01	1	1,01
michelle-bachelet	0,01	1	1,01
cívico	0,01	1	1,01
empatía	0,01	1	1,01
ejército	0,01	1	1,01
incapacidad	0,01	1	1,01
marco-ominami	0,01	1	1,01
coalición	0,01	1	1,01

diseño-de-vivir	0,01	1	1,01
libremercado	0,01	1	1,01
principios	0,01	1	1,01
binominal	0,01	2	2,02
movilización	0,01	2	2,02
política-de-estado	0,01	1	1,01
izquierda-revolucionaria	0,01	1	1,01
conflicto	0,01	1	1,01
evolución	0,01	1	1,01
martillo	0,01	1	1,01
contacto	0,01	1	1,01
conflicto-con-perú	0,01	1	1,01
sarkozy	0,01	1	1,01
lucro	0,01	1	1,01
injusticias	0,01	1	1,01
caudillo	0,01	1	1,01
inversión	0,01	1	1,01
norma	0,01	1	1,01
social-democracia	0,01	1	1,01
mentir	0,01	1	1,01
renovación	0,01	1	1,01
sebastián-piñera-	0,01	1	1,01
razón	0,01	1	1,01
logística	0,01	1	1,01
altruismo	0,01	1	1,01

indiferencia	0,01	1	1,01
dueño	0,01	1	1,01
agencia-de-inteligencia	0,01	1	1,01
ego	0,01	1	1,01
ilicitos	0,01	1	1,01
vocación	0,01	1	1,01
verborrea	0,01	1	1,01
mitín	0,01	1	1,01
intención	0,01	1	1,01
genocidio	0,01	1	1,01
república	0,01	2	2,02
boleta	0,01	2	2,02
pps	0,01	1	1,01
ciudad	0,01	1	1,01
izquierda-libertaria	0,01	1	1,01
estrategia-de-cambio	0,01	1	1,01
curioso	0,01	1	1,01
consejales	0,01	1	1,01
referente	0,01	1	1,01
vocero-de-gobierno	0,01	1	1,01
mando	0,01	1	1,01
arreglo	0,01	1	1,01
plata	0,01	1	1,01
autoritarismo	0,01	1	1,01
usura	0,01	1	1,01

popularidad	0,01	1	1,01
igualdad	0,01	1	1,01
millón	0,01	1	1,01
nueva-constitución	0,01	1	1,01
plan	0,01	1	1,01
sentido-común	0,01	1	1,01
relación	0,01	1	1,01
poco-bueno	0,01	1	1,01
debate	0,01	1	1,01
constitucional	0,01	1	1,01
sintonía	0,01	1	1,01
arte	0,01	1	1,01
trampa	0,01	2	2,02
neoliberalismo	0,01	2	2,02
tóxico	0,01	1	1,01
alessandri	0,01	1	1,01
compromisos	0,01	1	1,01
opinión-pública	0,01	1	1,01
cartel	0,01	1	1,01
líder	0,01	1	1,01
lord	0,01	1	1,01
no-trabajo	0,01	1	1,01
pobreza	0,01	1	1,01
división	0,01	1	1,01
coima	0,01	1	1,01

consensuar	0,01	1	1,01
empanadas	0,01	1	1,01
revancha	0,01	1	1,01
estudiante	0,01	1	1,01
ilegalidad	0,01	1	1,01
lucksic	0,01	1	1,01
santiago	0,01	1	1,01
billete	0,01	1	1,01
campaña-política	0,01	1	1,01
ibañez-del-campo	0,01	1	1,01
ferrocarriles	0,01	1	1,01
ajuste-de-sueldo	0,01	1	1,01
ximena-rincón	0,01	1	1,01
propaganda	0,01	1	1,01
movimiento	0,01	1	1,01
fiscal	0,01	1	1,01
nazismo	0,01	1	1,01
mir	0,01	1	1,01
promesa	0,01	1	1,01
enriquecimiento	0,01	1	1,01
programa	0,01	1	1,01
segmentación	0,01	1	1,01
inamovilidad	0,01	1	1,01
#¿NOMBRE?	0,01	1	1,01
conglomerado-político	0,01	1	1,01

