



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Studi Linguistici e Letterari

CORSO DI DOTTORATO DI RICERCA IN
SCIENZE LINGUISTICHE, FILOGICHE E LETTERARIE
CICLO XXXII

**Disponibilidad y riqueza léxica en un grupo de estudiantes
universitarios italianos de ELE**

Coordinatore: Ch.mo Prof. Rocco Coronato

Supervisore: Ch.ma Prof.ssa Carmen Castillo Peña

Dottoranda: Giulia Nalesso

A mi madre.

AGRADECIMIENTOS

En primer lugar, quiero expresar mi gratitud a mi directora, la Profesora Carmen Castillo Peña, por toda la confianza que depositó en mí y en este proyecto.

Para la recogida de datos fue esencial la colaboración de los Profesores de español del *Dipartimento di Studi Linguistici e Letterari* de la Universidad de Padua que me cedieron parte de sus clases: gracias por contribuir con esta investigación. Igualmente, mi gratitud va a los alumnos que cumplimentaron las encuestas.

A la Dra. Anna Polo debo agradecerle los buenos consejos durante estos años.

Un sincero agradecimiento a la Profesora Isabel García Parejo y al Profesor Narciso M. Contreras Izquierdo, que me acogieron en sus Universidades y permitieron que mi estancia en España fuera fácil y provechosa.

Mi reconocimiento también va a los investigadores que contribuyeron a resolver muchas de mis dudas, me concedieron utilizar los programas de cálculo necesarios para la realización de este trabajo y por el interés demostrado, en particular a Marta Sánchez-Saus Laserna, Matías Hidalgo Gallardo y Natividad Hernández Muñoz.

Más allá de lo académico, esta Tesis no habría llegado a buen fin sin mi familia. El agradecimiento más profundo se dirige hacia mis padres, Maria Grazia y Arcadio Giuseppe, y mi compañero de vida, Alessandro, por la confianza que han tenido en mí desde hace ya tantos años, por su paciencia y apoyo incondicional para que yo pudiera recorrer este camino. Agradezco también a mis tíos y a mi primo Enrico por estar siempre orgullosos. Para terminar, otro agradecimiento merece una persona sin la cual todo esto no hubiera sido posible: gracias, papá.

A mis amigas de toda la vida, Elettra y Silvia, y a todos los amigos que tanto me han querido y animado.

Por último, no me quiero olvidar de mis colegas, en particular Diana, por haber compartido conmigo este recorrido.

ÍNDICE

Índice de tablas, gráficos y figuras	12
Siglas y acrónimos utilizados	23
Sinossi.....	24
Abstract.....	25
Introducción.....	26
Capítulo I. Lingüística aplicada a la enseñanza de las lenguas extranjeras.....	38
1. Introducción.....	38
2. Teorías lingüísticas sobre la enseñanza de lenguas extranjeras.....	39
2.1 La teoría conductista.....	39
2.2 La teoría nativista	40
2.3 La teoría cognitivista	41
2.4 La teoría ambientalista.....	42
2.5 La teoría interaccionista.....	42
3. La didáctica del léxico de las lenguas extranjeras	43
3.1 Primera etapa	45
3.2 Segunda etapa	48
3.3 Tercera etapa.....	51
3.4 Tendencias actuales	56
4. Materiales para la didáctica de ELE	59
4.1 Sobre los manuales	59
4.2 Sobre el diccionario	67
5. La selección léxica.....	73
5.1 Criterios básicos.....	74

5.2	Criterios ligados a la situación de enseñanza-aprendizaje.....	76
5.3	Criterios del MCER y del PCIC	77
5.4	La disponibilidad léxica como criterio de selección.....	77
6.	La competencia léxica	79
Capítulo II.	La disponibilidad y la riqueza léxica: orígenes y desarrollo actual	84
1.	Premisa terminológica	84
1.1	Palabras y vocablos.....	84
1.2	Tipos de léxico	85
2.	La disponibilidad léxica.....	87
2.1	Introducción	87
2.2	El estudio pionero	88
2.3	Los sucesores del <i>Français Fondamental</i>	94
2.4	La disponibilidad léxica en el mundo hispánico.....	95
2.5	La disponibilidad léxica en ELE.....	100
2.6	Otras aplicaciones de la disponibilidad léxica.....	105
3.	La riqueza léxica	108
3.1	Introducción	108
3.2	Estudios pioneros.....	108
3.3	La riqueza léxica en el mundo hispánico.....	113
3.4	La riqueza léxica en ELE.....	116
3.5	Otras aplicaciones de la riqueza léxica	118
Capítulo III.	Metodología de la investigación	122
1.	Pautas generales	122
2.	La muestra	122
2.1	Selección de los informantes	122
2.2	Las variables sociolingüísticas	124

2.3 Codificación de las variables	127
3. La encuesta	128
3.1 El cuestionario sociológico	129
3.2 La prueba de disponibilidad léxica	131
3.3 La prueba de riqueza léxica	134
4. Análisis de la disponibilidad léxica	135
4.1 Los índices de disponibilidad léxica	135
4.2 Tratamiento informático de los datos	138
4.3 La edición de los datos	139
5. Análisis de la riqueza léxica	151
5.1 Los índices de riqueza léxica	151
5.2 Tratamiento informático de los datos	154
5.3 La edición de los datos	155
Capítulo IV. Análisis cuantitativo de la disponibilidad léxica	158
1. Introducción	158
2. Análisis transversal	159
2.1 Resultados generales	159
2.2 Conclusiones parciales	170
2.3 Resultados por variable	171
2.4 Conclusiones del análisis transversal	192
3. Análisis longitudinal	194
3.1 Número de palabras	195
3.2 Número de vocablos	199
3.3 Índice de cohesión y densidad léxica	202
3.4 Conclusiones del análisis longitudinal	204
4. Análisis comparativo	204

4.1 Análisis descriptivo	206
4.2 Conclusiones del análisis comparativo	211
5. Conclusiones del análisis cuantitativo de la disponibilidad léxica	213
Capítulo V. Análisis cualitativo de la disponibilidad léxica	216
1. Introducción	216
2. Análisis transversal	216
2.1 Resultados generales.....	216
2.2 Conclusiones parciales.....	229
2.3 Resultados por variable.....	231
2.4 Conclusiones del análisis transversal.....	263
3. Análisis longitudinal	264
3.1 Análisis descriptivo	264
3.2 Conclusiones parciales.....	275
4. Análisis comparativo	278
4.1 Análisis comparativo con el CREA.....	278
4.2 Análisis comparativo con el CAES	279
4.3 Conclusiones del análisis comparativo	281
5. Conclusiones del análisis cualitativo de la disponibilidad léxica	281
Capítulo VI. Análisis cuantitativo de la riqueza léxica	284
1. Introducción	284
2. Análisis transversal	284
2.1 Resultados generales.....	284
2.2 Conclusiones parciales.....	292
2.3 Resultados por variable.....	293
2.4 Conclusiones del análisis transversal.....	320
3. Análisis longitudinal	322

3.1 Diversidad léxica	325
3.2 Densidad léxica.....	328
3.3 Conclusiones del análisis longitudinal.....	330
4. Análisis comparativo	331
4.1 Análisis descriptivo	333
4.2 Conclusiones del análisis comparativo	335
5. Conclusiones del análisis cuantitativo de la riqueza léxica	336
Capítulo VII. Aproximación a un análisis cualitativo de la riqueza léxica	340
1. Introducción	340
2. Listas de frecuencia	340
2.1 Análisis general	341
2.2 Análisis de las palabras nocionales.....	342
3. Palabras clave	348
4. Conclusiones del análisis cualitativo de la riqueza léxica	352
Conclusiones.....	354
Referencias bibliográficas	374
Anexos.....	401
Anexo I – Test de medición de la competencia léxica	401
Anexo II – Lista de parada.....	408
Anexo III – Listado general de léxico disponible (Corpus general).....	409
Anexo IV – Listado alfabético de léxico disponible (Corpus general).....	456
Anexo V – Listado general de léxico disponible (Corpus B2_ b).....	503
Anexo VI – Listado alfabético de léxico disponible (Corpus B2_ b).....	543
Anexo VII – Datos de riqueza léxica (Corpus general).....	583
Anexo VIII – Datos de riqueza léxica (Corpus B2_ b).....	586

ÍNDICE DE TABLAS, GRÁFICOS Y FIGURAS

Tabla 1. Distribución de los informantes en función de la variable “sexo”.	125
Tabla 2. Distribución de los informantes en función de la variable “nivel de ELE”. ..	126
Tabla 3. Percepción de los informantes de su nivel de ELE.	126
Tabla 4. Opinión de los informantes sobre la dificultad de aprender español.....	127
Tabla 5. Distribución de los informantes en función de la variable “conocimiento de otras LE”.....	127
Tabla 6. Codificación de las variantes de las variables.	128
Tabla 7. Pregunta 9: ¿Qué cree que es más importante conocer de una LE?	130
Tabla 8. Pregunta 10: ¿Cuánto cree que es importante conocer el vocabulario de una LE?	130
Tabla 9. Pregunta 11: ¿Cómo cree que aprende más vocabulario?	131
Tabla 10. Índices generales de DL.	159
Tabla 11. Número total de palabras y media por informante.	160
Tabla 12. CI ordenados en función de su productividad.	161
Tabla 13. Número total de vocablos.	163
Tabla 14. CI ordenados en función de su variedad.....	164
Tabla 15. Índice de cohesión y densidad léxica por CI.	165
Tabla 16. CI ordenados en función de su cohesión y densidad.	166
Tabla 17. Índices de DL (recapitulación).	168
Tabla 18. Número de vocablos más disponibles.	169
Tabla 19. Índices generales de DL según la variable “sexo”.	172
Tabla 20. Índices de DL según la variante “mujer”.	172
Tabla 21. Índices de DL según la variante “hombre”.....	173
Tabla 22. Estadísticos descriptivos para la variable “sexo”.	175
Tabla 23. Índice de cohesión por CI según la variable “sexo”.....	176

Tabla 24. Densidad léxica por CI según la variable “sexo”.	177
Tabla 25. Índices generales de DL según la variable “nivel de ELE”.	179
Tabla 26. Índices de DL según la variante “nivel B1”.	180
Tabla 27. Índices de DL según la variante “nivel B2”.	180
Tabla 28. Estadísticos descriptivos para la variable “nivel de ELE”.	182
Tabla 29. Índice de cohesión y densidad léxica por CI según la variable “nivel de ELE”.	184
Tabla 30. Índices generales de DL según la variable “conocimiento de otras LE”.	186
Tabla 31. Índices de DL según la variante “LE =2”.	187
Tabla 32. Índices de DL según la variante “LE >2”.	187
Tabla 33. Estadísticos descriptivos para la variable “conocimiento de otras LE”.	189
Tabla 34. Índice de cohesión por CI según la variable “conocimiento de otras LE”.	190
Tabla 35. Densidad léxica por CI según la variable “conocimiento de otras LE”.	191
Tabla 36. Índices generales de DL según el análisis longitudinal.	194
Tabla 37. Número de palabras por CI según el análisis longitudinal.	195
Tabla 38. CI ordenados en función de su productividad léxica según el análisis longitudinal.	196
Tabla 39. Promedio de palabras por informante por CI según el análisis longitudinal.	197
Tabla 40. Estadísticos descriptivos de las dos suministraciones de la prueba.	198
Tabla 41. Número de vocablos por CI según el análisis longitudinal.	199
Tabla 42. CI ordenados en función de su variedad léxica según el análisis longitudinal.	200
Tabla 43. Promedio de vocablos por informante por CI según el análisis longitudinal.	201
Tabla 44. Índice de cohesión y densidad léxica por CI según el análisis longitudinal.	202
Tabla 45. Datos generales de DL de las investigaciones cotejadas.	206
Tabla 46. Promedios de palabras de las investigaciones cotejadas.	207
Tabla 47. Estadísticos descriptivos de las investigaciones cotejadas.	208
Tabla 48. Promedios de vocablos de las investigaciones cotejadas.	209

Tabla 49. Índice de cohesión de las investigaciones cotejadas.	211
Tabla 50. Vocablos más disponibles en el CI01.....	217
Tabla 51. Vocablos más disponibles en el CI02.....	218
Tabla 52. Vocablos más disponibles en el CI03.....	218
Tabla 53. Vocablos más disponibles en el CI04.....	219
Tabla 54. Vocablos más disponibles en el CI05.....	220
Tabla 55. Vocablos más disponibles en el CI06.....	221
Tabla 56. Vocablos más disponibles en el CI07.....	221
Tabla 57. Vocablos más disponibles en el CI08.....	222
Tabla 58. Vocablos más disponibles en el CI09.....	223
Tabla 59. Vocablos más disponibles en el CI10.....	224
Tabla 60. Vocablos más disponibles en el CI11.....	225
Tabla 61. Vocablos más disponibles en el CI12.....	226
Tabla 62. Vocablos más disponibles en el CI13.....	226
Tabla 63. Vocablos más disponibles en el CI14.....	227
Tabla 64. Vocablos más disponibles en el CI15.....	228
Tabla 65. Vocablos más disponibles en el CI16.....	229
Tabla 66. Número de vocablos más disponibles por CI.....	230
Tabla 67. Vocablos más disponibles del CI01 según la variable “sexo”.	231
Tabla 68. Vocablos más disponibles del CI02 según la variable “sexo”.	232
Tabla 69. Vocablos más disponibles del CI03 según la variable “sexo”.	233
Tabla 70. Vocablos más disponibles del CI04 según la variable “sexo”.	233
Tabla 71. Vocablos más disponibles del CI05 según la variable “sexo”.	234
Tabla 72. Vocablos más disponibles del CI06 según la variable “sexo”.	235
Tabla 73. Vocablos más disponibles del CI07 según la variable “sexo”.	235
Tabla 74. Vocablos más disponibles del CI08 según la variable “sexo”.	236
Tabla 75. Vocablos más disponibles del CI09 según la variable “sexo”.	236
Tabla 76. Vocablos más disponibles del CI10 según la variable “sexo”.	237

Tabla 77. Vocablos más disponibles del CI11 según la variable “sexo”.	238
Tabla 78. Vocablos más disponibles del CI12 según la variable “sexo”.	238
Tabla 79. Vocablos más disponibles del CI13 según la variable “sexo”.	238
Tabla 80. Vocablos más disponibles del CI14 según la variable “sexo”.	239
Tabla 81. Vocablos más disponibles del CI15 según la variable “sexo”.	240
Tabla 82. Vocablos más disponibles del CI16 según la variable “sexo”.	240
Tabla 83. Cardinalidad del conjunto por CI según la variable “sexo”.	241
Tabla 84. Vocablos más disponibles del CI01 según la variable “nivel de ELE”.....	243
Tabla 85. Vocablos más disponibles del CI02 según la variable “nivel de ELE”.....	243
Tabla 86. Vocablos más disponibles del CI03 según la variable “nivel de ELE”.....	244
Tabla 87. Vocablos más disponibles del CI04 según la variable “nivel de ELE”.....	244
Tabla 88. Vocablos más disponibles del CI05 según la variable “nivel de ELE”.....	245
Tabla 89. Vocablos más disponibles del CI06 según la variable “nivel de ELE”.....	246
Tabla 90. Vocablos más disponibles del CI07 según la variable “nivel de ELE”.....	246
Tabla 91. Vocablos más disponibles del CI08 según la variable “nivel de ELE”.....	247
Tabla 92. Vocablos más disponibles del CI09 según la variable “nivel de ELE”.....	247
Tabla 93. Vocablos más disponibles del CI10 según la variable “nivel de ELE”.....	248
Tabla 94. Vocablos más disponibles del CI11 según la variable “nivel de ELE”.....	249
Tabla 95. Vocablos más disponibles del CI12 según la variable “nivel de ELE”.....	249
Tabla 96. Vocablos más disponibles del CI13 según la variable “nivel de ELE”.....	249
Tabla 97. Vocablos más disponibles del CI14 según la variable “nivel de ELE”.....	250
Tabla 98. Vocablos más disponibles del CI15 según la variable “nivel de ELE”.....	251
Tabla 99. Vocablos más disponibles del CI16 según la variable “nivel de ELE”.....	251
Tabla 100. Cardinalidad del conjunto por CI según la variable “nivel de ELE”.....	252
Tabla 101. Vocablos más disponibles del CI01 según la variable “conocimiento de otras LE”.....	253
Tabla 102. Vocablos más disponibles del CI02 según la variable “conocimiento de otras LE”.....	254

Tabla 103. Vocablos más disponibles del CI03 según la variable “conocimiento de otras LE”.....	254
Tabla 104. Vocablos más disponibles del CI04 según la variable “conocimiento de otras LE”.....	255
Tabla 105. Vocablos más disponibles del CI05 según la variable “conocimiento de otras LE”.....	256
Tabla 106. Vocablos más disponibles del CI06 según la variable “conocimiento de otras LE”.....	256
Tabla 107. Vocablos más disponibles del CI07 según la variable “conocimiento de otras LE”.....	257
Tabla 108. Vocablos más disponibles del CI08 según la variable “conocimiento de otras LE”.....	257
Tabla 109. Vocablos más disponibles del CI09 según la variable “conocimiento de otras LE”.....	258
Tabla 110. Vocablos más disponibles del CI10 según la variable “conocimiento de otras LE”.....	258
Tabla 111. Vocablos más disponibles del CI11 según la variable “conocimiento de otras LE”.....	259
Tabla 112. Vocablos más disponibles del CI12 según la variable “conocimiento de otras LE”.....	259
Tabla 113. Vocablos más disponibles del CI13 según la variable “conocimiento de otras LE”.....	260
Tabla 114. Vocablos más disponibles del CI14 según la variable “conocimiento de otras LE”.....	260
Tabla 115. Vocablos más disponibles del CI15 según la variable “conocimiento de otras LE”.....	261
Tabla 116. Vocablos más disponibles del CI16 según la variable “conocimiento de otras LE”.....	262
Tabla 117. Cardinalidad del conjunto por CI según la variable “conocimiento de otras LE”.....	263

Tabla 118. Vocablos más disponibles del CI01 según la fecha de la prueba.	265
Tabla 119. Vocablos más disponibles del CI02 según la fecha de la prueba.	266
Tabla 120. Vocablos más disponibles del CI03 según la fecha de la prueba.	266
Tabla 121. Vocablos más disponibles del CI04 según la fecha de la prueba.	267
Tabla 122. Vocablos más disponibles del CI05 según la fecha de la prueba.	268
Tabla 123. Vocablos más disponibles del CI06 según la fecha de la prueba.	268
Tabla 124. Vocablos más disponibles del CI07 según la fecha de la prueba.	269
Tabla 125. Vocablos más disponibles del CI08 según la fecha de la prueba.	270
Tabla 126. Vocablos más disponibles del CI09 según la fecha de la prueba.	270
Tabla 127. Vocablos más disponibles del CI10 según la fecha de la prueba.	271
Tabla 128. Vocablos más disponibles del CI11 según la fecha de la prueba.	272
Tabla 129. Vocablos más disponibles del CI12 según la fecha de la prueba.	272
Tabla 130. Vocablos más disponibles del CI13 según la fecha de la prueba.	273
Tabla 131. Vocablos más disponibles del CI14 según la fecha de la prueba.	274
Tabla 132. Vocablos más disponibles del CI15 según la fecha de la prueba.	274
Tabla 133. Vocablos más disponibles del CI16 según la fecha de la prueba.	275
Tabla 134. Cardinalidad del conjunto por CI según la fecha de la prueba.	276
Tabla 135. Número de vocablos compartidos con el CREA.	278
Tabla 136. Número de vocablos compartidos con el CAES.	280
Tabla 137. Datos generales de RL.	285
Tabla 138. Índices de RL.	288
Tabla 139. Datos generales de RL según la variable “sexo”.	293
Tabla 140. Índices de RL según la variante “mujer”.	296
Tabla 141. Índices de RL según la variante “hombre”.	296
Tabla 142. Estadísticos descriptivos del TTR para la variable “sexo”.	297
Tabla 143. Estadísticos descriptivos de la variación léxica para la variable “sexo”. ...	298
Tabla 144. Estadísticos descriptivos del índice de hápax para la variable “sexo”.	299
Tabla 145. Estadísticos descriptivos de la densidad léxica para la variable “sexo”.	300