patria	0,01	1	1,01
armonizar	0,01	1	1,01
clasismo	0,01	1	1,01
escándalo	0,01	1	1,01
pi	0,01	1	1,01
portavoz	0,01	1	1,01
doctrina	0,01	1	1,01
vendido	0,01	1	1,01
kirchner	0,01	1	1,01
cd	0,01	1	1,01
foro	0,01	1	1,01
fec	0,01	1	1,01
levantamiento	0,01	1	1,01
negatividad	0,01	1	1,01
medio	0,01	1	1,01
cargo-público	0,01	1	1,01
vínculo	0,01	1	1,01
sueldo-reguleques	0,01	1	1,01
colusión	0,01	1	1,01
turbio	0,01	1	1,01
enojo	0,01	1	1,01
tradición	0,01	1	1,01
comunicación	0,01	1	1,01
falta-de-respeto	0,01	1	1,01
defensa	0,01	1	1,01

generar	0,01	1	1,01
cuota	0,01	1	1,01
financiamientoirregular	0,01	1	1,01
padrón	0,01	1	1,01
grupo	0,01	1	1,01
seremi	0,01	2	2,02
manifestación	0,00	2	2,02
pe	0,00	1	1,01
decreto	0,00	1	1,01
washington	0,00	1	1,01
transparencia	0,00	1	1,01
obama	0,00	1	1,01
fiscal-nisman	0,00	1	1,01
clase-social	0,00	1	1,01
demócrata	0,00	1	1,01
banco	0,00	1	1,01
año	0,00	1	1,01
amarillo	0,00	1	1,01
mejora	0,00	1	1,01
iglesia-católica	0,00	1	1,01
regulación	0,00	1	1,01
hitler	0,00	1	1,01
periodimso	0,00	1	1,01
sincero	0,00	1	1,01
malestar	0,00	1	1,01

burbuja	0,00	1	1,01
protesta	0,00	1	1,01
colectivización	0,00	1	1,01
aporte-	0,00	1	1,01
registroelectoral	0,00	1	1,01
primer-ministro	0,00	1	1,01
desorden	0,00	1	1,01
che-guevara	0,00	1	1,01
ulloa	0,00	1	1,01
tomar	0,00	1	1,01
desiciones	0,00	1	1,01
meo	0,00	1	1,01
republicano	0,00	1	1,01
compromiso	0,00	1	1,01
stalin	0,00	1	1,01
censura	0,00	1	1,01
desilgualdad	0,00	1	1,01
ayuda	0,00	1	1,01
diferencia-social	0,00	1	1,01
pro	0,00	1	1,01
core	0,00	1	1,01
palacio	0,00	1	1,01
preocupación	0,00	1	1,01
nepotismo	0,00	1	1,01
memoria	0,00	1	1,01

irregular	0,00	1	1,01
urna	0,00	1	1,01
israel	0,00	1	1,01
operador	0,00	1	1,01
avance	0,00	1	1,01
comuna	0,00	1	1,01
tv-senado	0,00	1	1,01
constancia	0,00	1	1,01
conferencia	0,00	1	1,01
letelier	0,00	1	1,01
gladys-marín	0,00	1	1,01
marchar	0,00	1	1,01
diferencia	0,00	1	1,01
bajo	0,00	1	1,01
estatuto	0,00	1	1,01
nacional-socialismo	0,00	1	1,01
anti-lider	0,00	1	1,01
poco	0,00	1	1,01
capuchinos	0,00	1	1,01
sebastián-dávalos	0,00	1	1,01
injusticia	0,00	1	1,01
descarado	0,00	1	1,01
sentido	0,00	1	1,01
ilegal	0,00	1	1,01
gasto-público	0,00	1	1,01

boicot	0,00	1	1,01
desarrollo	0,00	1	1,01
cadenas	0,00	1	1,01
región	0,00	1	1,01
michelle-obama	0,00	1	1,01
frei	0,00	1	1,01
gobernar	0,00	1	1,01
legislación	0,00	1	1,01
educación-gratis	0,00	1	1,01
convivir	0,00	1	1,01
defectos	0,00	1	1,01
guerra	0,00	1	1,01
repetición	0,00	1	1,01
oportunista	0,00	1	1,01
provado	0,00	1	1,01
proyectedeley	0,00	1	1,01
activista	0,00	1	1,01
1810	0,00	1	1,01
banda-presidencial	0,00	1	1,01
kennedy	0,00	1	1,01
pasión	0,00	1	1,01
alto-cargo	0,00	1	1,01
asco	0,00	1	1,01
despido	0,00	1	1,01
lujo	0,00	1	1,01

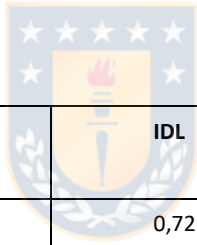
los-mismos-de-siempre	0,00	1	1,01
socialista	0,00	1	1,01
fuero	0,00	1	1,01
solución	0,00	1	1,01
deuda	0,00	1	1,01
factura	0,00	1	1,01
votantes	0,00	1	1,01
oposición	0,00	1	1,01
individualidad	0,00	1	1,01
consejeros	0,00	1	1,01
independencia	0,00	1	1,01
junta-de-vecinos	0,00	1	1,01
desonestidad	0,00	1	1,01
inconformidad	0,00	1	1,01
organizar	0,00	1	1,01
globalización	0,00	1	1,01
debilidad	0,00	1	1,01
partidos-político	0,00	1	1,01
encima-de-la-ley	0,00	1	1,01
buena-vida	0,00	1	1,01
sueldo-alto	0,00	1	1,01
ilícito	0,00	1	1,01
subsecretaria	0,00	1	1,01
radicales	0,00	1	1,01
popular	0,00	1	1,01

traje	0,00	1	1,01
oficialismo	0,00	1	1,01
quejas	0,00	1	1,01
proyecto-de-ley	0,00	1	1,01
vallejos	0,00	1	1,01
plazo	0,00	1	1,01
humor	0,00	1	1,01
lavín	0,00	1	1,01
obsoletos	0,00	1	1,01
soborno	0,00	1	1,01
feo	0,00	1	1,01
social	0,00	1	1,01
prensa	0,00	1	1,01
afiliación	0,00	1	1,01
malo	0,00	1	1,01
firma	0,00	1	1,01
opinión	0,00	1	1,01
caso	0,00	1	1,01
beige	0,00	1	1,01
pollos	0,00	1	1,01
word	0,00	1	1,01
riqueza	0,00	1	1,01
fluctuación	0,00	1	1,01
financiamiento	0,00	1	1,01
subsecretario	0,00	1	1,01

comisiones	0,00	1	1,01
greca	0,00	1	1,01
1973	0,00	1	1,01
creencia	0,00	1	1,01
aporte	0,00	1	1,01
dirigentes	0,00	1	1,01
mercado	0,00	1	1,01
sócrates	0,00	1	1,01
familia	0,00	1	1,01
subsecretarios	0,00	1	1,01
universidad	0,00	1	1,01
quehacer	0,00	1	1,01
dificultad	0,00	1	1,01
negociación	0,00	1	1,01
género	0,00	1	1,01
élite	0,00	1	1,01
demanda	0,00	1	1,01
apoyo	0,00	1	1,01
cotidiano	0,00	1	1,01
representante	0,00	1	1,01
habitual	0,00	1	1,01
sesiones	0,00	1	1,01
representativo	0,00	1	1,01
vínculos	0,00	1	1,01
bienestar	0,00	1	1,01

serviciopúblico	0,00	1	1,01
conglomerado	0,00	1	1,01
biencomún	0,00	1	1,01
encuentro	0,00	1	1,01
mandato	0,00	1	1,01
ordenamientoterritorial	0,00	1	1,01
nacionalismo	0,00	1	1,01
plurinacionalismo	0,00	1	1,01

3. Deporte.



Vocablo	IDL	Frecuencia	% Aparición
fútbol	0,72	78	78,8
tenis	0,52	64	64,6
básquetbol	0,49	59	59,6
natación	0,31	45	45,5
atletismo	0,25	35	35,4
voleibol	0,24	31	31,3
salud	0,21	26	26,3
rugby	0,21	30	30,3
golf	0,16	25	25,3
ciclismo	0,14	22	22,2
handbol	0,14	21	21,2

gimnasia	0,13	20	20,2
karate	0,12	21	21,2
pelota	0,10	15	15,2
correr	0,10	16	16,2
pingpong	0,10	15	15,2
estadio	0,09	14	14,1
béisbol	0,08	13	13,1
patinaje	0,08	14	14,1
competencia	0,07	12	12,1
gimnasio	0,07	12	12,1
ajedrez	0,07	11	11,1
pasión	0,07	10	10,1
remo	0,07	12	12,1
colocolo	0,07	9	9,1
fútbol-americano	0,07	11	11,1
polo	0,06	11	11,1
vida	0,06	8	8,1
gimnasia-artística	0,05	9	9,1
cancha	0,05	10	10,1
hockey	0,05	11	11,1
esfuerzo	0,05	9	9,1
selección	0,05	7	7,1
bienestar	0,05	6	6,1
vida-sana	0,05	8	8,1
equipo	0,05	9	9,1