Tabla 146. Estadísticos descriptivos del IAT para la variable “sexo”.....	301
Tabla 147. Datos generales de RL según la variable “nivel de ELE”.....	302
Tabla 148. Índices de RL según la variante “nivel B1”.....	304
Tabla 149. Índices de RL según la variante “nivel B2”.....	305
Tabla 150. Estadísticos descriptivos del TTR para la variable “nivel de ELE”.....	306
Tabla 151. Estadísticos descriptivos de la variación léxica para la variable “nivel de ELE”.....	307
Tabla 152. Estadísticos descriptivos del índice de hápax para la variable “nivel de ELE”.....	307
Tabla 153. Estadísticos descriptivos de la densidad léxica para la variable “nivel de ELE”.....	308
Tabla 154. Estadísticos descriptivos del IAT para la variable “nivel de ELE”.....	309
Tabla 155. Datos generales de RL según la variable “conocimiento de otras LE”.....	311
Tabla 156. Índices de RL según la variante “LE =2”.....	312
Tabla 157. Índices de RL según la variante “LE >2”.....	314
Tabla 158. Estadísticos descriptivos del TTR para la variable “conocimiento de otras LE”.....	314
Tabla 159. Estadísticos descriptivos de la variación léxica para la variable “conocimiento de otras LE”.....	315
Tabla 160. Estadísticos descriptivos del índice de hápax para la variable “conocimiento de otras LE”.....	316
Tabla 161. Estadísticos descriptivos de la densidad léxica para la variable “conocimiento de otras LE”.....	317
Tabla 162. Estadísticos descriptivos del IAT para la variable “conocimiento de otras LE”.....	318
Tabla 163. Mejores resultados de RL del análisis transversal.....	320
Tabla 164. Promedios de RL del análisis transversal.....	320
Tabla 165. Datos generales de RL según la fecha de la prueba.....	322
Tabla 166. Índices de RL en la I administración de la prueba.....	324

Tabla 167. Índices de RL en la II administración de la prueba.	325
Tabla 168. Estadísticos descriptivos del TTR según la fecha de la prueba.	326
Tabla 169. Estadísticos descriptivos de la variación léxica según la fecha de la prueba.	327
Tabla 170. Estadísticos descriptivos del índice de hápax según la fecha de la prueba.	328
Tabla 171. Estadísticos descriptivos de la densidad léxica según la fecha de la prueba.	329
Tabla 172. Estadísticos descriptivos del IAT según la fecha de la prueba.	330
Tabla 173. Datos generales de RL de las investigaciones cotejadas.	332
Tabla 174. Promedios del TTR de las investigaciones cotejadas.	333
Tabla 175. Promedios del IAT de las investigaciones cotejadas.	334
Tabla 176. Lista de las 25 palabras más frecuentes del corpus general de RL.	342
Tabla 177. Lista de las 25 palabras nocionales más frecuentes del corpus general de RL.	343
Tabla 178. Lista de los 25 sustantivos más frecuentes del corpus general de RL.	344
Tabla 179. Lista de los 25 verbos más frecuentes del corpus general de RL.	346
Tabla 180. Lista de los 25 adjetivos más frecuentes del corpus general de RL.	347
Tabla 181. Lista de los 25 adverbios más frecuentes del corpus general de RL.	348
Tabla 182. Lista de las palabras clave del corpus general de RL.	349
Gráfico 1. Número total de palabras.	160
Gráfico 2. Porcentaje de palabras actualizadas por CI.	160
Gráfico 3. Promedio de palabras por informante.	162
Gráfico 4. Número total de vocablos.	163
Gráfico 5. Porcentaje de vocablos actualizados por CI.	164
Gráfico 6. Índice de cohesión por CI.	167
Gráfico 7. Densidad léxica por CI.	167
Gráfico 8. Promedio de palabras según la variable “sexo”.	174
Gráfico 9. Diagrama de cajas para la variable “sexo”.	175

Gráfico 10. Promedio de vocablos según la variable “sexo”.	175
Gráfico 11. Índice de cohesión por CI según la variable “sexo”.	177
Gráfico 12. Promedio de palabras según la variable “nivel de ELE”.	181
Gráfico 13. Diagrama de cajas para la variable “nivel de ELE”.	182
Gráfico 14. Promedio de vocablos según la variable “nivel de ELE”.	183
Gráfico 15. Índice de cohesión según la variable “nivel de ELE”.	184
Gráfico 16. Densidad léxica según la variable “nivel de ELE”.	185
Gráfico 17. Promedio de palabras según la variable “conocimiento de otras LE”.	188
Gráfico 18. Diagrama de cajas para la variable “conocimiento de otras LE”.	189
Gráfico 19. Promedio de vocablos según la variable “conocimiento de otras LE”.	189
Gráfico 20. Índice de cohesión según la variable “conocimiento de otras LE”.	191
Gráfico 21. Número total de palabras según la fecha de la prueba.	196
Gráfico 22. Promedio de palabras por informante según la fecha de la prueba.	198
Gráfico 23. Diagrama de cajas según la fecha de la prueba.	198
Gráfico 24. Número total de vocablos según la fecha de la prueba.	200
Gráfico 25. Promedio de vocablos por informante según la fecha de la prueba.	201
Gráfico 26. Índice de cohesión según la fecha de la prueba.	203
Gráfico 27. Densidad léxica según la fecha de la prueba.	203
Gráfico 28. Promedios de palabras por informante en las investigaciones cotejadas.	207
Gráfico 29. Diagrama de cajas de las investigaciones cotejadas.	208
Gráfico 30. Promedios de vocablos por informante en las investigaciones cotejadas.	210
Gráfico 31. Número de vocablos más disponibles por CI.	230
Gráfico 32. Cardinalidad del conjunto por CI según la variable “sexo”.	242
Gráfico 33. Cardinalidad del conjunto por CI según la variable “nivel de ELE”.	252
Gráfico 34. Cardinalidad del conjunto por CI según la variable “conocimiento de otras LE”.	263
Gráfico 35. Cardinalidad del conjunto por CI según la fecha de la prueba.	276
Gráfico 36. Categorías gramaticales de los vocablos compartidos con el CREA.	279

Gráfico 37. Categorías gramaticales de los vocablos compartidos con el CAES.	280
Gráfico 38. Distribución de los textos según el promedio del TTR.	289
Gráfico 39. Distribución de los textos según el promedio de la variación léxica.	289
Gráfico 40. Distribución de los textos según el promedio del índice de hápax.	290
Gráfico 41. Distribución de los textos según el promedio de la densidad léxica.	291
Gráfico 42. Distribución de los textos según el promedio del IAT.	292
Gráfico 43. Diagrama de cajas del TTR para la variable “sexo”.	297
Gráfico 44. Diagrama de cajas de la variación léxica para la variable “sexo”.	298
Gráfico 45. Diagrama de cajas del índice de hápax para la variable “sexo”.	299
Gráfico 46. Diagrama de cajas de la densidad léxica para la variable “sexo”.	300
Gráfico 47. Diagrama de cajas del IAT para la variable “sexo”.	301
Gráfico 48. Diagrama de cajas del TTR para la variable “nivel de ELE”.	306
Gráfico 49. Diagrama de cajas de la variación léxica para la variable “nivel de ELE”.	307
Gráfico 50. Diagrama de cajas del índice de hápax para la variable “nivel de ELE”. .	308
Gráfico 51. Diagrama de cajas de la densidad léxica para la variable “nivel de ELE”. .	309
Gráfico 52. Diagrama de cajas del IAT para la variable “nivel de ELE”.	310
Gráfico 53. Diagrama de cajas del TTR para la variable “conocimiento de otras LE”. .	315
Gráfico 54. Diagrama de cajas de la variación léxica para la variable “conocimiento de otras LE”.	316
Gráfico 55. Diagrama de cajas del índice de hápax para la variable “conocimiento de otras LE”.	317
Gráfico 56. Diagrama de cajas de la densidad léxica para la variable “conocimiento de otras LE”.	318
Gráfico 57. Diagrama de cajas del IAT para la variable “conocimiento de otras LE”. .	319
Gráfico 58. Diagrama de cajas del TTR según la fecha de la prueba.	326
Gráfico 59. Diagrama de cajas de la variación léxica según la fecha de la prueba.	327
Gráfico 60. Diagrama de cajas del índice de hápax según la fecha de la prueba.	328
Gráfico 61. Diagrama de cajas de la densidad léxica según la fecha de la prueba.	329

Gráfico 62. Diagrama de cajas del IAT según la fecha de la prueba.	330
Gráfico 63. Promedios del TTR de las investigaciones cotejadas.....	333
Gráfico 64. Promedios del IAT de las investigaciones cotejadas.....	334
Gráfico 65. Distribución de las palabras en la lista de frecuencia del corpus general de RL.	341
Gráfico 66. Distribución de las palabras nocionales en el corpus general de RL.	342
Figura 1. Diagrama de la enseñanza centrada en el alumno.....	44
Figura 2. Clasificación de las unidades léxicas.	81
Figura 3. Uso del vocablo <i>vacación</i> en el corpus general de RL.	350
Figura 4. Uso del vocablo <i>visitar</i> en el corpus general de RL.....	351
Figura 5. Uso del vocablo <i>estupendo</i> en el corpus general de RL.....	351
Figura 6. Uso del vocablo <i>allí</i> en el corpus general de RL.	352

SIGLAS Y ACRÓNIMOS UTILIZADOS

- ASL: Adquisición de segundas lenguas
- CAES: Corpus de aprendices de español
- CI: Centro de interés
- CREA: Corpus de referencia del español actual
- DL: Disponibilidad léxica
- DRAE: Diccionario de la lengua española
- ELE: Español lengua extranjera
- ESO: Educación Secundaria Obligatoria
- IAT: Intervalo de aparición de palabras nocionales
- ID: Índice de disponibilidad léxica
- L1: Primera lengua
- L2: Segunda lengua
- L3: Tercera lengua
- LE: Lengua extranjera
- LFP: Perfil de Frecuencia Léxica
- LM: Lengua materna
- LO: Lengua objeto
- MCER: Marco común europeo de referencia para las lenguas
- PCIC: Plan curricular del Instituto Cervantes
- PPDHL: Proyecto Panhispánico de Disponibilidad Léxica
- RAE: Real Academia Española
- RL: Riqueza léxica
- TTR: *Type/Token Ratio*
- UD: Unidad didáctica

SINOSI

Inserita negli ambiti della linguistica applicata e della didattica della lingua spagnola, la presente Tesi esamina la disponibilità e la ricchezza lessicale di un gruppo di studenti italiani di spagnolo LS a livello universitario.

Partendo dai primi studi finalizzati all'elaborazione del *Proyecto Panhispánico de Disponibilidad Léxica* dedicati a ispanofoni nativi, dai progetti pionieristici di Carcedo (2000c) e Samper Hernández (2002) che si sono occupati di apprendenti non nativi di spagnolo ed infine dai contributi più recenti di Sánchez-Saus (2011), Hidalgo (2017b) e Del Barrio y Vann (2018), è stato elaborato un test che ha permesso di svolgere un'indagine linguistico-statistica trasversale e longitudinale del lessico disponibile del campione. Suddetto consisteva nella prova standardizzata di disponibilità lessicale, in cui si presentano alcuni stimoli semantici per attivare i vocaboli che i soggetti intervistati associano a tali *input*. Inoltre, nello stesso test è stata inserita un'ulteriore prova destinata alla misurazione e alla valutazione della ricchezza lessicale, dove era richiesta la stesura di un breve testo narrativo che è stato analizzato in base a indici standard (Read, 2000).

L'esperimento, proponendosi come un progetto innovativo che tratta il complesso fenomeno della competenza lessicale, ha conciliato due metodologie complementari capaci di determinare il livello di apprendimento e la conoscenza del vocabolario spagnolo dei partecipanti.

I risultati emersi dalle analisi quantitativa, qualitativa e comparativa dei dati hanno condotto a nuove ed interessanti riflessioni derivanti dall'unione della disponibilità e della ricchezza lessicale per la prima volta in un unico studio.

In più, gli approcci trasversale e longitudinale hanno portato a rinnovare il *modus operandi* di ricerca applicato finora in questo ambito, improntato solamente su studi trasversali.

ABSTRACT

This thesis, framed within the areas of Applied Linguistics and Didactics of Spanish, investigates lexical availability and lexical richness in a group of Italian university students learning Spanish as a FL.

The theoretical framework consists of *Proyecto Panhispánico de Disponibilidad Léxica*, involving native Spanish speakers, pioneering works of Carcedo (2000c) and Samper Hernández (2002) that dealt with non-native Spanish learners and the latest findings of Sánchez-Saus (2011), Hidalgo (2017b) and Del Barrio y Vann (2018).

It was designed an associative test which allowed to conduct a linguistic-statistical survey on the available lexicon of informants through a cross-sectional and longitudinal analysis. This test focuses on the vocabulary produced by respondents when prompted with stimuli and collects their mental lexicon. In addition, it was included a further test for measuring and evaluating lexical richness, which required the writing of a short narrative text that was examined on the basis of standard indexes (Read, 2000).

The experiment aims to be an innovative project which deals with the complex phenomenon of lexical competence, combining two complementary methodologies capable of ascertaining the participants' language proficiency and knowledge of Spanish vocabulary.

Results of the quantitative, qualitative and comparative analysis of data led to significant new considerations arising from the combination of lexical availability and lexical richness merged for the first time in a single study.

Moreover, the transverse and longitudinal approaches brought about a renewal of the research method in this field, which so far was solely based on cross-sectional studies.

INTRODUCCIÓN

Esta Tesis se enmarca en el ámbito de los estudios sobre el aprendizaje del léxico en español lengua extranjera (ELE) y tiene como objetivos prioritarios la medición y la evaluación de la competencia léxica de un grupo de aprendices itálofonos de nivel universitario.

En primera instancia, la elección de este tema se debe al interés que en el panorama investigativo actual despierta la *competencia léxica*, que asumimos como el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo (MCER, 2002: 108). Dicha competencia se sitúa dentro de la más amplia competencia lingüística y posibilita el desarrollo de una eficiente destreza comunicativa, que coincide con el propósito último del aprendizaje de una lengua extranjera: comunicar con éxito. Análogamente, concordamos con la literatura sobre un concepto de competencia léxica dinámico, procesual y capaz de determinar la capacidad de uso de las palabras de cada individuo.

En este sentido, la principal aportación de nuestra investigación pretende mejorar nuestro conocimiento de cómo se construye y cómo funciona la competencia léxica en un conjunto de discentes italianos de ELE. Para ello, hemos abordado el asunto de una manera novedosa, ya que conjuga dos métodos experimentales encaminados a medir y valorar el vocabulario que el aprendiz sabe y, por tanto, puede utilizar en una situación comunicativa dada. El primer método consiste en el análisis de la *disponibilidad léxica*, el cual se refiere al cómputo de palabras que un sujeto es capaz de actualizar ante un estímulo temático dado; el segundo se denomina análisis de la *riqueza léxica* y contabiliza la variedad del vocabulario utilizado en un texto, así como su densidad, esto es, el empleo de las unidades funcionales frente a la presencia de palabras nocionales.

En esta Tesis se plantea que la correlación entre estos dos métodos posibilita un conocimiento más profundo de la competencia léxica, ya que la perspectiva de la disponibilidad se coloca en el plano paradigmático del léxico y permite analizar no solo de qué palabras dispone un aprendiz, sino cómo se organizan conceptualmente en su lexicón en un momento dado. Por su parte, la perspectiva de la riqueza se ubica en el

plano sintagmático y permite observar cuántas palabras distintas el aprendiz utiliza en un texto.

Efectivamente, con ambas metodologías se han diseñado numerosos experimentos cuyos resultados consisten fundamentalmente en tres tipos de productos: (i) diccionarios de léxico disponible, elaborados a partir del conjunto de unidades léxicas que los participantes actualizan en los tests; (ii) análisis cuantitativos y estadísticos realizados a partir de índices que miden la cantidad de unidades léxicas que el grupo produce; y, por último, (iii) estudios de corte cualitativo sobre el tipo de léxico activado y utilizado, y sobre la naturaleza de las relaciones conceptuales que gobiernan el vocabulario de los informantes.

La literatura ha confirmado ampliamente la indudable utilidad de estos análisis para el conocimiento de la competencia léxica, subrayando los importantes resultados alcanzados, especialmente en la última década, pero a la vez poniendo de manifiesto algunas cuestiones críticas que hasta ahora no han sido adecuadamente resueltas y a las que pretendemos dar respuesta. En primer lugar, aplicamos las medidas de disponibilidad y riqueza léxica al estudio del bagaje léxico de los informantes, algo que ninguno de los proyectos de esta índole ha realizado hasta ahora. De hecho, uno de los puntos fuertes de la Tesis es la capacidad de reunir bajo un mismo marco metodológico procedimientos distintos para lograr una mejor comprensión de un fenómeno complejo como es el de la competencia léxica, proporcionando un instrumento de análisis empírico innovador para la comunidad científica.

En segundo lugar, hemos realizado la recogida de datos en dos momentos distintos para la misma muestra de estudiantes, lo cual nos ha permitido observar la evolución del vocabulario disponible y activo de los informantes tras la asistencia a un curso de español lengua extranjera de un año. Este tipo de análisis, del que no tenemos constancia que se haya realizado hasta ahora, pretende obtener las claves necesarias para observar la posible evolución temporal en la adquisición del léxico y, por tanto, la efectividad de la acción didáctica.

Asimismo, y por último, hemos cotejado nuestros datos con los obtenidos en otros trabajos desarrollados con aprendientes de ELE y con dos corpus de referencia para el español (el *Corpus de Referencia del Español Actual* y el *Corpus de aprendices de español*).

En definitiva, la originalidad de esta investigación consiste en conciliar las pautas metodológicas para la obtención del léxico disponible con las de la riqueza léxica en un estudio de tipo transversal y longitudinal: nuestra intención es estudiar la competencia léxica desde una perspectiva que sea lo más aglutinadora posible. Hemos programado, por esta razón, los siguientes objetivos investigadores:

- (i) analizar cuantitativa y cualitativamente el léxico disponible de una muestra suficiente de estudiantes universitarios italófonos de ELE;
- (ii) analizar cuantitativa y cualitativamente el léxico activo utilizado en textos escritos por dicha muestra;
- (iii) evaluar la incidencia de tres factores sociolingüísticos (sexo, nivel de ELE, conocimiento de otras lenguas extranjeras) en la competencia léxica del grupo;
- (iv) analizar la evolución en el tiempo de la disponibilidad léxica y del léxico activo;
- (v) comparar nuestros datos con los obtenidos en otros proyectos dedicados a estudiantes de ELE con objeto de determinar diferencias y carencias en función de la nacionalidad y del entorno educativo;
- (vi) elaborar un diccionario del léxico disponible de los aprendices italianos universitarios de ELE.