disciplina	0,05	7	7,1
gimnasia-rítmica	0,05	8	8,1
equitación	0,04	8	8,1
olimpiada	0,04	7	7,1
juego	0,04	6	6,1
uchile	0,04	6	6,1
crossfit	0,04	5	5,1
ejercicio	0,04	6	6,1
boxeo	0,04	9	9,1
surf	0,04	6	6,1
bicicleta	0,04	6	6,1
jugador	0,04	6	6,1
energía	0,04	6	6,1
perseverancia	0,04	6	6,1
balón	0,04	7	7,1
actividad-física	0,04	4	4,0
mundial	0,04	8	8,1
maratón	0,04	7	7,1
actividad	0,04	5	5,1
halterofilia	0,04	6	6,1
constancia	0,04	5	5,1
danza	0,04	7	7,1
movimiento	0,04	4	4,0
trote	0,03	7	7,1
judo	0,03	7	7,1

trabajo	0,03	6	6,1
copa-américa	0,03	6	6,1
saludable	0,03	5	5,1
físico	0,03	4	4,0
gol	0,03	5	5,1
taekwondo	0,03	6	6,1
triatlón	0,03	5	5,1
pesa	0,03	5	5,1
salto	0,03	7	7,1
running	0,03	4	4,0
carrera	0,03	6	6,1
zapatilla	0,03	6	6,1
dinero	0,03	4	4,0
vela	0,03	4	4,0
atleta	0,03	5	5,1
anfp	0,03	4	4,0
yoga	0,03	6	6,1
fuerza	0,03	4	4,0
arco	0,03	5	5,1
sky	0,03	5	5,1
juegos-olimpicos	0,03	4	4,0
water-polo	0,03	4	4,0
dedicación	0,03	5	5,1
medalla	0,03	5	5,1
canotaje	0,03	4	4,0

pasto	0,03	4	4,0
católica	0,02	4	4,0
jabalina	0,02	4	4,0
alegría	0,02	4	4,0
esgrima	0,02	3	3,0
saltar	0,02	4	4,0
cansancio	0,02	4	4,0
tenis-de-mesa	0,02	4	4,0
superación	0,02	3	3,0
lesión	0,02	4	4,0
amor	0,02	4	4,0
meta	0,02	4	4,0
felicidad	0,02	4	4,0
treking	0,02	4	4,0
sánchez	0,02	4	4,0
zumba	0,02	5	5,1
pesca	0,02	4	4,0
punto	0,02	3	3,0
entrenamiento	0,02	4	4,0
aeróbica	0,02	4	4,0
compañerismo	0,02	3	3,0
sudor	0,02	4	4,0
NBA	0,02	3	3,0
distracción	0,02	3	3,0
sano	0,02	3	3,0

entrenador	0,02	4	4,0
MMA	0,02	3	3,0
premio	0,02	3	3,0
diversión	0,02	2	2,0
automovilismo	0,02	3	3,0
baile	0,02	4	4,0
recreación	0,02	2	2,0
obediencia	0,02	2	2,0
skaters	0,02	2	2,0
estado-físico	0,02	3	3,0
fórmula-1	0,02	2	2,0
américa	0,02	3	3,0
caminata	0,02	3	3,0
massú	0,02	2	2,0
balón-mano	0,02	2	2,0
tiempo	0,02	2	2,0
raqueta	0,02	3	3,0
sedentarismo	0,02	2	2,0
estilo	0,02	3	3,0
multitud	0,02	2	2,0
tiro-con-arco	0,02	3	3,0
triunfo	0,02	3	3,0
derrota	0,02	3	3,0
logro	0,02	2	2,0
joven	0,02	2	2,0