Todo esto debería permitir corroborar una serie de hipótesis didácticas, psicolingüísticas y sociolingüísticas:

- La competencia léxica aumenta según avance el nivel de español, los estudiantes más avanzados presentarán una mayor productividad y riqueza léxicas, tanto en la activación de léxico disponible como en la producción escrita. Igualmente, el rendimiento será mejor al final del año académico.
- Existe una relación asociativa entre el nivel de ELE y la cantidad de lexías aprendidas, a saber, activadas y utilizadas en las dos partes del test de medición al que se ha sometido el alumnado.
- De entre los factores sociolingüísticos, (i) el sexo no influirá en la producción de los informantes, ni desde la perspectiva cuantitativa ni de la cualitativa; (ii) el nivel de ELE estará correlacionado con la competencia léxica del grupo ya que se espera una evolución de un nivel a otro, la cual implica que la disponibilidad y la riqueza léxica de los estudiantes de nivel más avanzado

sean mayores tanto en cantidad de palabras como de vocablos con respecto a los que tienen un menor dominio; (iii) los participantes que conocen o estudian más lenguas conseguirán mejores resultados porque ya tienen desarrollado un buen sistema lingüístico: cuantos más idiomas extranjeros conozcan, además del español, mayor será la competencia léxica, aunque se detectará la influencia de dichas lenguas, en particular del inglés al ser esta la primera lengua extranjera de estudio.

- En lo referente a la edición de los datos, suponemos que (i) las interferencias procedentes del italiano serán frecuentes debido a la afinidad lingüística con el español; (ii) se detectarán errores de ortografía; (iii) aparecerán palabras de la sintopía hispanoamericana y del sociolecto coloquial debido a la programación curricular de los cursos que prevén el conocimiento de los rasgos lingüísticos peculiares de dichas variedades¹.
- El análisis comparativo demostrará que la competencia léxica de nuestros informantes es mayor que la de otros aprendices de distinta procedencia debido a la afinidad lingüística entre lengua materna y lengua extranjera, aunque existirá un desfase entre las listas de léxico disponible de los que hayan estudiado en contexto de inmersión (que deberían poseer un conocimiento superior).
- En lo que atañe al análisis del léxico disponible, los centros de interés producirán un mayor número de palabras y de vocablos en función de su rentabilidad para el alumnado. Algunos de estos estímulos se revelan poco apropiados para cumplir con las necesidades de un aprendiz del siglo XXI, como ya han planteado otros autores.
- Teniendo en cuenta las investigaciones ya publicadas, los sustantivos serán la clase gramatical predominante y solo en menor medida aparecerán otras categorías morfosintácticas. Igualmente, aparecerá una escasa cantidad de unidades multipalabra.
- Considerado el nivel de ELE, los indicadores de riqueza léxica expresarán una apreciable variación de vocabulario gracias al desarrollo de una adecuada

¹ Véase, a este propósito, Contreras (2018) que trata con detalle la cuestión de la enseñanza de las variedades diatópicas del español, con particular atención al lenguaje coloquial.

competencia léxica que se repercute positivamente en la competencia comunicativa de los estudiantes.

- Discentes no nativos que comparten el dominio lingüístico, pero de distinta procedencia, poseen iguales capacidades expresivas y, a más y mejor, tienen desarrollada la misma competencia en ELE.
- La correlación de los datos procedentes de los análisis de disponibilidad léxica, basados en parámetros de tipo paradigmático, y de riqueza léxica, basados en parámetros de tipo sintagmático, posibilitarán un pormenorizado estudio de la competencia léxica de los aprendices itálofonos de ELE.

En sintonía con los planteamientos expuestos, esta Tesis se estructura en dos partes, constituidas por siete capítulos encaminados a la consecución de los objetivos y la comprobación de las hipótesis que nos marcamos en principio.

La primera parte está dedicada a los cimientos del trabajo, los patrones teóricos y metodológicos que rigen nuestro estudio.

El Capítulo I sirve de marco para la investigación, resume las teorías sobre la adquisición de las lenguas extranjeras a raíz de los postulados de la lingüística aplicada y de otras disciplinas que han contribuido al desarrollo de las tendencias actuales, según un recorrido historiográfico que parte de los años cincuenta del siglo pasado hasta la fecha. Dichas teorías han evolucionado de manera sobresaliente a partir de las premisas del conductismo, pasando por una lectura innatista del aprendizaje lingüístico, hasta llegar al cognitivismo y al interaccionismo, que lo describen tanto desde una perspectiva individual como social, centrándose en la finalidad comunicativa del lenguaje (Hymes, 1972). Es más, la tendencia hacia lo psicológico que se ha sumado a los planteamientos de la lingüística ha llevado a dedicar mucha atención a los procesos cognitivos que subyacen a la adquisición de una lengua (recordamos los estudios dedicados al lexicón mental).

Todo esto ha influido en la enseñanza de las lenguas extranjeras por lo que, a continuación, se ha llevado a cabo un repaso histórico de los métodos y enfoques desarrollados a partir de las citadas teorías, centrado en el tratamiento que se ha dedicado al léxico, eje central de nuestro trabajo. Hemos repartido este espacio temporal en tres etapas según la atención que se otorga al vocabulario: (i) las listas de palabras; (ii) la revalorización del léxico; (iii) el protagonismo de sus aspectos cualitativos, aun vigente

en la actualidad (Morante Vallejo, 2005).

A la luz de los últimos hallazgos, hemos estudiado la repercusión de sus postulados en la práctica didáctica diaria. Para ello revisamos algunos manuales de ELE que aplican los principios de los métodos y enfoques mencionados arriba. Por último, hacemos hincapié en el uso de otro recurso esencial para el aprendizaje de una lengua y de su vocabulario, el diccionario.

Tras la revisión historiográfica, a raíz de la importancia que ha cobrado el componente léxico, hemos destinado un apartado a la selección del vocabulario que ocupa (o debería ocupar) una parte importante de la tarea docente, de los investigadores y de los creadores de materiales. Se han mostrado los criterios que hasta la fecha han guiado los procesos para una adecuada elección encaminada al aprendizaje lingüístico (determinar cuál y cuánto vocabulario enseñar). Este primer capítulo se cierra con una breve sección dedicada a la competencia léxica.

El Capítulo II desarrolla el marco teórico. En primer lugar, fijamos los conceptos fundamentales que se han de tener en consideración para la comprensión del trabajo empírico descrito a lo largo de la siguiente parte de la Tesis consistente en el análisis cuantitativo y cualitativo de la disponibilidad y de la riqueza léxica. Se trata de una serie de conceptos básicos como la diferencia entre palabra y vocablo y la taxonomía léxica. En segundo lugar, expone el estado de la cuestión en lo que se refiere a los orígenes y el desarrollo actual de los estudios sobre disponibilidad y riqueza léxica.

A partir de la segunda mitad del siglo XX, en Francia, se sentaron las bases de una metodología que solucionase las limitaciones del criterio de frecuencia y posibilite una adecuada selección del léxico con fines pedagógicos. Se elabora el concepto de *disponibilidad léxica* y se establecen los principios metodológicos que permiten recoger datos auténticos para elaborar listas de unidades léxicas definidas *disponibles*, útiles para la enseñanza de la lengua extranjera. Así nace el *Français Fondamental* en 1964 (Gougenheim *et al.*). Esta metodología se funda en pruebas asociativas diseñadas para contabilizar la cantidad de palabras que una persona conoce y tiene mentalmente a su disposición en relación con un estímulo temático (denominado *centro de interés*). Se asume que los ítems aportados en estas pruebas pertenecen al léxico activable en una situación comunicativa concreta y se deben enseñar en las primeras etapas de aprendizaje por ser muy rentables.

En lo que se refiere al español, el punto de inflexión lo marca el *Proyecto Panhispánico de Disponibilidad Léxica*, iniciado por López Morales en 1999, cuyos planteamientos han interesado a numerosos investigadores que han propuesto estudios sobre el habla española. Si bien estos primeros trabajos estaban dedicados a nativos, rápidamente se percibe el interés que pueden tener en la adquisición del español LE. Se empieza una nueva línea de investigación centrada en el léxico disponible de hablantes no nativos, con un fecundo desarrollo sobre todo en las últimas dos décadas en distintas partes del mundo, aunque, pese a los importantes resultados de tales proyectos, solo recientemente Italia se ha acercado a este campo.

Además de la didáctica de lenguas, la disponibilidad léxica cuenta con aplicaciones en otros ámbitos como la sociolingüística, la dialectología, la psicolingüística y la etnolingüística, en los que tiene un vasto abanico de posibilidades teóricas y prácticas que abren puertas a aspectos inéditos o todavía no tratados con profundidad.

Por otro lado, la *riqueza léxica* es una metodología que ha sido definida y utilizada de forma diferente por varios autores, pero que, en general, sirve para la medición del léxico activado y empleado en una muestra lingüística en virtud de su variedad y densidad. Los estudios pioneros han surgido a mediados del siglo XX en Francia para analizar textos producidos por hablantes nativos (Guiraud, 1954), cuyos postulados siguen siendo válidos.

Se parte de la distinción entre (i) *tokens* y *types*, (ii) palabras semánticas y gramaticales y se aplican fórmulas matemáticas dirigidas a computar índices de naturaleza cuantitativa y comparativa que expresan el grado de riqueza léxica. La intención es cuantificar la variación y el tamaño del vocabulario mediante un *modus operandi* encaminado a la recogida de diferentes indicadores que concurren a establecer si un texto es “rico”. Se trata, por lo general, de las siguientes medidas (Read, 2000): (i) la densidad léxica; (ii) la variación léxica; (iii) la sofisticación léxica; (iv) la individualidad (o originalidad) léxica; (v) el número de errores.

López Morales (1984) es el primero que realiza un trabajo de riqueza léxica en ámbito hispano, introduciendo notables modificaciones al protocolo empírico. Ha reformulado las operaciones matemáticas y ha aportado nuevos indicadores, la *diversidad léxica* y el *intervalo de aparición de palabras de contenido nocional*. De ahí, se han

realizado múltiples estudios sobre la producción, escrita u oral, de hispanohablantes nativos.

Sin embargo, en lo que se refiere a hablantes no nativos del español, disponemos todavía de modestas contribuciones en este marco.

Como la disponibilidad, también la riqueza léxica puede ser utilizada en distintas áreas, además de la lingüística y de la didáctica de lenguas materna y extranjeras (especialmente en los procedimientos de evaluación, en la programación didáctica, en la selección y graduación del *input*). En particular, se aprecia el interés de diversos trabajos de corte sociolingüístico, cuyo objetivo es analizar la influencia de factores sociales, culturales y económicos en la habilidad expresiva de un grupo o de una comunidad de habla.

El Capítulo III está dedicado a los aspectos metodológicos. En concreto, describe (i) la selección y la estratificación de la muestra; (ii) la elección de los centros de interés y de los índices de medición; (iii) las técnicas de elicitación de los datos; (iv) los criterios de edición y el tratamiento –informático y estadístico– del material recogido.

La población estudiada es la de los estudiantes de ELE del *Dipartimento di Studi Linguistici e Letterari* de la Universidad de Padua, de la cual se ha extraído una muestra de cien sujetos, estratificados en función de tres variables: sexo (mujer u hombre), nivel de español (B1 o B2, a partir de los niveles del MCER) y conocimiento de otras lenguas extranjeras (dos o más de dos).

La selección de los centros de interés incluidos en el test de disponibilidad léxica sigue la del *Proyecto Panhispánico de Disponibilidad Léxica*, que a su vez corresponde a la propuesta elaborada en el *Français Fondamental* (Gougenheim *et al.*, 1964). Aunque cada estudio haya añadido otros centros, hemos optado por mantener los tradicionales tanto para garantizar la comparabilidad entre nuestros resultados y los de otros investigadores, como para evaluar la congruencia de estos campos semánticos con otros trabajos realizados, a pesar de ser conscientes de la relativa falta de adecuación de algunas de estas áreas en una investigación dedicada a informantes no nativos. La lista se compone de los siguientes centros (CI): CI01, “partes del cuerpo”; CI02, “la ropa”; CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI05, “alimentos y bebidas”; CI06, “objetos colocados en la mesa para la comida”; CI07, “la cocina y sus utensilios”; CI08, “la escuela: muebles y materiales”; CI09, “iluminación y calefacción”;

CI10, “la ciudad”; CI11, “el campo”; CI12, “medios de transporte”; CI13, “trabajos del campo y del jardín”; CI14, “los animales”; CI15, “juegos y distracciones”; CI16, “profesiones y oficios”.

Los indicadores que hemos utilizado para determinar el léxico disponible de los encuestados son el total de palabras y el total de vocablos en función de cada centro de interés y de las medias por informante. Asimismo, hemos medido dos índices que arrojan luz sobre el grado de concreción semántica de un campo: el *índice de cohesión* y la *densidad léxica*. En el análisis cualitativo hemos tenido en cuenta también la *cardinalidad del conjunto* (López Chávez, 1992) para averiguar el porcentaje de compatibilidad del vocabulario aportado entre las distintas agrupaciones.

Por su parte, las medidas que nos han permitido evaluar la riqueza léxica son: *variación léxica*, *Type/Token Ratio*, *índice de hápax*, *densidad léxica* e *intervalo de aparición de palabras nocionales*. Asimismo, hemos estudiado la lista de las palabras extraídas del corpus de textos para localizar qué unidades se emplean de manera más consistente y examinar la alternancia de ítems gramaticales y léxicos en la producción de los discentes. Esto nos ha permitido proponer un análisis cualitativo de los resultados.

La encuesta que hemos diseñado para la medición de la competencia léxica se divide en tres partes:

- (i) el cuestionario sociológico, encaminado a obtener informaciones personales sobre los participantes;
- (ii) la prueba de disponibilidad léxica, organizada según las pautas tradicionales;
- (iii) la prueba de riqueza léxica, consistente en la redacción de un texto escrito.

En lo referente al análisis del léxico disponible, seguimos los criterios de edición fijados para el *Proyecto Panhispánico* (Samper Padilla, 1998) y para trabajos sobre aprendices extranjeros de español (Samper Hernández, 2002), sometiendo el corpus de trabajo a una minuciosa edición según las prácticas convencionales que se emplean en la bibliografía sobre el estudio de la producción escrita en lengua extranjera. Por último, el tratamiento informático de los datos se ha realizado mediante los programas *Dispogen II* (Echeverría *et al.*, 2005), *AntConc* (Anthony, 2018) y *Sketch Engine* (Kilgarriff, *et al.*, 2004).

La segunda parte constituye el bloque central de la Tesis y consiste en el análisis de los datos sobre la disponibilidad y la riqueza léxica en ELE obtenidos en nuestra

muestra.

El apartado analítico propiamente dicho inicia en el Capítulo IV, que presenta los datos y la discusión de los resultados cuantitativos sobre el léxico disponible extraído de las cien pruebas realizadas. Como todos los trabajos de disponibilidad léxica, se empieza por los datos generales de la muestra: número total de palabras y número total de vocablos recogidos, media de palabras por sujeto y por centro de interés, índice de cohesión y densidad léxica.

La siguiente parte del análisis transversal se dedica a examinar la incidencia de las tres variables consideradas según el mismo protocolo, complementado por el estudio de los datos estadísticos descriptivos de cada condicionante para obtener información sobre el comportamiento intragrupal e intergrupalo de los encuestados, cuya significatividad resulta de gran interés a la hora de aplicar los resultados a la mejora de la enseñanza del español (adaptación de la selección de los contenidos, de la propuesta didáctica y de los materiales), en particular al considerar el nivel lingüístico y el conocimiento de otras lenguas extranjeras.

A continuación, el análisis longitudinal, llevado a cabo gracias a la segunda aplicación de la encuesta a los participantes de nivel B2 (otras cincuenta pruebas), nos ha permitido observar la evolución en el tiempo de la competencia léxica y deducir en qué medida los procesos didácticos inciden en la adquisición del vocabulario.

Por último, hemos cotejado nuestros resultados generales con los de otras investigaciones realizadas con aprendices de ELE de idéntica competencia, pero de distinta procedencia y lengua materna. El objetivo es indagar si el bagaje léxico de las muestras es igual, en qué medida y cuáles son sus puntos comunes, diferencias o carencias.

El Capítulo V se divide en dos partes y aborda el análisis cualitativo de las listas de léxico desde las perspectivas transversal –análisis general y por variable– y longitudinal. Primeramente, hemos estudiado el tipo de léxico activado en los centros de interés a partir del estudio de los índices de disponibilidad léxica, frecuencia de aparición y frecuencia acumulada. Hemos analizado, en particular, los vocablos más disponibles (tomando como medida de corte los que alcanzan un índice de disponibilidad igual o mayor de 0,1) para detectar las temáticas predominantes en cada campo y las categorías gramaticales más difundidas en las respuestas. En segundo lugar, hemos cuantificado la

cardinalidad del conjunto con el propósito de medir la compatibilidad existente entre las variantes de una misma variable o entre dos grupos.

La última parte consiste en un estudio comparativo de las unidades que forman las listas de léxico disponible con el *Corpus de Referencia del Español Actual* y el *Corpus de aprendices de español* finalizado a verificar si las aportaciones de los informantes están incluidas en estos bancos de datos tomados como corpus de referencia.

El estudio de la riqueza léxica prosigue el bloque analítico para complementar la información adquirida hasta ahora sobre la competencia de los aprendices, como ya hemos explicado, así, en el Capítulo VI mostramos los cómputos y análisis que nos han permitido medir la riqueza de vocabulario de los encuestados, siguiendo el mismo orden de los capítulos anteriores. Para ello, hemos examinado el corpus de trabajo mediante los indicadores de la variación y de la densidad léxica. El paso siguiente ha supuesto el estudio del mismo muestreo en función de los resultados por variable a raíz de las diferencias cuantitativas entre *tokens* y *types* y entre palabras funcionales y nocionales. De nuevo, la descripción de los datos estadísticos ha permitido afinar la discusión de los resultados demostrando la simetría o la dispersión del comportamiento de los participantes. También para este índice se han comparado los datos procedentes de la primera prueba con los de la segunda –ambas suministradas a los informantes de nivel B2–. Se ha propuesto, por primera vez en este marco, un análisis longitudinal, introduciendo una novedad metodológica que ha permitido analizar el grado de aumento de la competencia léxica de los aprendices en relación con el tiempo de aprendizaje lingüístico.

Se concluye este capítulo con el análisis comparativo con los datos de investigaciones centradas en la producción escrita de otros estudiantes de ELE, cuya finalidad es corroborar la hipótesis de que discentes no nativos que comparten el dominio lingüístico, pero de distinta procedencia, poseen iguales capacidades expresivas.

El Capítulo VII finaliza la Tesis ofreciendo un estudio sobre la riqueza léxica desde una perspectiva pionera: proponemos un acercamiento al análisis cualitativo de los datos. Nuestra pretensión, aun siendo conscientes de los límites que conlleva un trabajo preliminar, es abrir una nueva vía de encuesta para futuras investigaciones. Efectivamente, aunque los indicadores cuantitativos aportan un tipo de información sobre la riqueza léxica que va más allá de los datos numéricos, es decir la “calidad de la

escritura” (López Morales, 2011), hemos decidido plantear un estudio de las unidades léxicas aportadas en el conjunto textual rebasando las meras cifras. Para cumplir con este cometido, hemos articulado un estudio en torno a la lista de los veinticinco vocablos más frecuentes para determinar morfológicamente las lexías más utilizadas, el área semántica a la que pertenecen y el grado de asociación con la trama propuesta para la redacción del texto, cuyo tema debería poderse adivinar en función de las palabras clave, que hemos recopilado en otra lista.