cuerpo	0,02	3	3,0
real-madrid	0,02	2	2,0
campeonato	0,02	3	3,0
bravo	0,02	3	3,0
calor	0,02	2	2,0
buzo	0,02	3	3,0
tapia	0,02	2	2,0
piscina	0,02	3	3,0
partido	0,01	2	2,0
pasatiempo	0,01	2	2,0
ballet	0,01	2	2,0
trotar	0,01	2	2,0
lucha-libre	0,01	2	2,0
crossfit	0,01	2	2,0
ríos	0,01	2	2,0
soccer	0,01	2	2,0
ciclismo	0,01	2	2,0
jockey	0,01	2	2,0
vida-saludable	0,01	2	2,0
dolor	0,01	2	2,0
gonzález	0,01	2	2,0
spinning	0,01	2	2,0
entretenimiento	0,01	2	2,0
sacrificio	0,01	2	2,0
corrupción	0,01	2	2,0

mente	0,01	3	3,0
udec	0,01	2	2,0
waterpolo	0,01	3	3,0
baloncesto	0,01	2	2,0
artes-marciales	0,01	2	2,0
snowboard	0,01	3	3,0
ánimo	0,01	3	3,0
fútbolito	0,01	2	2,0
máquina	0,01	3	3,0
peso	0,01	3	3,0
delgadez	0,01	2	2,0
cardio	0,01	2	2,0
diario	0,01	2	2,0
entretención	0,01	2	2,0
orgullo	0,01	2	2,0
zamorano	0,01	2	2,0
pista	0,01	3	3,0
deportista	0,01	2	2,0
baby-fútbol	0,01	2	2,0
auspiciador	0,01	3	3,0
motivación	0,01	2	2,0
calorías	0,01	2	2,0
frío	0,01	2	2,0
jiujitsu	0,01	2	2,0
trofeo	0,01	2	2,0

pista-atlética	0,01	2	2,0
gimnasiaartística	0,01	2	2,0
corazón	0,01	2	2,0
red	0,01	2	2,0
salas	0,01	2	2,0
educación-física	0,01	2	2,0
mejora	0,01	2	2,0
ronaldo	0,01	2	2,0
messi	0,01	2	2,0
libertad	0,01	2	2,0
pie	0,01	2	2,0
rítmica	0,01	2	2,0
juegos-olímpicos	0,01	2	2,0
corrida	0,01	2	2,0
lanzamiento-de-bala	0,01	2	2,0
implemento	0,01	3	3,0
musculatura	0,01	2	2,0
escalada	0,01	2	2,0
alimentación	0,01	2	2,0
arquero	0,01	2	2,0
ski	0,01	2	2,0
vidal	0,01	2	2,0
ejercicios	0,01	1	1,0
flojera	0,01	1	1,0

hockey-sobre-césped	0,01	1	1,0
incentivo	0,01	1	1,0
impotante	0,01	1	1,0
recurso	0,01	1	1,0
gusto	0,01	1	1,0
sentimiento	0,01	1	1,0
zapatillas	0,01	1	1,0
fome	0,01	1	1,0
yugo	0,01	1	1,0
garra	0,01	3	3,0
torneo	0,01	3	3,0
barras	0,01	2	2,0
nadar	0,01	2	2,0
bala	0,01	2	2,0
ganar	0,01	2	2,0
cdf	0,01	2	2,0
mujer	0,01	2	2,0
músculo	0,01	2	2,0
copa-davis	0,01	1	1,0
hockey-sobre-hielo	0,01	1	1,0
independiente-de-cauquenes	0,01	1	1,0
dinámica	0,01	1	1,0
maradona	0,01	1	1,0
poco	0,01	1	1,0

albo	0,01	1	1,0
bádminton	0,01	1	1,0
nuevamente	0,01	1	1,0
amigo	0,01	2	2,0
regla	0,01	2	2,0
agua	0,01	2	2,0
campeón	0,01	2	2,0
estudio	0,01	1	1,0
chino-ríos	0,01	1	1,0
buena-alimentación	0,01	1	1,0
juegos-de-rol	0,01	1	1,0
test-de-naveta	0,01	1	1,0
u-de-chhile	0,01	1	1,0
bailarina	0,01	1	1,0
blanco-y-negro	0,01	1	1,0
quidish	0,01	1	1,0
alto-rendimietno	0,01	1	1,0
adicción	0,01	1	1,0
marcas	0,01	1	1,0
entusiasmo	0,01	1	1,0
nado-sincronizado	0,01	2	2,0
mancuerna	0,01	1	1,0
net	0,01	1	1,0
iberia	0,01	1	1,0
sabiduría	0,01	1	1,0