Finalmente, consagramos las últimas páginas a las conclusiones que destacan y discuten tanto los resultados alcanzados en este trabajo, como sus aspectos más novedosos, aun dentro de sus limitaciones.

Por último, presentamos la bibliografía citada y los anexos, en los que se recogen:

- (i) el modelo de la prueba suministrada;
- (ii) la lista de parada utilizada en el análisis de la riqueza léxica;
- (iii) los diccionarios de léxico disponible ordenados por índice de disponibilidad y por orden alfabético;
- (iv) los datos completos de cada texto utilizados para computar los índices de riqueza léxica.

Esperamos que la aquí presente Tesis pueda abrir nuevos caminos, en parte relacionados con la exploración de la competencia léxica de los aprendientes italianos de español y, en general, del aprendizaje de ELE; en parte dirigidos a la proposición de un nuevo protocolo experimental aplicable a diferentes entornos investigativos y educativos.

Por eso, confiamos en que cumpla los objetivos que nos fijamos en un principio y que nos han acompañado durante estos años de trabajo.

CAPÍTULO I.

LINGÜÍSTICA APLICADA A LA ENSEÑANZA DE LAS LENGUAS EXTRANJERAS

1. INTRODUCCIÓN

Aprender una lengua es un proceso complicado que requiere tiempo y motivación.

Esta afirmación, en apariencia banal, es el eje central de la investigación de distintas disciplinas, cuya finalidad es contestar a las preguntas que lingüistas, docentes y psicólogos se plantean al observar los aprendices de una lengua extranjera.

En la evolución natural del ser humano existe un hecho obvio: la lengua materna (L1 o LM) *se adquiere* desde los primeros meses de vida. Las otras lenguas, distintas de la LM, *se aprenden* cuando ya se posee un sistema cognitivo desarrollado² y se consideran segundas lenguas (L2) o lenguas extranjeras (LE)³.

Las cuestiones abiertas en torno a los procesos que subyacen al desarrollo lingüístico son múltiples y variadas. Entre otras, las que más interés despiertan se refieren, en primer lugar, a las diferencias entre el aprendizaje de una lengua nueva con respecto a la adquisición de la nativa, es decir a, las teorías sobre los mecanismos cognitivos empleados por los aprendientes extranjeros; en segundo lugar, a las metodologías adecuadas para una enseñanza fructífera de la lengua objeto (LO); y, por último, a los métodos de medición y de evaluación de dicha competencia, cuyos propósitos coinciden

² Krashen (1981) pone de relieve la diferencia entre los términos causada por el contexto formativo que produce dos tipos de conocimiento: la *adquisición* es el conocimiento de la lengua de forma natural en ámbito cotidiano, entendida como proceso biológico y natural; el *aprendizaje* coincide con el conocimiento formal adquirido en el aula, basado en la escolarización.

³ Se trata de conceptos distintos: según Larsen-Freeman y Long (1991) LE se refiere al estudio de una lengua *x* en el país nativo del aprendiz, por ejemplo, un italiano que estudia español en Italia; L2 se refiere al estudio de una lengua *x* en un país *x*, por ejemplo, un italiano que estudia español en España (contexto de inmersión). Baralo (2011 [1999]: 22-23) se refiere también al tipo de contexto: la L2 es el idioma que se aprende en contexto natural (de forma parecida a como se adquiere la LM), esto es, en una comunidad que dispone de dos sistemas lingüísticos en contacto (bilingüismo y diglosia); mientras que la LE se aprende de forma tutorizada en contexto institucional-formal; se trata, con todo, de categorías fluidas: si una LE se utiliza muchas horas (por ejemplo, al estudiar en un colegio bilingüe o al estar de vacaciones en el país nativo), dicha lengua puede pasar a ser una L2 más que una LE.

En el ámbito de la investigación en torno a LE/L2 se emplea la expresión, traducida del inglés “*second language acquisition*”, “adquisición o aprendizaje de una lengua segunda” (ASL), en relación con todos los procesos de aprendizaje lingüístico, subconscientes y conscientes, de los aprendices no nativos. En este trabajo utilizaremos indistintamente los términos *L2*, *LE*, *adquisición* y *aprendizaje* para referirnos a la lengua estudiada por el alumnado.

Cabe señalar, a estas alturas, que ambos idiomas se definen también lengua objeto (LO).

parcialmente con los de la presente Tesis. En efecto, los investigadores pretenden abarcar tanto el análisis de los resultados alcanzados por los discentes, como el estudio de la eficacia del método didáctico (entendido como manera planificada de enseñar un idioma, del que depende la acción docente), además de describir y explicar el proceso adquisitivo.

Todo esto confluye en el campo de la lingüística aplicada, la cual está encaminada a “mejorar la eficacia de una tarea práctica en la que la lengua es el componente central” (Santos Gargallo, 2017 [1999]: 10) valiéndose de la contribución de otras disciplinas debido a la complejidad del proceso de adquisición de una LE (la psicolingüística, la sociolingüística y, pese a ciertas controversias, las ciencias de la educación)⁴. De ahí que la lingüística aplicada se sirva de recursos teóricos y prácticos de distinta naturaleza (psicológico, lingüístico, social y educativo) y se mueva entre marcos teóricos distintos, pero con un ámbito de actuación común: garantizar que el aprendiz logre el resultado deseado y esperado de su ASL.

2. TEORÍAS LINGÜÍSTICAS SOBRE LA ENSEÑANZA DE LENGUAS EXTRANJERAS

En este apartado reseñamos, sin la pretensión de ser exhaustivos, los rasgos generales de las teorías lingüísticas sobre la ASL propuestas desde los años cincuenta del siglo pasado hasta la fecha, cuyo propósito es, fundamentalmente, explicar cómo funcionan los mecanismos responsables de la adquisición de las LE⁵, en las que se fundan los métodos y enfoques didácticos aplicados a la enseñanza.

2.1 La teoría conductista

El conductismo parte de la premisa de que el proceso de aprendizaje de una LE se desarrolla de la misma manera en la que se adquiere la LM. Se concibe el aprendizaje como un iter de formación de hábitos por medio de la imitación y de la repetición de un aducto lingüístico recibido del exterior que permite la producción de un educto en la LO. Skinner (1957), uno de sus mayores representantes, defiende que la conducta lingüística de un aprendiz se explica mediante el modelo *estímulo-respuesta-refuerzo*, que consiste

⁴ Véase la obra completa de Santos Gargallo (2017 [1999]).

⁵ Al no perseguir una revisión historiográfica de la cuestión, no abordamos el estudio de periodos anteriores. Para una discusión detallada véase López García (1988); Larsen-Freeman y Long (1991); Ellis (1994); Doughty *et al.* (2002) y, para el caso específico de ELE, Liceras Muñoz (1992).

en la sucesión de los siguientes procesos:

- (i) observación del *input* recibido y entendido;
- (ii) creación de hipótesis;
- (iii) comprobación de las hipótesis⁶;
- (iv) fijación de normas;
- (v) reflexión metalingüística.

Se aboga por la naturaleza comportamental del lenguaje subrogando su adquisición a las leyes del aprendizaje ya que el proceso es igual tanto en la L1 como en la L2.

Los años sesenta se van a caracterizar por la polémica mantenida entre los seguidores de estas tesis y los defensores del modelo generativista propuesto por Chomsky (1959).

2.2 La teoría nativista

Los principios en que se apoya el conductismo vacilan porque se asume que la adquisición lingüística no es solo un proceso de imitación: hay que aceptar que los seres humanos nacen biológicamente programados para el lenguaje, es decir que tienen una capacidad innata de adquirir una lengua gracias a un procesador mental (de ahí que la teoría se nombre también *innatista*)⁷.

A raíz de este planteamiento se considera el lenguaje como construcción creativa y la adquisición como proceso constituido por:

- (i) formación de hipótesis lingüísticas;
- (ii) verificación de las hipótesis;
- (iii) automatización a través de la práctica.

Estos pasos pueden aplicarse al aprendizaje de cualquier lengua porque:

[...] el dispositivo de adquirir lenguas capacita al aprendiz a relacionar datos que percibe de su entorno con un sistema lingüístico universal, una especie de plantilla que le permite categorizar elementos para poder seleccionar los más apropiados según la tarea comunicativa inmediata (White, 1996 *apud* Griffin, 2005: 36).

⁶ Este paso fomenta la autonomía del aprendiz que intenta llegar por sí mismo a la comprobación de su hipótesis.

⁷ Se trata de una dotación biológica del ser humano, definida como un mecanismo computacional o programa genético que contiene todos los principios compartidos por todas las lenguas (los “universales tipológicos”). Véase Chomsky (1986) para una detallada argumentación del asunto.

De igual manera, Corder (1967) afirma que, si el niño está predispuesto a adquirir la LM en virtud de una predisposición innata, la misma adquisición podrá verificarse para una cualquiera LE.

La *monitorización* de Krashen (1981, 1982, 1985), es uno de los modelos más representativos de las teorías nativistas. Este asume que el ser humano posee un talento lingüístico específico a partir de cinco hipótesis:

1. la adquisición de la LM es distinta del aprendizaje de una LE;
2. el aprendizaje es gestionado por un monitor que detecta los errores y los corrige;
3. el desarrollo de la LE obedece a un patrón universal, peculiar de cada lengua e independiente de la L1;
4. el *input* activador del proceso de adquisición ha de ser comprensible, y debe introducir alguna novedad o dificultad para el discente;
5. el proceso de adquisición de la L2 está sometido a un conjunto de variables que permiten o bloquean el proceso (definido *filtro afectivo*).

2.3 La teoría cognitivista

Esta corriente tiene sus orígenes en la interacción entre lingüística aplicada y psicología que da como resultado la aparición de una nueva disciplina, la psicolingüística, que trata la adquisición del lenguaje como una capacidad humana que se desarrolla en virtud de la relación entre la mente del sujeto aprendiente con el mundo externo. Según este paradigma el primer estadio de aprendizaje es social y después individual: Vygotsky (1962) habla de un proceso de interiorización que parte del aprendizaje en sociedad.

Ellis (1994: 356) inscribe dentro de este panorama la propuesta de Krashen (1981, 1982, 1985) ya que un sujeto solo puede adquirir una nueva lengua gracias a la información vehiculada por elementos externos (contexto, situación, conocimiento del mundo), desde el momento en el que se asume que para la construcción del conocimiento lingüístico es necesaria la intervención de: (i) el *input*, o aducto, que coincide con la entrada de información; (ii) el *intake*, o procesamiento de la información; (iii) el *output*, o educto lingüístico, que sería la producción en la LO.

Se pone de manifiesto que el estudio de una LE debe ser determinado por su función comunicativa porque la lengua es la única herramienta que posibilita la

comunicación y la interacción social, debido al impacto del contexto externo sobre el aprendiz.

2.4 La teoría ambientalista

Desde estos hallazgos, la ASL se explica a partir del papel que el entorno tiene en el aprendiz, focalizándose más en la incidencia de ciertos condicionantes externos que en sus capacidades, ya sean cognitivas ya sean innatas. Por eso, las teorías ambientalistas consideran que la experiencia social tiene más influencia en el desarrollo del aprendizaje que las cualidades naturales de los sujetos porque la adquisición de la L2 es un subproducto de la asimilación cultural.

Una de las propuestas más representativas en este contexto es la de Schumann (1978), quien sostiene que el grado de distancia social y psicológica entre el aprendiz no nativo y los hablantes nativos determina el éxito o el fracaso del aprendizaje, ya que este es fruto de la adaptación a una nueva cultura.

Se incluyen en esta corriente aquellas teorías que dejan de lado los procesos cognitivos del discente como apuntan Larsen-Freeman y Long (1991) porque se concede más importancia a la experiencia personal y a otras variables externas que a las dotes innatas.

2.5 La teoría interaccionista

Como síntesis de los enfoques posteriores, las teorías interaccionistas de los años ochenta del siglo XX intentan explicar la adquisición de la LE recurriendo tanto a aspectos cognitivos (factores innatos) como ambientales (factores externos). Doughty (2000: 167) señala:

[...] la teoría interaccionista social de la adquisición de la lengua tiene un gran atractivo al incorporar tanto la fascinante capacidad del lenguaje innata del ser humano como la fuerza impulsora y determinante del propósito comunicativo en la adquisición del lenguaje.

Se tiene en consideración que la sola Gramática Universal no es suficiente para obtener un buen conocimiento de la LO: las aportaciones de la sociolingüística conciben las lenguas como medio de comunicación, pues se sitúan dentro de un marco comunicativo en el que el ambiente externo y la socialización se revelan elementos fundamentales durante el iter adquisicional (Long, 1985).

Si bien las teorías interaccionistas presentan grandes diferencias unas de otras⁸, Stern (1983) resume todos los postulados en un modelo integrador, cuya característica principal es concebir la ASL como multidimensional y multifactorial porque la competencia lingüística se consigue mediante las facultades individuales del aprendiz, de un lado, y de la dimensión social, del otro.

Así pues, la renovación teórica de este marco teórico reside en la centralidad del componente comunicativo del lenguaje, según la cual la interacción hace posible la comprensión del *input* y su interiorización: el conocimiento de la LE “no ha de entenderse como algo que se queda en el cerebro del aprendiz sino como algo que se construye en la interacción” (Young, 1999).

Desde un punto de vista práctico, el discente debe (i) sentir la necesidad de comunicarse en la L2; (ii) disponer de una buena cantidad de *input*; (iii) conocer el tema de la conversación; (iv) estar expuesto a diversas funciones comunicativas; (v) tener oportunidad de interactuar de forma desinhibida (Ellis, 1994).

3. LA DIDÁCTICA DEL LÉXICO DE LAS LENGUAS EXTRANJERAS

Hemos observado que los marcos teóricos han evolucionado notablemente a partir de las primeras hipótesis que consideraban el aprendizaje de LE como la formación de hábitos lingüísticos (conductismo), pasando por una interpretación innatista, hasta llegar a una explicación de los procesos adquisicionales tanto desde una perspectiva individual (cognitivismo) como social (interaccionismo).

Según la psicolingüística el conocimiento de una LE es el resultado de una actividad cognitiva sometida a las características propias de cada sujeto e influenciada por el entorno que le rodea mediante la entrada de la información (*input*), la subsiguiente apropiación de los datos (*intake*) y, por último, la salida del producto almacenado (*output*)⁹. Todo ello implica aceptar que cualquier procedimiento didáctico gire en torno al aprendiz: quién es, cuáles son sus motivaciones y qué operaciones mentales se activan para que se desarrolle la competencia comunicativa.

⁸ Cfr. Givón (1979); Preston (1989); Long (1996); Gass (1997); Young (1999).

⁹ Se trata de los siguientes factores internos: aptitud, motivación, actitud, personalidad, estilo cognitivo, especialización hemisférica, estrategias de aprendizaje, voluntad, interés, sexo, experiencias previas (Gardner, 1985), sin mencionar las disfunciones lingüísticas u otros trastornos. En segundo lugar, los factores externos son: la situación de aprendizaje (natural o formal), el contexto de uso de la lengua meta, el tipo de exposición lingüística, la frecuencia de uso, los tipos de interacción y de interlocutores.

Resumimos esquemáticamente estos elementos:

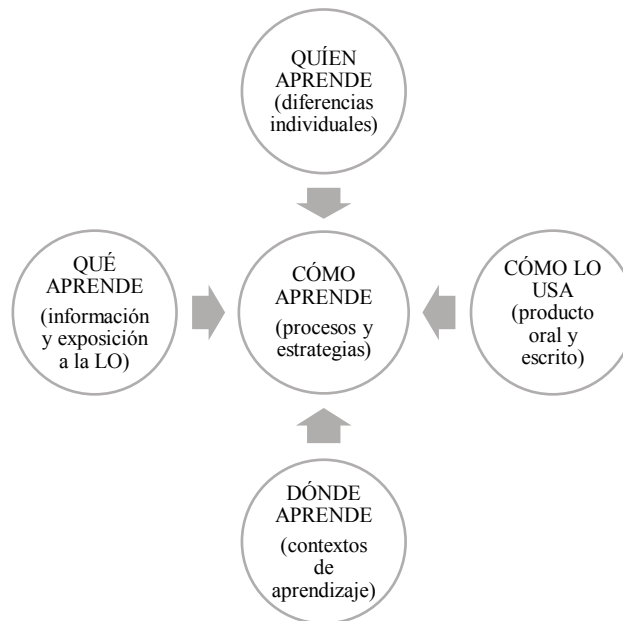


Figura 1. Diagrama de la enseñanza centrada en el alumno.

En definitiva, la didáctica de la LE se concibe como un

Conjunto de decisiones que se toman –en relación con un contexto docente concreto– a la hora de determinar los objetivos del aprendizaje, el conjunto de *contenidos* que son pertinentes en función de los objetivos previamente establecidos, los *procedimientos* que se emplearán para desarrollar los contenidos y asegurar el logro de los objetivos, los *materiales* que concretarán los tres elementos mencionados, y que serán instrumento fundamental para llevar a cabo este proceso; y los criterios para la *evaluación* de la competencia comunicativa. (Santos Gargallo, 2017 [1999]: 29)

Este esquema es consecuente de varias propuestas metodológicas que han ido conformando el proceso de enseñanza-aprendizaje, como veremos a continuación en lo referente al tratamiento del léxico.

En los siguientes apartados llevamos a cabo la revisión de la didáctica del vocabulario en la enseñanza de las LE, describimos los cambios que ha experimentado el componente léxico en los distintos métodos y enfoques didácticos, planteados según los postulados teóricos que hemos reseñado. Seguimos la división temporal de Morante Vallejo (2005), quien parte de la introducción de los métodos tradicionales en los años cincuenta hasta llegar al eclecticismo de la actualidad¹⁰ y ampliamos la información con

¹⁰ Definida “era post-método” por Sánchez Pérez (1992: 344) ya que la didáctica de LE actual no aplica

el trabajo analítico de Kumaravadivelu (2006a).

Morante Vallejo (2005) destaca tres etapas bien distintas en las que el foco de la enseñanza del vocabulario se desplaza de las estructuras lingüísticas y las listas de palabras, hacia una revaloración del léxico, pasando por los aspectos cualitativos de la competencia léxica. Cada una de estas fases se basa en una teoría lingüística de referencia, cuyos planteamientos afectan a los estudios de adquisición y enseñanza. De ahí que se reflejen en la concepción del léxico, la cual tiene repercusiones en las prácticas docentes, el diseño de materiales y los temas tratados en ámbito investigativo.

Todo esto confluye en los métodos de enseñanza empleados y las estrategias de aprendizaje potenciadas¹¹.

3.1 Primera etapa

Hasta los años setenta el vocabulario se trata como componente secundario de la lengua, supeditado a la gramática y la sintaxis. Las palabras se aprenden mediante pequeñas listas organizadas por paradigmas flexivos y recopiladas según el criterio de frecuencia. Según el estructuralismo vigente, la lengua es un conjunto de estructuras por lo que la ASL está finalizada a su identificación y adquisición.

Por otra parte, la *Hipótesis del Análisis Contrastivo*¹² se apoya en las teorías conductistas, que sostienen que el aprendizaje lingüístico consiste en la creación de hábitos que se desarrollan al transferir formas y significados de la LM a la LO.

Los aspectos básicos de la enseñanza (el único eje de la investigación en el ámbito de las LE) son la sintaxis y la fonética. En lo que se refiere al léxico, no es importante conocer grandes cantidades de palabras, sino que se requiere la memorización de lo suficiente para practicar las estructuras sintácticas. El bagaje del vocabulario se

rígurosamente un único modelo, sino que mezcla diferentes propuestas con el objetivo de ofrecer al alumnado las actividades y estrategias más adecuadas a sus necesidades (Larsen-Freeman, 1986: 2).

¹¹ De manera parecida, Richards y Rogers (2001) establecen que cada metodología destinada a la enseñanza debe necesariamente componerse de tres elementos: una teoría lingüística; un enfoque, o aproximación; un conjunto de procedimientos que regulen el quehacer diario de docentes y alumnos.

¹² Este término agrupa una serie de procedimientos encaminados a mejorar la enseñanza de LE a través de la identificación de las áreas de mayor dificultad para el aprendizaje y susceptibles de una mayor frecuencia de error. Se hipotiza que el contraste sistemático entre la LM y la LO permitiría establecer una jerarquía de dificultades mediante la identificación de las similitudes y diferencias estructurales entre los idiomas. La interferencia con la L1 se considera la principal causante de las potenciales dificultades en la ASL.

A este propósito, Lado (1957: 2), uno de los autores más representativos de la hipótesis, afirma que “aquellos elementos que sean parecidos en su lengua nativa le resultarán más sencillos y los que sean distintos le serán más difíciles”.

profundiza en una fase más avanzada de la ASL, de manera casual ya que se presentan las palabras a medida que aparecen en actividades descontextualizadas y sin proponer estrategias de aprendizaje.

3.1.1 *El Método Tradicional (o de Gramática y Traducción)*

El Método Tradicional asume que los alumnos aprenden el vocabulario de la LO mediante listas bilingües de palabras organizadas por campos semánticos. El léxico está subordinado a la gramática, se estudia por separado y se selecciona exclusivamente en función de su utilidad para la comprensión de textos escritos. La enseñanza está finalizada a la lectura y la traducción de obras literarias. La lengua se aprende para traducir, no para hablar.

La secuenciación didáctica se desarrolla según el siguiente esquema:

- (i) presentación de reglas gramaticales;
- (ii) ejemplificación;
- (iii) ejercicios de traducción.

3.1.2 *El Movimiento de Reforma*

Como respuesta al modelo anterior, se dejan de utilizar listas de palabras descontextualizadas. Los lexemas se presentan mediante textos reales, realizando prácticas en contextos significativos y rechazando la traducción (de hecho, el uso de la LM se limita a comprobar la comprensión o, si necesario, explicar palabras nuevas).

El Movimiento de Reforma constituye una importante contribución a la enseñanza de lenguas gracias a la nueva valorización del uso de la lengua oral frente a la escrita: la lengua hablada cobra una relevancia desconocida hasta entonces.

Los contenidos gramaticales y el vocabulario siguen estudiándose por separado: queda todavía mucho para que la integración de ambos conocimientos alcance valor.

3.1.3 *El Método Directo (o Natural)*

Los preceptos de este método prevén el uso exclusivo de la LO durante las horas de clase (lo cual significa evitar la traducción) y la revalorización de las destrezas orales. De ahí que la gramática empiece a supeditarse a las necesidades comunicativas. El vocabulario que se trabaja es concreto, contextualizado y está relacionado con la vida

cotidiana. El énfasis de la enseñanza está en el uso de la lengua y no en la memorización: el alumno deja de aprender de memoria reglas y listas de palabras.

Los procedimientos didácticos consisten en una interacción continua entre el profesor y los alumnos (preguntas-respuestas), dirigida por *el principio de la secuencia metodológicamente adecuada*, es decir, se trata de una propuesta graduada: por ejemplo, las unidades léxicas nuevas solo se enseñan una vez que se han adquirido aquellas presentadas previamente.

3.1.4 *Los Métodos Estructurales*

Entre los Métodos Estructurales contamos con la aparición del *Reading Method*, según el cual el conocimiento del vocabulario está finalizado al alcance de la capacidad lectora del alumnado. Por tanto, proporcionar una adecuada selección léxica es el objetivo más importante, que se logra utilizando listas de frecuencia para la creación de materiales didácticos.

En segunda instancia, el *Situational Language Teaching* (Método Situacional) pone los cimientos para el futuro desarrollo de los métodos orales (en particular, el Enfoque Comunicativo) que supondrán el recurso a todos aquellos soportes que puedan ayudar a la comunicación.

Por primera vez, se considera el vocabulario un componente fundamental en la ASL por lo que una buena selección léxica es una necesidad imperiosa. Se proponen el empleo de criterios más objetivos y la nivelación de los contenidos conforme al dominio lingüístico del estudiantado. A la luz de estos presupuestos, se presenta una cantidad preestablecida de unidades léxicas, basada en listas de frecuencia recopiladas y simplificadas según el nivel de aprendizaje, omitiendo las léxias consideradas difíciles (de ahí la aparición de las lecturas graduadas).

3.1.5 *El Método Audiolingual*

Los principios de las corrientes estructuralista y conductista confluyen en este método, cuyo foco privilegiado es la construcción de buenos hábitos lingüísticos por medio del desarrollo de la destreza oral y de un conocimiento limitado del vocabulario. Parece que proponer demasiadas unidades léxicas lleva al discente a concebir el aprendizaje como un proceso de acumulación de palabras.

El vocabulario, de nuevo, se califica como componente secundario: se aboga por la repetición mecánica y controlada de patrones estructurales para conseguir los nuevos y correctos hábitos verbales.

Los procedimientos didácticos se centran en la memorización de estructuras y en la repetición, dejando el mínimo espacio a los errores.

3.2 Segunda etapa

A principios de los años setenta, se formula un nuevo paradigma metodológico, cuyo cambio de perspectiva lleva a una inédita valoración del léxico que cobra la misma importancia de los demás elementos de una lengua. Estudiar vocabulario significa adquirir significados y no memorizar palabras: gana protagonismo la semántica léxica.

El objetivo de la ASL es comunicarse. Se respalda la utilidad de las actividades que suponen una interacción auténtica a través de una visión más abarcadora de la competencia lingüística, que permite la producción de oraciones correctas y adecuadas a una situación y un contexto. De ahí que se fomente la integración de la competencia gramatical y de la pragmática que confluyen en la *competencia comunicativa* (Hymes, 1972), definida por Canale (1983 *apud* Llobera, 1995: 65)

el dominio de los sistemas subyacentes de conocimiento y habilidad requeridos para la comunicación (por ejemplo, conocimiento del vocabulario y habilidad de usar las convenciones sociolingüísticas de una lengua dada) [...] El conocimiento hace referencia a lo que uno sabe, consciente o inconscientemente, sobre el lenguaje y sobre otros aspectos del uso comunicativo del lenguaje; la habilidad hace referencia a lo bien o mal que se utiliza este conocimiento en la comunicación real.¹³

Asimismo, la etnografía de la comunicación empieza a ejercitar su influencia y a destacar el mayor peso de los factores sociolingüísticos en la ASL, alejándose de la centralidad que tenía antes la creación de hábitos.

Todo esto se caracteriza por el interés que despiertan la interacción y la sociedad, dos conceptos hasta este momento marginados por la lingüística. Las contribuciones del

¹³ Asume que el conocimiento y la habilidad son aspectos independientes y complementarios a cada una de las cuatro subcompetencias lingüísticas que componen la *competencia comunicativa*:

- (i) la *competencia gramatical* coincide con el conocimiento de formas y estructuras lingüísticas;
- (ii) la *competencia sociolingüística* está relacionada con aquellas reglas socioculturales que permiten actuar adecuadamente en un contexto determinado;
- (iii) la *competencia discursiva* consiste en el manejo eficaz de la cohesión y la coherencia;
- (iv) la *competencia estratégica* se compone de elementos gramaticales, sociolingüísticos y discursivos que mejoran la comunicación y solucionan los fallos comunicativos.

cognitivismo, por su parte, subrayan el hecho que es la formación de hipótesis la clave que conduce a la competencia lingüística, por tanto, el discente tiene que involucrarse activamente en el proceso de aprendizaje para que sea significativo. En efecto, es compartida la idea de que cada proceso es singular e individual, por lo que hay que poner atención en las estrategias que utilizan los aprendices para llegar al éxito (Šifrar Kalan, 2000: 903).

Estos hallazgos hacen necesario encontrar estrategias *ad hoc* para la enseñanza-aprendizaje del vocabulario ya que:

De acuerdo con este nuevo planteamiento, el léxico se convierte en un elemento imprescindible para lograr dicho objetivo de la recién nacida metodología comunicativa. Si comunicar es transmitir significado, como afirma Lewis (1993, p. 389), las palabras contienen más significado que la gramática. Esta afirmación se hace eco en la conocida cita de Wilkins (1972, p. 111) que plantea que sin gramática poco se puede transmitir, pero sin vocabulario, nada [...]. (Rufat, 2018: 76)

3.2.1 *El Método Nocio-funcional*

Para el Método Nocio-funcional la clase de idiomas se desarrolla en función de las necesidades y expectativas de comunicación del alumnado, esto significa que los contenidos lingüísticos se enseñan por su relevancia comunicativa¹⁴. Se distinguen dos tipos de contenidos trabajados en aula: los nocionales, de carácter semántico-gramatical y los funcionales, que abordan modalidad, evaluación, persuasión, argumentación, exposición, relaciones emocionales e interpersonales.

Dentro de este marco, el vocabulario se aprende en contexto, a saber, se incluye en situaciones comunicativas adaptadas al nivel y los intereses de los discentes.

3.2.2 *El Enfoque Comunicativo*

Ahora el objeto privilegiado por la enseñanza es el uso comunicativo de la LO, así que el potenciamiento y la integración de todas las destrezas lingüísticas (comprensión oral y lectora, producción oral y escrita) están en la base de las actividades pedagógicas. El Enfoque Comunicativo supone que la lengua coincide con el discurso. Se proponen ejercicios basados en situaciones comunicativas contextualizadas y deja de ser útil

¹⁴ En 2001, el Consejo de Europa planteará el proyecto del *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (MCER) a partir de las directrices teóricas y prácticas de este método.

aprender piezas léxicas aisladas porque, como ya se había hecho en el Método Nocio-funcional, el contexto guía todo tipo de interacción.

A estas alturas, el significado y la fluidez prevalecen sobre la forma y la precisión: no se requiere que los alumnos sepan cosas sobre la lengua o dominen sus estructuras, sino que sean capaces de comunicarse adecuadamente en entornos reales y conforme a sus necesidades.

La acción docente se organiza a raíz de las situaciones que se quieren presentar al estudiantado y que determinan los asuntos que se van a tratar en el aula. La selección de los contenidos se lleva a cabo en función de campos semánticos que tienen en consideración, simultáneamente, criterios cuantitativos (frecuencia) y cualitativos (rentabilidad, utilidad) con el propósito de impulsar un uso real de la LO, esto es, imitar el habla de los nativos. La autenticidad del *input* y de los materiales es fundamental (se introducen el uso de los *realia*). Los procedimientos didácticos se secuencian en las fases de *presentación-práctica-producción* que conducen a la construcción del conocimiento.

La renovación aportada por esta corriente se aprecia en la tipología de actividades propuestas (ejercicios de vacío de información, juegos de rol, simulaciones y debates) que fomentan la interacción y permiten aprovechar el aula como un espacio de comunicación. Allí, se recrean usos auténticos de la LO a partir de los siguientes elementos: las áreas temáticas sobre los que versa la comunicación; las situaciones en que dicha comunicación tiene lugar; las funciones concretas que se realizan; los registros lingüísticos implicados por el uso de tales funciones (Sánchez Pérez, 1992).

3.2.3 *El Enfoque Natural*

Retomando muchas de las pautas anteriores, el Enfoque Natural sigue acentuando la importancia del aprendizaje del léxico porque conlleva los significados que se quieren expresar, pues se basa en la exposición a un *input* comprensible, significativo y auténtico que favorezca la comunicación oral.

El vocabulario se adquiere mediante la enseñanza formal (organizada y sistematizada) y estrategias específicas. Se ponen en tela de juicio las listas de palabras y el criterio de frecuencia en la elección de los contenidos ya que no respetan ni el principio de utilidad ni de relevancia ni de cobertura; es más, varían en función del corpus y del contexto (oral o escrito) del cual se han extraído.

Los contenidos léxicos, por ende, se escogen conforme a una valoración más subjetiva, dejando de un lado los recuentos estadísticos que hasta la fecha habían gobernado los procesos de selección léxica y la creación de materiales: los aprendices estudian piezas léxicas contextualizadas y graduadas a su nivel. Al principio de la ASL, se fomenta el estudio de un vocabulario básico y estratégico que posibilite solucionar dificultades comunicativas (utilizar palabras generales, parafrasear, adivinar e inferir significados) y, solo en una fase más avanzada, se aconseja la ampliación hacia un tipo de léxico más específico y terminológico.

3.3 Tercera etapa

Desde mediados de los años ochenta, en particular desde la publicación de *Vocabulary in a second language* de Meara (1983), tiene lugar una revaloración del léxico como componente esencial de la ASL. El aprendizaje pasa a ser cualitativo: no se trata de la mera acción de memorizar palabras, sino que son importantes también las formas, los usos y las relaciones entre las unidades léxicas, cuyo concepto se extiende más allá de la simple palabra-forma (el material gráfico comprendido entre dos espacios en blanco) hasta abarcar las nociones de polisemia, colocaciones, lenguaje idiomático, fraseología.

La intervención de disciplinas como el análisis del discurso y la pragmática contribuye a sentar la idea de que el vocabulario determina los contenidos de la comunicación. De la misma manera, la lingüística computacional participa en esta renovación destacando una inédita concepción del papel del léxico, pues, se hipotiza que la producción del lenguaje consiste en la recuperación de unidades almacenadas en la memoria.

Surge el concepto de *lexicón mental*¹⁵, esencial para el impulso del componente léxico en el aula de segundas lenguas.

¹⁵ El lexicón mental se concibe como una galería de palabras almacenadas en el sistema cognitivo al que se accede para comprender un mensaje y para expresarse. La lengua se procesa en el lexicón según los siguientes pasos: (i) reconocimiento visual o auditivo de las piezas léxicas; (ii) activación de las unidades que lo componen; (iii) selección para la comprensión o la producción. Cabe señalar que no se trata de un almacén estático, sino que es un proceso dinámico (Baralo, 2001).

Asimismo, se trata de una red con múltiples conexiones semánticas, sintácticas, fonológicas, ortográficas, morfológicas, cognitivas, culturales y autobiográficas (Thornbury, 2002) que permiten retener las palabras con toda la información morfosintagmática necesaria para su conocimiento. Los seres humanos poseen la facultad de almacenar cadenas de material lingüístico en una representación semántica y fónica, y formar conexiones entre estas. De ahí que se hable de la metáfora de la red (Morante Vallejo, 2005: 47).

El interés de la investigación se desplaza hacia la adquisición y la figura del discente, se abren cuestiones relacionadas con la organización de lexicón en la mente humana: *¿qué significa aprender una palabra?* y *¿en qué consiste el desarrollo del léxico?*¹⁶ Para dar respuesta a tales preguntas, en primer lugar, se acuña la noción de *competencia léxica* (Canale y Swain, 1980; Marconi, 1999; Jiménez Catalán, 2002)¹⁷.

En segundo lugar, se plantean teorías sobre la adquisición y los mecanismos de

Miranda García (1993: 94) asevera que las diferentes concepciones sobre el lexicón mental se pueden repartir en dos modelos: de un lado, el “modelo telaraña” ya que este tipo de organización evoca la imagen de una tela de araña en la que las palabras están interconectadas; del otro lado, el “modelo atómico” implica que las palabras son moléculas que se forman a partir de un inventario de átomos de significado y aquellas que están semánticamente relacionadas presentan átomos comunes. En todo caso, estas asociaciones facilitan la comprensión y la producción del vocabulario. Las redes ayudan a visualizar las relaciones entre las piezas léxicas.

En resumidas cuentas, es, a la vez, un almacén de piezas léxicas y un modelo abstracto que representa la organización de los ítems léxicos (cómo se guardan en la memoria y se recuperan).

¹⁶ En palabras de Rufat (2018: 81):

Para entender el proceso de adquisición del vocabulario es preciso analizar la organización del lexicón mental y los mecanismos del procesamiento del lenguaje. [...] se acepta comúnmente que la organización y el funcionamiento del lexicón son cualitativamente semejantes en la L1 y en la L2, si bien el proceso de adquisición del léxico en la L2 pasa por el filtro del conocimiento preexistente: la L1.

Así pues, no obstante parezca que la LM y la LE funcionan de la misma manera, el aprendizaje de una LE parte de un sistema cognitivo previo, por lo que los aprendientes extranjeros tienen una memoria semántica a su disposición, la cual ejerce cierto influjo en la ASL ya que el nuevo *input* se interpreta bajo los parámetros ya fijados por la LM. En lo que se refiere a la adquisición del léxico también hay que tener en consideración los procesos de influencia y de transferencia positiva y negativa de la L1.

De cualquier modo, parece que el aprendizaje del vocabulario es el mismo en ambos casos: empieza con el reconocimiento de una forma y sigue con el establecimiento de conexiones entre dicha forma con uno o más significados. Para ello, según el modelo propuesto por Aitchison (1994), y posteriormente por Baralo (2001), hay tres fases interconectadas: (i) etiquetado; (ii) empaquetado; (iii) construcción de redes semánticas. Tales redes implican, a su vez, el establecimiento de relaciones entre unidades léxicas. Por tanto, el aprendiz incorpora nuevos elementos en una red ya construida y crea diferentes asociaciones paradigmáticas, sintagmáticas, culturales e idiosincrásicas según su conocimiento.

En este sentido, los modelos de acceso léxico tratan de explicar los procesos de comprensión del *input* y cómo se añaden nuevos elementos al lexicón mental. De entre las varias teorías (*teoría de rasgos semánticos*, *teoría de prototipos*, *teoría de las redes semánticas*), se prefiere la *hipótesis de redes semánticas* debido al comportamiento combinatorio de las unidades léxicas en el lexicón.

De acuerdo con esto, es necesaria la intervención de ciertos factores que hagan el léxico recordable (Rufat, 2018: 85-86) y recuperable de la memoria a largo plazo: (i) exposición repetida a un mismo *input* en diferentes contextos; (ii) reglas mnemotécnicas; (iii) vinculación emocional; (iv) esfuerzo cognitivo. En lo que atañe a las estrategias para auxiliar a la memoria, que facilitan la fijación y la recuperación de la pieza léxica, cabe recordar que el lexicón no solo alberga las palabras, sino también las reglas que organizan su comportamiento, por eso se insiste en la necesidad de implementar la descripción semántica de una entrada léxica también con sus posibilidades combinatorias (Pérez Serrano, 2017).

En suma, el vocabulario se adquiere en contexto, con la mayor cantidad y calidad de información referencial posible. Los aprendices deben estudiar la morfología, las reglas de combinación, los significados y los usos de las palabras para que se establezcan asociaciones más firmes entre la estructura fonológica y léxico-conceptual de la pieza léxica. Por su parte, los profesores deben comprometerse a seleccionar el vocabulario tanto por su nivel de dificultad como por la relación con el conocimiento del mundo y del tema que tenga su alumnado.