relajo	0,01	1	1,0
universidad-de-chhile	0,01	1	1,0
color	0,01	1	1,0
progreso	0,01	1	1,0
cobreola	0,01	1	1,0
ayuda	0,01	1	1,0
individualismo	0,01	1	1,0
club	0,01	1	1,0
andinismo	0,01	1	1,0
deportiva	0,01	1	1,0
descanso	0,01	1	1,0
tiro	0,01	2	2,0
penal	0,01	2	2,0
juventud	0,01	2	2,0
desarrollo	0,01	1	1,0
ricos	0,01	1	1,0
comida	0,01	1	1,0
ace	0,01	1	1,0
victoria	0,01	1	1,0
subdesarrollo	0,01	1	1,0
desconexión	0,01	1	1,0
v/s	0,01	1	1,0
fifa	0,01	1	1,0
union-española	0,01	1	1,0

kitesurf	0,01	1	1,0
oro	0,01	1	1,0
basquetbol	0,01	1	1,0
gozo	0,01	1	1,0
billar	0,01	1	1,0
minuto	0,01	1	1,0
palín	0,01	2	2,0
música	0,01	2	2,0
levantamiento-de-pesa	0,01	1	1,0
élite	0,01	1	1,0
match-point	0,01	1	1,0
innovación	0,01	1	1,0
brasil	0,01	1	1,0
entrega	0,01	1	1,0
fanatismo	0,01	1	1,0
criquet	0,01	1	1,0
evasión	0,01	1	1,0
nacional	0,01	1	1,0
espacios	0,01	1	1,0
jordan	0,01	1	1,0
participantes	0,01	1	1,0
divertido	0,01	1	1,0
generosidad	0,01	1	1,0
oportunidades	0,01	1	1,0

empate	0,01	1	1,0
runner	0,01	1	1,0
record	0,01	1	1,0
trailrunning	0,01	1	1,0
ritmo	0,01	1	1,0
montañismo	0,01	1	1,0
lucha	0,01	2	2,0
odesur	0,01	1	1,0
flaute	0,01	1	1,0
set	0,01	1	1,0
equitación	0,01	1	1,0
proteínas	0,01	1	1,0
tiro-al-arco	0,01	1	1,0
cricket	0,01	1	1,0
barcelona	0,01	1	1,0
vilchez	0,01	1	1,0
champions	0,01	1	1,0
aprendizaje	0,01	1	1,0
playlist	0,01	1	1,0
squach	0,01	1	1,0
responsabilidad	0,01	1	1,0
chuncho	0,01	1	1,0
conflicto	0,01	1	1,0
respiración	0,01	1	1,0
bañarse	0,01	1	1,0

dt	0,01	1	1,0
fracaso	0,01	1	1,0
inseguridad	0,01	1	1,0
aro	0,01	1	1,0
barroso	0,01	1	1,0
coordinación	0,01	1	1,0
descenso	0,01	1	1,0
ropa-deportiva	0,01	1	1,0
cycling	0,01	1	1,0
gente-delgada	0,01	1	1,0
rutina	0,01	1	1,0
balonmano	0,01	1	1,0
babyfútbol	0,01	1	1,0
pilates	0,01	2	2,0
muschin-dog	0,01	1	1,0
indio	0,01	1	1,0
adrenalina	0,01	1	1,0
apariencia	0,01	1	1,0
actitud	0,01	1	1,0
buena-calidad-de- vida	0,01	1	1,0
body-board	0,01	1	1,0
natalia-ducco	0,01	1	1,0
beausejour	0,01	1	1,0
europa	0,01	1	1,0

industria	0,01	1	1,0
copa	0,01	1	1,0
posta	0,01	2	2,0
skate	0,01	1	1,0
parapente	0,01	1	1,0
esquiar	0,01	1	1,0
longevidad	0,01	1	1,0
comodidad	0,01	1	1,0
corredora	0,01	1	1,0
musculoso	0,01	1	1,0
ayrton-sena	0,01	1	1,0
off-side	0,01	1	1,0
larga-vida	0,01	1	1,0
pasear	0,01	1	1,0
federer	0,01	1	1,0
estereotipo	0,01	1	1,0
toxinas	0,01	1	1,0
mundial-de-fútbol	0,01	1	1,0
valdés	0,01	1	1,0
champions-league	0,01	1	1,0
financiamiento	0,01	1	1,0
bailar	0,01	1	1,0
calentamiento	0,01	1	1,0
malabares	0,01	1	1,0
gasto-de-energía	0,01	1	1,0

transpirar	0,01	1	1,0
televisión	0,01	1	1,0
bowling	0,01	1	1,0
representantes- nacionales	0,01	1	1,0
djokovic	0,01	1	1,0
canopy	0,01	1	1,0
plata	0,01	1	1,0
medel	0,01	1	1,0
control	0,01	1	1,0
hombre	0,01	1	1,0
pinto	0,01	1	1,0
pavez	0,01	1	1,0
tristeza	0,01	1	1,0
rol-social	0,01	1	1,0
roja	0,01	1	1,0
aire-libre	0,01	1	1,0
práctica	0,01	1	1,0
motociclismo	0,01	1	1,0
pinpon	0,01	1	1,0
surfing	0,01	1	1,0
poco-acceso	0,01	1	1,0
pinochet	0,01	1	1,0
estabilidad	0,01	1	1,0
entrada	0,01	1	1,0

Acondicionamiento-físico	0,01	1	1,0
woods	0,01	1	1,0
futbolista	0,01	1	1,0
transpiración	0,01	1	1,0
abdominal	0,01	1	1,0
kayak	0,01	1	1,0
sampaoli	0,01	1	1,0
gradería	0,01	1	1,0
habilidad	0,01	1	1,0
suazo	0,01	1	1,0
asia	0,01	1	1,0
emoción	0,01	1	1,0
básquetboll	0,01	1	1,0
saneamiento	0,01	1	1,0
integridad	0,01	1	1,0
posibilidad	0,01	1	1,0
cobresal	0,01	1	1,0
olímpico	0,01	1	1,0
autoestima	0,01	1	1,0
chueca	0,01	1	1,0
pausa	0,01	1	1,0
tranquilidad	0,01	1	1,0
preparación	0,01	1	1,0
racketball	0,01	1	1,0

gimnasia-olímpica	0,01	1	1,0
evaluación-física	0,01	1	1,0
marcador	0,01	1	1,0
apoyo	0,01	1	1,0
diagnóstico	0,01	1	1,0
trabajo-en-equipo	0,01	1	1,0
lasarte	0,01	1	1,0
calidad	0,01	1	1,0
delantero	0,01	1	1,0
cesta	0,01	1	1,0
frescura	0,01	1	1,0
delgado	0,01	1	1,0
oceanía	0,01	1	1,0
droga	0,01	1	1,0
minusválidos	0,01	1	1,0
velleybol	0,01	1	1,0
complemento	0,01	1	1,0
salir-adelante	0,01	1	1,0
niño	0,01	1	1,0
tenisdemesa	0,01	1	1,0
desafío	0,01	1	1,0
profesionalismo	0,01	1	1,0
aikido	0,01	1	1,0
poder	0,00	1	1,0

lanzamiento-de- arco	0,00	1	1,0
tiro-al-aire	0,00	1	1,0
puntuación	0,00	1	1,0
unión	0,00	1	1,0
cabalgar	0,00	1	1,0
decatlón	0,00	1	1,0
pivot	0,00	1	1,0
rebote	0,00	1	1,0
paredes	0,00	1	1,0
áfrica	0,00	1	1,0
odiosidad	0,00	1	1,0
cuerpo-sano	0,00	1	1,0
oportunidad	0,00	1	1,0
segunda	0,00	1	1,0
niña	0,00	1	1,0
viento	0,00	1	1,0
copa-europea	0,00	1	1,0
voleybol	0,00	1	1,0
hábito	0,00	1	1,0
musculación	0,00	2	2,0
lanzamiento-de- garrocha	0,00	1	1,0
lugar	0,00	1	1,0
segundones	0,00	1	1,0
talento	0,00	1	1,0

barra-brava	0,00	1	1,0
calentamiento	0,00	1	1,0
gimnasia-deportiva	0,00	1	1,0
taca-taca	0,00	1	1,0
persona	0,00	1	1,0
medalla-olímpicas	0,00	1	1,0
medio	0,00	1	1,0
chile	0,00	1	1,0
veleibol	0,00	1	1,0
escalar	0,00	1	1,0
rendimiento	0,00	1	1,0
metros	0,00	1	1,0
reunión	0,00	1	1,0
extremo	0,00	1	1,0
capitán	0,00	1	1,0
mundo	0,00	1	1,0
gana	0,00	1	1,0
debilidad	0,00	1	1,0
acrobacia	0,00	1	1,0
airelibre	0,00	1	1,0
gimnasia-rítmica	0,00	1	1,0
pelotas	0,00	1	1,0
huachipato	0,00	1	1,0
mombo	0,00	1	1,0
salto-largo	0,00	1	1,0