¹⁷ Véase §6.

procesamiento y organización del lexicón aplicando los modelos generales de adquisición de la lengua a la adquisición del vocabulario (Appel, 1996 *apud* Morante Vallejo, 2005: 30-31). A todo esto, se añade una concepción innovadora del aprendizaje de la gramática, que deja de ser independiente y autónoma con respecto al vocabulario.

En resumidas cuentas, se ponen los cimientos de enfoques metodológicos de corte lexicalista, que sientan las bases de los desarrollos futuros de las propuestas de enseñanza, asumiendo que el léxico ejerce un rol fundamental en la producción del lenguaje, por lo que se pone en el centro de la acción docente y del proceso de aprendizaje.

Según Coady (1997 *apud* Morante Vallejo, 2005: 31-32) la enseñanza del léxico encaja en aquellas corrientes que se van sumando a las propuestas didácticas descritas hasta ahora y desembocan en el método lexicalista pionero, el *Lexical Approach* de Lewis (1993).

Pero, antes de detenernos en este enfoque, conviene mencionar otras propuestas que tienen en cuenta la *Teoría de los Actos de Habla* de Austin (1962), según las cuales el lenguaje no es solo un medio de expresión, sino una herramienta para llevar a cabo actos, o funciones.

En línea con la pragmática, la lengua no es un sistema de formas o de reglas, sino un aspecto de la acción humana que permite “realizar cosas”. Por otra parte, de acuerdo con la etnografía de la comunicación, Hymes (1972) afirma que es necesario recurrir a una visión más amplia de la lengua y de la competencia lingüística que incluya la capacidad de producir oraciones correctas y adecuadas al contexto (situación, participantes, finalidades, secuencia de actos, clave, normas y género).

Estos planteamientos confluyen en la propuesta de Halliday (1973, 1975), quien afirma que solo se puede entender el funcionamiento de la lengua como sistema si se tienen en cuenta los propósitos o funciones de la comunicación desde el momento en el que se trata de un medio de actuar en sociedad y, así, “solo a través de la interacción significativa en contextos de comunicación puede el aprendiz ampliar y profundizar su capacidad de uso de la lengua” (Kumaravadivelu, 2006a: 8).

3.3.1 *El Enfoque por tareas*

El denominado Enfoque por tareas es una variante del Enfoque Comunicativo, comparte todos los principios teóricos que subyacen a la relevancia de la comunicación

en la ASL que se han señalado anteriormente, pero introduce el concepto clave de *tarea*¹⁸ en la secuenciación didáctica. En este sentido, la finalidad y los medios del currículo son los mismos, a saber, alumnos que aprenden a comunicar comunicando.

La idea central es trasladar las acciones de la vida cotidiana al aula de idiomas de manera que los aprendientes reproduzcan actos representativos de los que podrían tener que resolver en situaciones reales. Estas acciones coinciden con el punto de partida, el objetivo y el eje en torno al cual gira todo el proceso de aprendizaje. La programación prevé una serie de pasos intermedios (*tareas posibilitadoras*) que culminan en una *tarea final*.

En este marco, el léxico es uno de los componentes principales que guían la enseñanza vista la concepción que tiene en ese momento: comunicar es transmitir significado, y el vocabulario contiene más significado que la gramática.

3.3.2 El Enfoque Léxico

El Enfoque Léxico (Lewis, 1993, 1997, 2000) se manifiesta como la aportación más novedosa entre las propuestas de corte lexicalista planteadas hasta la fecha. Higuera (2006) resume uno de sus principales patrones, destacando la primacía de la enseñanza del léxico en detrimento de la gramática.

Se defiende el aprendizaje del vocabulario como un proceso cualitativo y se basa su conocimiento tanto en el significado (contextual y cotextual), como en los registros, las restricciones de uso y las posibilidades combinatorias.

Es más, se concede mayor atención a la planificación y presentación del léxico y se aboga por la creación de oportunidades tanto para el aprendizaje directo (explícito) como indirecto (incidental)¹⁹. De acuerdo con esta teoría, el aprendizaje explícito está centrado en aprender palabras nuevas, pero, como requiere una gran inversión de tiempo,

¹⁸ Sobre este concepto la bibliografía es muy abundante, aquí nos limitamos a asumir que

[u]na tarea se define como cualquier acción intencionada que un individuo considera necesaria para conseguir un resultado concreto en cuanto a la resolución de un problema, el cumplimiento de una obligación o la consecución de un objetivo [...] como, por ejemplo, mover un armario, escribir un libro, obtener determinadas condiciones en la negociación de un contrato, jugar una partida de cartas, pedir comida en un restaurante, traducir un texto de una lengua extranjera o elaborar un periódico escolar mediante trabajo en grupos. (MCER, 2002: 9-10)

¹⁹ El aprendizaje directo almacena la información en la memoria intermedia y parece más rentable a corto plazo; mientras que el aprendizaje indirecto permite registrar las palabras junto con las relaciones que establecen entre sí, por lo que a largo plazo resulta mejor.

resultaría imposible adquirir de esta manera todo el vocabulario de una LE. Por esta razón, se aprovechan los resultados del aprendizaje incidental (gracias a la exposición a un *input* adecuado) aunque se trata de un proceso más lento y menos sistemático.

Todo esto supone también un cambio de la acción docente, el rechazo de la secuencia didáctica tradicional *presentación-práctica-producción* a favor de las fases *observar, hipotizar y experimentar*, según una secuenciación que se traduce en un aprendizaje ordenado, en línea con la sucesión *reconocimiento, recuerdo, comprensión y uso* (Read, 2000).

Otro de los fundamentos del *Lexical Approach*, posiblemente el más polémico con respecto a las tendencias precedentes, asume el *principio de idiomaticidad* de Sinclair (1991): parece que una lengua dispone de un número de bloques semiconstruidos, por alguna razón fijados en la norma, que el usuario puede combinar cuando habla o escribe. En palabras de Pérez Serrano (2017: 15), estas unidades “se almacenan en el lexicón de forma holística y se recuperan y procesan en bloque” y Troitiño (2017: 145) afirma²⁰:

Los hablantes de una lengua no comprendemos ni producimos enunciados o textos palabra por palabra (Schmitt y Carter, 2000), sino que procesamos segmentos completos de palabras. Este fenómeno de procesamiento, denominado *chunking* (Ellis, 2005), permite explicar nuestra fluidez en el manejo de la lengua.

En español se utilizan las expresiones *bloques léxicos, secuencias formulaicas, bloques prefabricados, frases léxicas* para referirse a este concepto que abarca otros tipos de unidades como las *colocaciones*; las *expresiones institucionalizadas, habitualizadas o estandarizadas*; las *expresiones idiomáticas* que responden al principio de idiomaticidad:

Secuencias de palabras –continuas o discontinuas– que parecen prefabricadas, es decir, que se almacenan y se recuperan de la memoria como un todo, en lugar de generarse desde cero cada vez que se producen o estar sujetas a las reglas de la gramática. (Wray, 2002: 9 *apud* Pérez Serrano, 2017: 14)

Desde este punto de vista aprender una lengua equivale a saber segmentar el léxico en unidades complejas (Higueras, 2006), que es oportuno presentar ya desde niveles iniciales de forma que los alumnos puedan reconocerlas, separarlas y combinarlas. Este tipo de conocimiento conlleva en sí la enseñanza de la gramática puesto que “la lengua no consiste en gramática lexicalizada, sino en léxico gramaticalizado” (Lewis, 1993: 51),

²⁰ Sin embargo, antes que Lewis, ya se opinaba que la producción de lenguaje consistía en segmentos léxicos prefabricados, que reciben el nombre de *chunks* (Nattinger y De Carrico, 1992).

es decir, está formada por un conjunto de piezas léxicas relacionadas mediante estructuras.

La propuesta de Lewis marca un punto de inflexión en lo referente a la concepción de la unidad de aprendizaje, pues, la palabra deja de ser la unidad básica del nivel léxico: se dirige la atención del aprendiz hacia aquellos conjuntos de lexías que suelen aparecer juntas, no se trata de aprender simples palabras. De ahora en adelante, se aboga por la enseñanza de la unidad léxica, que sustituye la unidad palabra “limitada”.

Son muchos los autores que siguen estos nuevos principios, por esta razón Pérez Serrano (2017) habla en plural de *enfoques léxicos*. Entre estos, mencionamos a Bogaards (1996; 2001) por ofrecer una clara distinción entre lo que es una palabra (unidad formal) y una unidad léxica (unidad de significado) ya que “[...] durante el aprendizaje del vocabulario de una segunda lengua, se pueden dar diversos casos” (Morante Vallejo, 2005: 67):

- (i) aprender una forma desconocida y un significado nuevo;
- (ii) aprender un nuevo significado para una forma que ya se conoce;
- (iii) aprender un nuevo significado para una combinación de formas ya conocidas;
- (iv) aprender relaciones semánticas y morfológicas entre unidades léxicas;
- (v) aprender usos correctos de unidades léxicas al nivel gramatical y pragmático;
- (vi) aprender las colocaciones usuales.

En este sentido, el conocimiento del vocabulario supone la combinación de formas y significados bajo la estructura de cadenas multipalabras prefabricadas que facilitan la recuperación de las unidades léxicas y la producción correcta de eductos en la LO.

3.4 Tendencias actuales

La práctica docente actual es una continuación de los métodos y enfoques precedentes, pues se presenta como una ampliación de las propuestas ya presentes en el panorama de la enseñanza-aprendizaje de LE, conllevando un gran eclecticismo en cuanto a sus aplicaciones didácticas y, al mismo tiempo, ningún cambio significativo con respecto a los modelos anteriores²¹:

²¹ Algunos autores apuntan a la “era post-método”: parece que en el aula ya no se utilizan metodologías delimitadas, se ha pasado a unas propuestas más flexibles y organizadas conforme a las necesidades y expectativas del alumnado (Sánchez Pérez, 1992: 344). En los últimos años, se ha empezado a usar también

El siglo XXI está profundamente enraizado en la situación heredada del siglo XX, como no podía ser de otra manera. La metodología comunicativa sigue aún vigente. Pero emergen nuevas ideas, nuevos enfoques, nuevas maneras de hacer las cosas y nuevas perspectivas. (Sánchez Pérez, 1992: 345).

En 2001, la publicación del MCER por parte del Consejo de Europa supone un hito importante en el campo de la enseñanza de LE desde el momento en el que se deja de pensar en el aprendiz como a un mero usuario de la lengua porque toma el “carga” de agente social.

El proceso de ASL pasa a ser integrador de toda la persona y está finalizado al incremento de aquellas competencias que posibilitan actuar como miembro de la sociedad, se trata de “la suma de conocimientos, destrezas y características individuales que permiten a una persona realizar acciones” (MCER, 2002: 20), a saber: los recursos (medios) que permiten al alumno realizar acciones (objetivo). En consecuencia, el desarrollo de “una serie de competencias, tanto generales como competencias comunicativas lingüísticas, en particular” (MCER, 2002: 9) es uno de los factores esenciales²².

El vocabulario sigue ganando importancia en la perspectiva historiográfica que hemos adoptado. La competencia léxica, que tiene ahora el mismo peso de otros componentes lingüísticos, se alcanza de manera dual: se aboga por un desarrollo cuantitativo y cualitativo del bagaje léxico. Además, es una habilidad en sí misma que requiere metodologías, materiales y estrategias *ad hoc*. Cervero y Pichardo (2000: 110) asumen que dichas estrategias pueden ser enseñadas, si se concibe la enseñanza como proceso de concienciación del estudiantado, desde el momento en el que se trata de conocimientos ya utilizados por los alumnos en la LM, si bien de forma inconsciente.

A este propósito, el Consejo de Europa y el Instituto Cervantes, que en 2006 se ha encargado de redactar el *Plan Curricular del Instituto Cervantes* (PCIC) como

el término “postcomunicativo” para referirse a modelos metodológicos surgidos con el objetivo de facilitar la adquisición de la competencia comunicativa en LE (Baralo y Estaire, 2010). Dentro de estas tendencias se encajan, por ejemplo, el Enfoque Ecléctico, el Enfoque Accional y el Enfoque Competencial.

²² Las *competencias generales* abarcan el conocimiento declarativo (*saber*); las destrezas y las habilidades (*saber hacer*); la competencia existencial (*saber ser*); la capacidad de aprender (*saber aprender*). Del otro lado, las *competencias comunicativas* incluyen la *competencia lingüística* (la capacidad de utilizar correctamente la lengua gracias a una competencia léxica, gramatical, semántica, fonológica, ortográfica y ortoépica); la *competencia sociolingüística* (la capacidad de adecuación a la situación comunicativa y al contexto cultural); la *competencia pragmática* (la capacidad para resolver situaciones comunicativas y de alcanzar fluidez y precisión en el discurso como garantía del éxito funcional, es decir, la suma de las competencias discursiva, funcional y organizativa).

documento oficial de referencia para el aprendizaje del español, atribuyen un rol nuclear al vocabulario. Asumen en sus directrices los principios teóricos y metodológicos del Método Nocional-funcional y del Enfoque Léxico, abordan detenidamente el asunto del componente léxico y del desarrollo de su respectiva competencia, enmarcada en la enseñanza comunicativa de la lengua:

[...] en la relación de los objetivos generales de enseñanza y aprendizaje por niveles y en los doce inventarios de descripción de material, que derivan en conjunto de un análisis de los distintos componentes –gramatical, pragmático-discursivo, nocional, cultural y de aprendizaje– que conforman una visión amplia y matizada de la lengua desde la perspectiva de la comunicación. (PCIC, 2006: Introducción general)

Para ello, propician inventarios de léxico, o nociones²³, bajo la forma de listas de carácter orientativo, organizadas por áreas temáticas y por nivel lingüístico, que sirvan de guía al profesorado y permitan a los aprendices desenvolverse en toda situación comunicativa (ámbito personal, público, profesional, educativo).

A estas alturas, “es incuestionable que el desarrollo de la competencia léxico-semántica juega un papel primordial en el proceso de aprendizaje de una lengua extranjera o segunda lengua” (Pérez Serrano, 2017: 7) y tiene repercusiones tanto en la competencia lingüística como en la sociolingüística y la pragmática (MCER, 2002) ya que conocer una palabra implica saber cuál son su forma, su significado, su uso (Nation, 2001). Se concibe como una adquisición de tipo transversal que incluye el uso de estrategias específicas. Entre estas, facilitar el acceso a la combinatoria léxica de la LO (la preocupación principal de Lewis (2000) en la aplicación práctica del *Lexical Approach*) ha llevado a una

²³ La Introducción general del PCIC (2006) aclara que

el componente nocional se sitúa en línea con los análisis de la lengua, desarrollados sobre todo a partir de los 70, que identifican una serie de categorías de carácter semántico-gramatical que no son, propiamente, unidades léxicas –entendidas como elementos discretos portadores de significado concreto–, sino conceptos descriptivos generales que pueden incluir grupos de unidades léxicas. De este modo, el concepto de noción trasciende el concepto tradicional de vocabulario. Las nociones generales son las que un hablante puede necesitar cualquiera que sea el contexto en el que se dé el acto de comunicación, por lo que están referidas en gran medida a conceptos abstractos, relacionales –como tiempo, cantidad o lugar–, mientras que las nociones específicas son las que tienen que ver con detalles más concretos del «aquí y ahora» de la comunicación y se relacionan con transacciones, interacciones o temas determinados. La articulación de las funciones –que presentan un tratamiento sistemático de las intenciones comunicativas del hablante– y las nociones –los conceptos a los que el hablante se refiere al realizar un acto de habla– constituyó en su día el eje de los programas nocio-funcionales, que se incorporaron en seguida a los materiales didácticos y a los programas de enseñanza de lenguas y dieron carta de naturaleza al Enfoque Comunicativo. La base de experiencia adquirida en estas últimas décadas con este tipo de programas permite avalar la rentabilidad pedagógica de las descripciones de carácter nocional frente a los tradicionales repertorios de léxico o vocabulario.

“optimización del enfoque léxico” (Higueras, 2019: 146) y al planteamiento de nuevas propuestas.

Emblemático es el trabajo de Boers y Lindstromberg (2009), quienes basan su teoría sobre el aprendizaje horizontal de los bloques léxicos mediante un amplio abanico de técnicas aplicables en el aula, que favorecen un dominio lingüístico y una fluidez parecidos al habla nativa, además de tener efectos positivos en la entonación, la pronunciación y la evitación de errores gramaticales. Parece que estos elementos no se adquieren naturalmente con el tiempo, sino que son el fruto de la incorporación de estas unidades complejas en el lexicón mental, pues, “[...] las secuencias formulaicas (SF) forman parte de los segmentos léxicos cuyo tratamiento en el aula, según el Enfoque Léxico, mejora la naturalidad y fluidez de los aprendientes.” (Contreras y Rey, 2018: 20).

Desde la década final del siglo XX se duda de que un método único sea la respuesta a la pregunta de cómo enseñar una LE. Se difunde la idea de que este *método* (o *enfoque*) “debería dejar de ser considerado como un constructo valioso o viable y de que lo que se necesita no es un método alternativo, sino una alternativa al método” (Kumaravadivelu, 2006b: 67).

Asimismo, hay que tener en consideración la influencia de las disciplinas académicas y de la investigación en lingüística aplicada que han modificado los conceptos de lengua y aprendizaje, que a su vez han provocado cambios en la didáctica.

4. MATERIALES PARA LA DIDÁCTICA DE ELE

4.1 Sobre los manuales

Ilustramos ahora la evolución de la didáctica de LE mediante una breve reseña de algunos manuales de ELE concebidos a partir de los postulados de los métodos y enfoques que acabamos de exponer. La pregunta a la que queremos dar respuesta es: *¿qué hay detrás de un manual?* Comentamos, a modo de ejemplo, aquellos que destacan los cambios más evidentes y representan siete de estas corrientes metodológicas²⁴.

²⁴ No cabe en nuestras intenciones discutir el valor de ninguno de estos manuales y tampoco pretendemos hacer un repaso exhaustivo de la oferta del mercado editorial: nos limitamos a ofrecer una panorámica de la aplicación de las teorías lingüísticas en los enfoques que subyacen en la creación de los manuales de ELE.

De nuevo, nos basamos en el intervalo temporal de Morante Vallejo (2005), esto es, conforme a aquellas etapas que han llevado, paulatinamente, a una nueva concepción de la enseñanza-aprendizaje del vocabulario.

4.1.1 *Revisión de un manual según el Método Tradicional*

El conocimiento explícito de la gramática de la LE, en detrimento del vocabulario, es el objetivo principal de este método aplicado a la enseñanza. Esto se ve claramente en el manual *Curso breve de español para extranjeros* (1970 [1954]), donde la lengua queda reducida a las reglas gramaticales y estructurales que permiten la decodificación de textos literarios. Pese a que las unidades didácticas (UD) estén organizadas por campos semánticos y empiecen con una lista de palabra y una muestra de la LO, los contenidos de aprendizaje y los ejercicios predominantes son gramaticales y de traducción. Se aboga por el uso de la LM del alumnado y una reflexión contrastiva con la LE.

En la parte conclusiva del libro, se plantea una ampliación del vocabulario mediante nuevas listas de palabras y refranes, finalizadas a la memorización. Asimismo, el apéndice presenta algunos verbos españoles y un glosario multilingüe (italiano, francés, inglés, alemán) de las palabras que aparecen en la parte gramatical.

Dentro de la misma corriente, el manual *Español conversacional. Ejercicios de vocabulario* (1966) está finalizado a profundizar el bagaje léxico de los discentes y lo hace intentando acercarse a la lengua hablada, como declara su autor en el prólogo. Se plantea el aprendizaje del español común, “de la calle”, pero, al fin y al cabo, se trata de la mera recopilación de listas de palabras, locuciones y colocaciones, ordenadas por áreas temáticas sin la propuesta de actividades prácticas significativas, por lo que la memorización de unidades léxicas parece la única tarea que el aprendiz puede llevar a cabo.

4.1.2 *Revisión de un manual según el Método Situacional*

Los planteamientos lingüísticos de corte estructural cambian completamente la configuración del manual de ELE: desaparecen las listas de palabras y se diseñan materiales que puedan ayudar al desarrollo de la comunicación. En los manuales que traducen en práctica este método la cantidad de vocabulario presentada es limitada y se repite a lo largo de la UD.

En primer lugar, *Español 2000* (1987) se organiza en función de un contexto situacional cotidiano, graduado según el nivel de los aprendices, cuyo propósito es propiciar los mecanismos para acceder a la lengua y a la cultura española, según aseguran los autores. Los contenidos principales son la gramática y las estructuras lingüísticas, las cuales permiten formar hábitos mediante el binomio estímulo-respuesta. Los ejercicios, de práctica controlada y de comprensión, siguen los esquemas gramaticales propuestos a lo largo de la UD.

Lo gramatical es el punto de partida de cada actividad ya que los textos y los diálogos, que abren cada sección, presentan estructuras y, solo secundariamente, permiten adquirir léxico (siempre supeditado a la gramática): el uso de la LO que debe hacer el alumnado consiste en manejar dichas estructuras, el componente privilegiado por la enseñanza.

En la última parte del libro, se dan breves informaciones sobre el alfabeto (lectura, pronunciación, acentuación, entonación) y los verbos (auxiliares, regulares, irregulares).

En el caso específico del aprendizaje de ELE por parte de estudiantes italianos, disponemos de *Español situacional* (1981). Este manual se divide en UD graduadas y diseñadas a partir de situaciones de la vida cotidiana que faciliten el acercamiento a la lengua hablada, cuyos contenidos se centran, de nuevo, en estructuras de la LO.

Una clase-ejemplo se secuencia como sigue (Calvi *et al.*, 1981: 8):

1. lectura de un texto por parte del profesor;
2. lectura del mismo texto por parte de los alumnos;
3. explicación de significados de palabras y expresiones (sin recurrir al italiano, en la medida de lo posible);
4. explicación de las estructuras lingüísticas;
5. utilización de los elementos previos para una tarea de conversación.

En casa, el alumnado realiza ejercicios de gramática y traducción, que se encuentran en uno de los apartados finales del libro, junto a una sección dedicada a los verbos.

La clase sucesiva prevé:

1. la corrección de los ejercicios y el aclaramiento de dudas;
2. una actividad de dictado;
3. la lectura y explicación del texto.

La enseñanza sigue focalizándose en reglas gramaticales y estructurales ya que no se tratan directamente los contenidos léxicos, aunque el intercambio oral gane cierto protagonismo durante las actividades en aula.

4.1.3 *Revisión de un manual según el Método Nocio-funcional*

Los planteamientos de la corriente nocio-funcional constituyen un nuevo cambio de paradigma en lo que atañe a la enseñanza de LE que se interesa, ahora, por facilitar a los aprendices aquellos recursos necesarios para poder expresar funciones en la LO, es decir, aplicar los elementos lingüísticos de forma adecuada al contexto comunicativo.

El manual *Viaje al Español* (1993) sigue estos postulados fijando su atención en los requisitos comunicativos, no en las estructuras lingüísticas.

Las secuencias didácticas se agrupan en virtud de los actos de habla: todo contenido se inserta en el libro para satisfacer necesidades expresivas. Las reglas gramaticales (presentadas detenidamente en el último apartado del manual, a modo de compendio) y de uso son muy breves y están organizadas *ad hoc* siguiendo el esquema funcional. Por su parte, el léxico no tiene una sección dedicada, se aprende del contexto. Solo en la última parte del libro tiene un apartado propio, donde se presenta la lista de las palabras aparecidas. La batería de ejercicios es variada y encaminada a la expresión de las funciones orales (individuales, por pareja, en grupo), auditivas (con respuesta oral o escrita) y escritas.

Del mismo modo, *Itinerarios por el español* (1993), diseñado para un público italiano, pretende cubrir las necesidades comunicativas del estudiantado y también dedica un apartado al léxico (aunque las propuestas siguen basándose en la memorización de unidades léxicas).

Las funciones del lenguaje son el contenido de aprendizaje predominante y se trabajan por medio de reglas estructurales y morfosintácticas, con especial atención a la contrastividad entre LM y LE. Asimismo, se tienen en consideración las variedades y los registros de uso presentando muestras de lengua auténtica.

Tres partes componen las UD: (i) presentación de los contenidos y ejercicios de comprensión; (ii) exposición de las funciones y breves nociones gramaticales (un compendio final las recoge con más detenimiento); (iii) sección de léxico (falsos amigos, expresiones idiomáticas y palabras ligada al campo semántico de la UD).

4.1.4 Revisión de un manual según el Enfoque Comunicativo

Se pone de relieve la función de medio de comunicación de la lengua: la realización de actividades comunicativas facilita el alcance de la competencia comunicativa del alumnado, esto es, el objetivo de la ASL. Por eso, se reconsideran los contenidos de aprendizaje y los procedimientos didácticos para posibilitar el desarrollo de aquellas destrezas que permitan comunicarse al aprendiz. Para ello, es esencial ofrecer un *input* que sea representativo y auténtico del habla nativa, como propone *Español sin fronteras* (1999).

Cada UD parte de un pretexto que introduce los contenidos (funcionales, gramaticales, léxicos, culturales) que se trabajan después, ya que parece que es la formación de hipótesis lo que conduce a la competencia lingüística (la formación de hábitos deja de ser el foco de la enseñanza), por lo que las actividades se organizan según la secuencia *presentación-práctica-producción*.

Como la lengua es a la vez comunicación y sistema, se hace necesario contextualizar sus elementos conforme a la función comunicativa que se quiere cumplir. El vocabulario se aprende a partir del contexto y de las muestras lingüísticas que componen la UD. La gramática se explica y estudia a partir de los usos que se quieren hacer de la LE, de la función expresada²⁵. El resultado es un material variado formado por actividades de comprensión lectora y auditiva, ejercicios gramaticales, huecos por rellenar y se fomentan las actividades interactivas.

¡Claro que sí! (1989) está dirigido a estudiantes italófonos, y aplica las directrices del Enfoque Comunicativo “puro”, pues el procedimiento operativo se funda en el desarrollo de actividades de intercambio, cuyo objetivo privilegiado son la interacción y la simulación de una situación comunicativa concreta en el aula que favorece la creatividad del alumno. La atención reside en la transmisión de significados: se parte de criterios pragmáticos y discursivos para conseguir la habilidad de comunicarse.

Un diálogo abre cada UD introduciendo y contextualizando las estructuras lingüísticas que forman parte de la secuencia didáctica. Los contenidos léxicos se estudian mediante diferentes ejercicios y juegos finalizados a estimular la curiosidad de los

²⁵ Cabe señalar que se trata de un manual comunicativo y gramatical, según afirman los autores en la *Presentación*, que otorga la misma importancia a los elementos gramaticales y a los funcionales, motivos por el cual Santos Gargallo (2017 [1999]: 57 y 71-78) lo caracteriza como *enfoque comunicativo moderado*.

usuarios. Se incluyen, además, cuadros de gramática española y contrastiva español-italiano que posibilitan el aprendizaje de las estructuras morfosintácticas, así como documentos reales para la ampliación de los conocimientos culturales.

4.1.5 *Revisión de un manual según el Enfoque por tareas*

La secuenciación didáctica propia de este enfoque parte de la presentación de una tarea final que los estudiantes pueden realizar después de unos pasos previos (las tareas posibilitadoras) que les proporcionan las formas y los recursos necesarios para ella, integrando todo tipo de destreza lingüística y fomentando la cooperación del grupo.

Gente (1997) es el manual pionero en la aplicación del Enfoque por tareas diseñado para el aprendizaje del español con una clara orientación comunicativa: las funciones comunicativas están privilegiadas y determinan tanto la organización de la UD, como las destrezas trabajadas (con especial atención a la comprensión y la expresión oral). El orden de las actividades no sigue un criterio fijo en la progresión de los contenidos más simples a los más complejos.

En la subsiguiente edición de 2004 (*Gente 2. Nueva edición*) se sigue el mismo esquema, las UD se dividen en los apartados a continuación:

1. *Entrar en materia*, permite acercarse al tema y al vocabulario;
2. *En contexto*, desarrolla la comprensión mediante la exposición a la LE;
3. *Formas y Recursos*, trabaja elementos gramaticales;
4. *Tareas*, incluye todo tipo de actividades y ejercicios (la tarea final tiene carácter global ya que abarca todos los distintos componentes estudiados anteriormente: gramática, léxico, fonética, pragmática y cultura);
5. *Mundos en contacto*, propone una reflexión sobre la cultura hispana.

Por lo general, los contenidos lingüísticos intervienen secundariamente en las actividades comunicativas, implicando un proceso de inducción por parte del alumno (al final del libro se incluye una sección de gramática).

Por su parte, la competencia léxica desempeña un papel importante ya que se le dedica un apartado propio (*Vocabulario*) diseñado en función de campos semánticos lo más cercanos posible a los intereses del alumnado.

El léxico se presenta contextualizado y se aprende de manera significativa y variada, abandonando la memorización.

4.1.6 Revisión de un manual según el Enfoque Léxico

La aparición del Enfoque Léxico modifica, otra vez, la concepción de la didáctica de las LE: se confiere al vocabulario un rol central. El nuevo tratamiento que se otorga al léxico cambia el modo de dar clase y diseñar materiales, pues, se llama la atención sobre la adquisición de bloques léxicos mediante baterías de ejercicios finalizadas al aprendizaje implícito, explícito y de producción oral y escrita.

De entre los manuales que reflejan esta ideología, *Bitácora* (2016) es representativo (también de las tendencias eclécticas más recientes) colocándose a medio camino entre el Enfoque Léxico y el Enfoque orientado a la acción²⁶.

Dedica mucho espacio al vocabulario, se trabajan colocaciones y unidades multipalabra a partir de la idea de que la lengua es un conjunto de elementos léxicos relacionados en estructuras sintácticas. La ASL radica en un proceso de aprendizaje cualitativo del léxico finalizado al conocimiento y uso correcto de los *chunks* (destacan los apartados *Palabras en compañía* y el *Diccionario de construcciones verbales*).

Las UD disponen de una sección inicial en la que se establece un primer contacto con los contenidos. Se presenta el vocabulario esencial que favorece la recuperación de conocimientos previos y la activación de estrategias de inferencia. A continuación, una parte importante de los ejercicios está enfocada en el desarrollo de la competencia léxica desde múltiples perspectivas. Se ofrecen muestras auténticas de lengua (adecuadas al nivel de los usuarios) para entrenar la habilidad de comprensión y propiciar ejemplos de uso que servirán en las tareas finales. Cada UD cuenta, además, con una *Agenda de aprendizaje* donde los aprendices pueden tomar notas sobre sus progresos. Efectivamente, los propósitos de los autores alientan a la reflexión metalingüística y contrastiva con la LM y tienen en consideración los intereses del alumnado de modo que el proceso de aprendizaje sea realmente significativo.

4.1.7 Revisión de un manual según el Enfoque Ecléctico

Las tendencias actuales, es sabido, ponen el aprendiz en el centro de la enseñanza, se presta especial atención al desarrollo de aquellas estrategias que fomenten la ASL, la

²⁶ El Enfoque por tareas puede considerarse su precursor ya que el Enfoque Accional coincide con una versión renovada y ampliada a partir de las directrices del MCER, que considera al aprendiz un agente social que realiza otras tareas por encima de las lingüísticas.

autonomía e involucrar al alumno en la toma de decisiones en lo referente a la acción didáctica. El alumno es un agente social, por lo que la integración de las cuatro destrezas lingüísticas, que permiten actuar adecuadamente en la sociedad, es un paso imprescindible.

Se requiere que un manual contenga toda la información lingüística y cultural posible para formar un usuario competente de la LO²⁷. Las prácticas pedagógicas adoptan una visión eclética de la enseñanza, cuyos contenidos y actividades de aprendizaje resultan variados porque mezclan las directrices de los Enfoques Comunicativo, Accional y Competencial, entre otros.

Es este el caso de *Vía rápida* (2011), que abarca los niveles A1-B1 del MCER. Este manual, tras la exposición de los contenidos de la UD, las funciones comunicativas, los recursos léxicos y gramaticales que se estudiarán, propone actividades enfocadas al desarrollo de la capacidad productiva del alumno. Después, se trabajan la comprensión auditiva y lectora. A continuación, se aprende el léxico en una sección propia.

Las últimas páginas de la UD agrupan las reglas gramaticales aparecidas en cada una (la sección final del libro presenta, además, una recapitulación sobre cuestiones de fonología, ortografía y ortoépica).

En esta misma línea, *Español ELElab* (2013) permite el pleno acceso a un nivel B2. El foco de este manual es que los usuarios aprendan a comunicar y actuar eficaz y adecuadamente en español ya que la finalidad del ASL consiste en hacer cosas *con y en* la LE, según los autores. Así pues, los postulados del Enfoque Comunicativo y del Enfoque Accional guían las diferentes actividades, tareas y proyectos (lingüísticos y no lingüísticos) fomentando la autonomía del aprendiz.

La secuenciación didáctica es modular y sigue los siguientes pasos:

1. presentación de un proyecto inicial;
2. comprensión lectora;
3. comprensión auditiva;
4. presentación de formas y recursos lingüísticos;
5. expresión e interacción orales;

²⁷ La dimensión intercultural es un elemento crucial en los procesos de enseñanza y aprendizaje de lenguas extranjeras, se trata de la “competencia en comunicación intercultural” (García Parejo, 2018: 311) acuñado a raíz de la concepción de la ASL como un proceso de socialización que se desarrolla dentro de una red social (Ambadiang y García Parejo, 2006).

6. expresión e interacción escritas;
7. presentación de contenidos culturales.

Estas breves notas permiten observar el eclecticismo del manual y, en general, de la enseñanza que actualmente se propone en el aula de idiomas, que aboga por la combinación de todas las destrezas enumeradas arriba, encaminadas al dominio comunicativo, sociocultural, funcional y formal de la lengua española.

Se amplía la noción de *tarea* (procedente del Enfoque por tareas): no se detecta un contenido predominante, todo tipo de información lingüística (funcional, estructural, gramatical, discursiva, léxica, sociocultural, estratégica) tiene la misma importancia para instruir un usuario avanzado de ELE.

4.2 Sobre el diccionario

A luz de este recorrido historiográfico sobre la evolución del tratamiento del léxico en la enseñanza de lenguas y su aplicación en los manuales de ELE, tomamos como hilo conductor las palabras de Higuera (2017: 19) para introducir otro elemento protagonista de las aulas de LE, el diccionario:

[...] podemos afirmar que es mucho lo ya recorrido y avanzado. Así mismo, es alentador saber que hay bastantes áreas de mejora y gran cantidad de proyectos en los que embarcarse, tanto desde el punto de vista de la investigación como de la formación o el diseño de materiales (didácticos o lexicográficos), para conseguir que el léxico ocupe el lugar que se merece en el día a día de la enseñanza de ELE.

Nos parece un paso obligado mencionar este recurso y destacar la continua y necesaria interacción entre lexicografía y didáctica a partir de las emblemáticas primeras líneas de la *Introducción* del volumen coral editado por Domínguez Vázquez y Sanmarco Bande (2017: 5) sobre este asunto, en la que se hace hincapié en las ventajas de estrechar este vínculo:

[...] Los estudios desarrollados en ambas disciplinas muestran una notable desatención teórico-metodológica en cuanto al diccionario como recurso didáctico, cuestión trascendental desde la enseñanza y el aprendizaje de una segunda lengua. Por esta razón, consideramos imprescindible realizar una actualización de los estudios centrados en las aplicaciones didácticas del diccionario, con el objeto de ofrecer un recurso actualizado a quienes desean enseñar o aprender una segunda lengua.

Se pone manifiesto la importancia de la “didactización de los contenidos lexicográficos para su aplicación en el aula” desde el momento en el que el aprendizaje

del léxico no coincide con la mera apropiación y acumulación de unidades aisladas o descontextualizadas.

El conocimiento del vocabulario trae consigo el desarrollo de la competencia léxica que se compone de todas las propiedades de las palabras (formales, sintácticas, semánticas), sus variantes sociolingüísticas y su uso pragmático, a los que se añade un conocimiento enciclopédico, que permiten el alcance de aquellas habilidades que posibilitan la comunicación.

4.2.1 *El uso del diccionario en el aula de ELE*

Afirma Alvar Ezquerro (2003) que aprender el léxico de ELE significa estudiar: la ortofonía; la ortografía; las reglas gramaticales y sintácticas; las variedades diastráticas, diatópicas, diafásicas, diacrónicas; las relaciones semánticas; la morfología flexiva y la combinatoria de las palabras (palabras compuestas y fraseología)²⁸.

Un buen diccionario para no nativos, por tanto, debe incluir la información gramatical, cultural y diatópica de la LO.

Un aprendiente extranjero no conoce todas las acepciones y aplicaciones de una pieza léxica ya que no todo se puede aprender (y enseñar) a la vez, por eso, el apoyo del diccionario es crucial durante el aprendizaje por sus múltiples ventajas²⁹:

En la enseñanza-aprendizaje del léxico de la lengua [...] el diccionario, como obra en la que se recogen palabras, puede prestarnos una inestimable ayuda, aunque no hay un diccionario válido para todas las necesidades, no en vano el diccionario es un instrumento, y como tal instrumento está destinado a unos fines concretos, distintos en cada uno de ellos. (Alvar Ezquerro, 2003: 12)

Nadie duda de que es imprescindible para aprender una LE, siempre ha sido fiel compañero del aprendiz: “los diccionarios son los textos más identificados con la actividad del aprendizaje lingüístico” (Castillo Peña, 2008: 377). En efecto, ayuda la

²⁸ Cfr. Penadés Martínez (1999); Higuera (2006); Barrios (2015); Velázquez Puerto (2018).

²⁹ Dado que nuestro punto de partida es la didáctica de LE, nos referimos a diccionarios monolingües debido a su función de herramienta diseñada para ofrecer información variada sobre las unidades léxicas y no ser meros inventarios de palabras. Igualmente, aludimos a los productos de la lexicografía bilingüe como puente lingüístico y contrastivo entre la LM y la LO. Y, finalmente, aunque todos estos diccionarios responden a una función pedagógica, hay algunos que se han recopilado específicamente como apoyo a la enseñanza de LE, por tanto, no descuidamos el provecho que se puede sacar del diccionario didáctico (monolingüe y bilingüe) teniendo en cuenta que el concepto de “*didactismo* en un diccionario no significa ‘calidad’, sino ‘adecuación a las necesidades y a las capacidades de los usuarios’” (Maldonado, 2017: 163). Para más detalle sobre las características micro y macroestructurales de estas obras, véase, entre otros, Rey Castillo (2018), quien dedica especial atención al diccionario monolingüe de ELE y Müller-Spitzer (2017), que plantea las propiedades de un buen diccionario *online* (definido el diccionario del futuro).

ejecución de todo tipo de tarea: lectura, comprensión oral, expresión oral y escrita, traducción (Martín García, 1999), pues tiene simultáneamente un papel decodificador y codificador. Se trata de un instrumento que encierra todo tipo de información léxica sobre las nuevas palabras³⁰, por lo que es una herramienta para el profesor de LE, junto a la gramática y el manual (Romero Gualda, 2008: 95).