canasta	0,00	1	1,0
practicar	0,00	1	1,0
giro	0,00	1	1,0
cazely	0,00	1	1,0
marcelo-ríos	0,00	1	1,0
fortaleza	0,00	1	1,0
enfermería	0,00	1	1,0
ley-estadio-seguro	0,00	1	1,0
aplicación	0,00	1	1,0
guardiola	0,00	1	1,0
flecha	0,00	1	1,0
verano	0,00	1	1,0
fractura	0,00	1	1,0
destreza	0,00	1	1,0
paracaidismo	0,00	1	1,0
ropa	0,00	1	1,0
lanzamiento	0,00	1	1,0
futilidad	0,00	1	1,0
entrenamiento- canino	0,00	1	1,0
remar	0,00	1	1,0
pinilla	0,00	1	1,0
malla	0,00	1	1,0
pila	0,00	1	1,0
linesman	0,00	1	1,0

terapia	0,00	1	1,0
ensaladas	0,00	1	1,0
masa-muscular	0,00	1	1,0
mantener	0,00	1	1,0
pena	0,00	1	1,0
ganador	0,00	1	1,0
trotadora	0,00	1	1,0
zapatos	0,00	1	1,0
ejercitar	0,00	1	1,0
jugada	0,00	1	1,0
kinesiología	0,00	1	1,0
piquero	0,00	1	1,0
seguro	0,00	1	1,0
aire	0,00	1	1,0
nado	0,00	1	1,0
guardalínea	0,00	1	1,0
contrato	0,00	1	1,0
verde	0,00	1	1,0
patín	0,00	1	1,0
desestrés	0,00	1	1,0
tecnicaa	0,00	1	1,0
guatazo	0,00	1	1,0
odio	0,00	1	1,0
condición	0,00	1	1,0
rotación	0,00	1	1,0

torcedura	0,00	1	1,0
espíritu	0,00	1	1,0
periodista- deportivo	0,00	1	1,0
esquí	0,00	1	1,0
noche	0,00	1	1,0
física	0,00	1	1,0
rueda	0,00	1	1,0
aburrido	0,00	1	1,0
femenino	0,00	1	1,0
polera	0,00	1	1,0
estrategia	0,00	1	1,0
copa-libertadores	0,00	1	1,0
kayack	0,00	1	1,0
garrocha	0,00	1	1,0
Transmisión- deportiva	0,00	1	1,0
valor	0,00	1	1,0
bote	0,00	1	1,0
relajación	0,00	1	1,0
reir	0,00	1	1,0
power-ade	0,00	1	1,0
concentración	0,00	1	1,0
saltolargo	0,00	1	1,0
mundialdefútbol	0,00	1	1,0
individualidad	0,00	1	1,0

energizante	0,00	1	1,0
relato	0,00	1	1,0
barra	0,00	1	1,0
gimnasia-aeróbica	0,00	1	1,0
juegosolímpicos	0,00	1	1,0
taller	0,00	1	1,0
fitness	0,00	1	1,0
aeroyoga	0,00	1	1,0
activo	0,00	1	1,0
plan	0,00	1	1,0
hichada	0,00	1	1,0
fanaticada	0,00	1	1,0
semilargas	0,00	1	1,0
mano	0,00	1	1,0
estrés	0,00	1	1,0
cheerleading	0,00	1	1,0
10k	0,00	1	1,0
clasificatoria	0,00	1	1,0
5k	0,00	1	1,0
camiseta	0,00	1	1,0
casco	0,00	1	1,0
vital	0,00	1	1,0
río-2016	0,00	1	1,0
caza	0,00	1	1,0
usain-bolt	0,00	1	1,0

animar	0,00	1	1,0
erika-olivera	0,00	1	1,0
perder	0,00	1	1,0
playa	0,00	1	1,0
riesgo	0,00	1	1,0
aventura	0,00	1	1,0
vacación	0,00	1	1,0
estudiante	0,00	1	1,0