En lo que atañe a las diferentes aplicaciones del diccionario en el aula de idiomas, en primer lugar, la adecuación al tipo de usuario es primordial, sobre todo en función de su nivel lingüístico (Alvar Ezquerro, 2003: 104-105) ya que a principios de la ASL se requiere que el discente domine la lengua general y solo en etapas más avanzadas consiga un conocimiento más profundo y preciso³¹.

En este ámbito, los diccionarios son “herramientas de consulta para usuarios que busquen información con el fin de resolver problemas o necesidades que surgen en una situación o contexto social específico” (Nomdedeu y Tarp, 2018: 52 *apud* Maldonado, 2019: 114). Presentan una triple aplicación práctica: son obras (i) de consulta que ayudan a entender el significado de un lexema; (ii) de reparación que permiten salir de un apuro; (iii) de aprendizaje que auxilian la producción en LE.

Para ello, las investigaciones de mercado en lo referente a los hábitos de consulta y de consumo del diccionario, mediante sondeos directos a docentes y aprendices, son esenciales para diseñar un producto que sea realmente valioso. Estos conocimientos servirán para adaptar los futuros diccionarios a las necesidades de los usuarios ya que la consideración del destinatario es un factor central en la planificación de la labor lexicográfica. Encuestas de esta naturaleza aportan informaciones sobre cuánto y cómo incide “en la definición de un proyecto lexicográfico (en el continente y en el contenido) el buen conocimiento del público objetivo al que iba dirigido (el target, en la jerga comercial)” (Maldonado, 2019: 103).

En segundo lugar, la guía del profesor es también importante para trabajar de la manera mejor posible con este instrumento. De ahí que deba tener una formación sólida en lexicografía para presentar el diccionario clara y eficazmente al alumnado, el cual podrá moverse sin dificultades entre sus páginas para averiguar el sentido de una pieza

³⁰ Se trata, *in primis*, de la definición o de la traducción de las entradas y los contextos sintácticos, paradigmáticos y sintagmáticos y, después, de la información gramatical, histórica, dialectal, literaria, normativa o estadística (Rey Castillo, 2017: 2).

³¹ Cfr. MCER (2002) y PCIC (2006).

léxica y ser guiado a su uso. En este proceso de discernimiento los usuarios deberían ir acompañados de un docente, que sea capaz de evaluar la calidad de un diccionario y discernir si es adecuado o no a las necesidades del grupo:

El empleo que hagan del diccionario los estudiantes durante el aprendizaje de la lengua dependerá de las instrucciones que le dicte su profesor. Por eso es por lo que resulta importante y necesario que el profesor conozca los diccionarios y sepa qué partido puede sacar de ellos. (Alvar Ezquerro, 2003: 13)

Pese al acuerdo de la mayor parte de los investigadores sobre la utilidad que tiene el diccionario en el aula de ELE, su uso real no ha recibido la atención que merece por parte de los docentes que, en muchos casos, confiesan abiertamente no saber qué hacer, exceptuando la consulta personal. Parece que el empleo del diccionario es una opción personal del aprendiz: algunos lo utilizan mucho y otros prácticamente nada. Análogamente, frente al complejo panorama lexicográfico, gran parte de los usuarios de hoy recurre a buscadores en internet y resuelve las dudas de manera intuitiva o casual a través de obras no fiables o simplemente inapropiadas para el tipo de usuario y la finalidad de su consulta.

Todo esto ha abierto cuestiones más o menos prácticas que es imperioso solucionar: algunos enseñantes creen que sus alumnos abusan del diccionario y descuidan otras técnicas o estrategias de aprendizaje válidas (deducir por el contexto o buscar palabras con una forma próxima); otros opinan que el alumno se distrae y pierde el hilo de la clase. Del otro lado, un problema muy frecuente es la decepción del aprendiz que no encuentra aquello que busca, lo que le lleva a dejar de usarlo: en este caso, la mediación de un profesor competente se confirma sustancial.

Superados tales inconvenientes, es provechoso un uso continuo del diccionario, no una consulta esporádica. Asimismo, las tareas pedagógicas que benefician de este recurso son muchas y se recomienda que el grupo posea el mismo diccionario para que se programen algunas sesiones de clase dedicadas a su utilización práctica (así, los resultados serán los mismos para todos).

Además de los ejercicios propiamente lingüísticos y aquellos de reflexión metalingüística, es útil trabajar la organización interna del diccionario (lo que se denomina macro y microestructura) con el objetivo de concienciar y familiarizar a los alumnos para permitirles búsquedas más rápidas y efectivas. Para ello, se debe tener en cuenta un conjunto de objetivos didácticos específicos que permitan el conocimiento y el

buen uso del diccionario, que llevan al desarrollo de aquellas destrezas relacionadas con su explotación (las destrezas referenciales). Estos objetivos podrían ser (Martínez Ezquerro, 2001: 317):

1. Conocer los diversos tipos de diccionarios: generales, enciclopédicos, bilingües, de sinónimos, ideológicos, etc.
2. Conocer la estructura del diccionario: ordenación, estructura de una entrada, informaciones contenidas, ejemplos, etc.
3. Saber interpretar la información semántica y gramatical que proporcionan las entradas de los diccionarios.
4. Comprender las definiciones y los ejemplos y saber elaborarlos.
5. Desarrollar actitudes activas respecto a la búsqueda de información y al interés por la precisión.

4.2.2 *Los trabajos sobre y para el diccionario de ELE*

Como ha ocurrido con el tratamiento del léxico, la preocupación por la lexicografía y el papel de los diccionarios en la ASL es una rama joven en el marco investigativo de la didáctica de lenguas pese a que los temas y los métodos de encuestas y estudios empíricos realizables sean muchos.

A principios de los años ochenta, se revoluciona el panorama lexicográfico con un proyecto dedicado al inglés LE que culmina con la publicación del *Collins Cobuild* (Sinclair, 1987), un diccionario monolingüe para aprendientes no nativos, cuya principal novedad consiste en definir todas las entradas utilizando solo palabras básicas (necesariamente conocidas por un aprendiz de nivel inicial) y recopilarlas exclusivamente a partir de muestras reales de la LO.

A raíz de esta innovación metodológica las casas editoras y los equipos encargados de redactar materiales lexicográficos para LE/L2 elaboran diccionarios orientados al aprendizaje de estudiantes extranjeros. Actualmente, la oferta ha crecido enormemente y existe un amplio abanico de recursos de esta índole.

El mercado español presenta numerosos diccionarios para estudiantes extranjeros que, en realidad, son reediciones de diccionarios escolares adaptados³², mientras que la primera obra creada específicamente para discentes de ELE es el *Diccionario de Español*

³² El primero en publicarse es el *Diccionario para la Enseñanza de la Lengua Española* (1995). Aunque estaba encaminado inicialmente al uso por parte de hispanohablantes nativos, en la edición posterior se ha modificado para el uso de los no nativos (de ahí que se haya cambiado el nombre en *Diccionario Español para Extranjeros para la Enseñanza de la Lengua Española*). El *Diccionario Salamanca de la Lengua Española* (1996) sigue un camino similar ya que en su segunda edición se ha incluido el subtítulo *español para extranjeros*.

para *Extranjeros* (2002) de la editorial SM y la Universidad Complutense de Madrid³³.

En la misma línea, los hallazgos del Enfoque Léxico y, más en general, de la corriente lexicalista que privilegian la enseñanza de la combinatoria léxica, han fomentado el diseño y la propuesta de los diccionarios combinatorios y de colocaciones (además de los numerosos corpus electrónicos y bancos de datos disponibles en línea) que permiten observar y aprender la lengua real en uso (Pérez Serrano, 2017: 71-76). De entre estos, el *Diccionario combinatorio del español contemporáneo. REDES* (2004) y el *Diccionario combinatorio práctico del español contemporáneo* (2006) de SM, ambos dirigidos por Bosque, pueden resultar muy útiles en distintos niveles de ELE³⁴ desde el momento en el que explican cómo utilizar una palabra³⁵.

Asimismo, cabe señalar la intervención de las nuevas tecnologías al servicio de la lexicografía y la didáctica. No se puede dudar de la nueva concepción de los procesos de enseñanza-aprendizaje si los asociamos a la revolución digital que estamos viviendo. Sabemos que las nuevas tecnologías, las denominadas TICs³⁶, y todos los demás recursos en red (diccionarios *online*, materiales didácticos, interfaces de consulta en línea de corpus de nativos, tecnologías lingüísticas, etc.)³⁷ son muy rentables en la didáctica del léxico (Cruz Piñol, 2015).

Estas herramientas abarcan una multitud de diccionarios electrónicos, que conllevan una gran cantidad de propuestas prácticas, cuya principal finalidad es facilitar

³³ Cabe señalar que en 2011 se ha iniciado un proyecto (*in fieri*) finalizado a la redacción de un diccionario monolingüe para el aprendizaje de ELE en la Universidad Pompeu Fabra de Barcelona, patrocinada por la Fundación Comillas y el Ministerio de Ciencia e Innovación de España, el *Diccionario de Aprendizaje del Español como Lengua Extranjera*.

³⁴ Véase Lewis (2000); Higuera (2006; 2008); Ruiz Martínez (2006) y Minervini (2008) para observar una UD concreta basada en el aprovechamiento de estos recursos.

³⁵ Del mismo modo, el Grupo DiCE, bajo la dirección de Alonso Ramos, está trabajando a la realización de un diccionario especializado en las colocaciones léxicas del español (y en su explotación didáctica), el *Diccionario de colocaciones del español*, cuya nomenclatura se limita, por el momento, al campo semántico de los nombres de sentimiento.

Las características principales de obras lexicográficas de esta naturaleza se pueden encontrar en Ferrando (2013 *apud* Pérez Serrano, 2017: 73).

³⁶ Entre otros, Contreras (2008) plantea una reflexión sobre la integración de las TICs en el proceso de enseñanza-aprendizaje de ELE, analizando la calidad y la idoneidad de los recursos informáticos a las necesidades cotidianas del alumnado ya que las nuevas tecnologías, es sabido, han experimentado un gran desarrollo en las últimas décadas hasta ser consideradas un elemento de transformación en la educación (Hernández Muñoz y Román-Mendoza, 2018).

³⁷ Si bien en esta sección nos limitamos a mencionar el diccionario, somos conscientes de que la lexicografía del siglo XXI ofrece otros recursos para el aula, atendiendo especialmente a su relación con la didáctica de LE y la sociedad digital que ha de adaptarse a una compleja y cambiante realidad y, para ello, debe valerse de los instrumentos propios de su tiempo (de los que prescindimos en este trabajo pese a ser igualmente importantes en el aprendizaje de una lengua).

informaciones útiles y actualizadas sobre el vocabulario, aplicables a través de un vasto abanico de posibilidades.

A estas alturas, en lo referente al desarrollo de las destrezas referenciales de los estudiantes de LE que permiten el correcto manejo del diccionario, en un marco tan amplio como es la lexicografía, insistimos en la relación entre la obra lexicográfica, su didáctica y la sociedad digital. Es necesario atender a este nuevo tipo de destinatario, la denominada “Generación Z”, y las constantes evoluciones tecnológicas. En efecto:

La enseñanza de segundas lenguas no puede ser ajena a esta nueva manera de construir nuestra naturaleza social. Integrar las tecnologías en el aprendizaje es reconocer y asumir cómo somos y empoderar al alumno para que desarrolle plenamente sus capacidades de agente social en la nueva lengua. (Fernández Juncal y Hernández Muñoz, 2018: 2)

Es más, uno de los principales desafíos en el campo de la didáctica de los diccionarios de la actualidad es la formación adecuada y efectiva de las destrezas de los usuarios:

La lexicografía es una disciplina dinámica y la era digital se caracteriza por la existencia de cambios e innovaciones continuos. Por ello la relación entre lexicografía y sociedad no es estática, aunque no cabe duda de que algunas características tendrán una naturaleza duradera. Por lo que respecta tanto a los aspectos estáticos como dinámicos de la sociedad y a la lexicografía, debemos asegurarnos de realizar los esfuerzos necesarios para lograr que esta relación sea la mejor posible. (Gouws, 2017: 18)

5. LA SELECCIÓN LÉXICA

Como hemos visto, el léxico ha pasado de ser un componente marginal, presentado bajo listas de palabras descontextualizadas, a cobrar protagonismo como componente nuclear en los procesos de aprendizaje de las LE, trabajado en el aula mediante una multitud de actividades y técnicas. Esto requiere una enseñanza formal y bien organizada, que supone un cuidadoso proceso de selección de los contenidos.

En particular, desde el Enfoque Comunicativo, el léxico ha adquirido una gran importancia, revelándose fundamental para el desarrollo de la competencia comunicativa al ser vehículo de significados y, desde los planteamientos del Enfoque Léxico, se ha puesto en el centro de la ASL para el alcance de la competencia léxica.

De ahí que se haya comprendido la necesidad de programar y planificar una enseñanza cualitativa del vocabulario, la cual ha abierto un nuevo interrogante: *¿qué léxico enseñar?* Contestar a esta pregunta no es cuestión baladí ya que cada lengua cuenta con una gran cantidad de unidades léxicas, que varían en función del entorno (variación

diatópica, diafásica, diastrática); del asunto (variación temática); de la evolución lingüística (variación diacrónica).

Asimismo, el léxico engloba el conjunto de las palabras de un idioma y es imposible que una persona, incluso un hablante nativo, alcance su dominio completo. Así pues, a la hora de programar la enseñanza se debe reflexionar sobre los elementos que necesitan conocer los estudiantes de manera que se trabaje con una cantidad y una tipología de ítems adecuada a la situación comunicativa o la tarea que se desea practicar.

El objetivo de la selección es contextualizar el léxico para que el *input* inicial se convierta primero en *léxico pasivo* (aquellas unidades que pueden ser reconocidas al ser escuchadas o leídas) y más tarde en *léxico activo* (las unidades que el aprendiz puede recuperar en su memoria y usar correctamente en un educto). Se respalda la importancia de un buen proceso de selección léxica unido a las siguientes fases de planificación, secuenciación y presentación del vocabulario, así como al diseño de los materiales finalizados a la ASL, porque, de lo contrario, la primera por sí sola no serviría de mucho.

5.1 Criterios básicos

Hasta la primera mitad del siglo XX, se asumía que para poder enseñar el léxico de una LE era esencial conocer las voces más frecuentes de esa lengua, las más *rentables* para fines comunicativos³⁸.

En efecto, una unidad léxica con un alto grado de frecuencia tiene una gran *utilidad* y a nivel pedagógico resulta de mayor interés que otra, menos utilizada (Ávila Martín, 2010: 169); por ende, es mucho más rentable aprender una unidad altamente usada y polisémica que otra que se emplea menos (Almela *et al.*, 2005).

A este propósito, existe una estrecha correlación entre frecuencia y polisemia: si una palabra tiene más sentidos, presenta también más ocurrencias, por lo que su aprendizaje resulta más conveniente.

Las palabras frecuentes son las que cubren un mayor número de campos semánticos, aparecen en diferentes registros y deberían incluirse en los contenidos didácticos. Sin embargo, sabemos que el criterio de frecuencia por sí mismo no es ni

³⁸ La *rentabilidad* comunicativa se mide a través de la *cobertura*: parece que las cinco mil palabras más frecuentes de una lengua cubren el 90% de un texto, lo que permite desenvolverse en todas las situaciones comunicativas cotidianas (Haensch, 1982: 159; Alvar Ezquerro, 2003: 100).

suficiente ni proficuo para el aprendiz extranjero ya que los ítems están descontextualizados y, en la mayoría de los casos, se trata de palabras funcionales, que suponen la ausencia de unidades de contenido nocional (Ávila Muñoz, 2017: 73).

En segunda instancia, la *utilidad* es un patrón de selección alternativo (Bartol Hernández, 2010: 88), pero, al mismo tiempo, es difícil establecer qué significa que una palabra es útil y cómo cuantificar dicha utilidad, lo que ha llevado a plantear otras propuestas.

Por eso, contamos con algunos criterios básicos que juntos permiten lograr una selección léxica que proporcione un *input* adecuado.

(i) La frecuencia.

Se basa en las unidades léxicas más utilizadas en una lengua y en el mayor número de situaciones comunicativas. Sin embargo, la elección no debe basarse exclusivamente en este criterio ya que “[...] la ecuación *frecuente = útil*, se ha demostrado imprecisa” (Hidalgo, 2017b: 165) porque las listas de frecuencia dependen en gran medida de las fuentes elegidas (de la mayor o menor presencia de textos orales o escritos), del registro y del tema; deben ser constantemente actualizadas para no correr el riesgo de elegir formas obsoletas; no ofrecen ninguna información complementaria sobre las mencionadas unidades léxicas.

(ii) El rango.

Sirve como criterio de corrección de la frecuencia, combinándola con el parámetro de dispersión ya que permite detectar las unidades que se encuentran en el mayor número de textos, abarcando tanto la cantidad como la diversidad textual. Por ende, cuanto mayor sea el rango, mayor utilidad se le otorga a una palabra que consigue mayor estabilidad.

(iii) La productividad.

Existen unidades léxicas más productivas que otras, es decir que son capaces de generar una más amplia familia léxica o que tienen una enorme posibilidad de combinación con otras palabras (unidades fraseológicas, colocaciones, secuencias formulaicas, etc.). Obviamente conocer este tipo de unidad resulta beneficioso para el discente que puede emplear y adaptar tales piezas en diferentes discursos y situaciones.

(iv) La rentabilidad.

Se trata de palabras que pueden aparecer en distintos contextos debido a su

significado genérico por lo que están estrechamente relacionadas con el desarrollo de la competencia estratégica del aprendiz: permiten compensar dificultades y deficiencias comunicativas cuando no se dispone de los términos exactos que se requerirían (“capacidad de negociación” según Izquierdo Gil, 2005).

5.2 Criterios ligados a la situación de enseñanza-aprendizaje

Otros parámetros en los que la selección léxica debería sentar sus bases están ligados a la situación de enseñanza-aprendizaje, esto es, dependen del entorno educativo, de la acción docente y de la proximidad lingüística entre L1 y LE.

(i) Las necesidades lingüísticas y los intereses del alumnado.

A la luz del hecho de que la enseñanza está centrada en el alumno, el discente es el sujeto de su propio aprendizaje, por tanto, debería tener la posibilidad de manifestar sus expectativas y exigencias. Es necesario que el profesor tenga en cuenta las características y las aficiones de su grupo clase para propiciar contenidos ajustados a las temáticas más interesantes y motivadoras (Nation, 1990).

(ii) La regularidad y la facilidad de aprendizaje.

En primera instancia, se trata de contabilizar el tiempo del que se dispone para el curso y la realización de las actividades, así como de plantear de manera adecuada al grupo la metodología didáctica para presentar unidades léxicas que faciliten el aprendizaje e incrementen el interés.

En segunda instancia, no se debería menospreciar el criterio de facilidad de aprendizaje: empezar el estudio con palabras familiares al aprendiz, por morfología o contexto, que le ayuden en el proceso adquisitivo. A este propósito, Izquierdo Gil (2005) pone de relieve el concepto de *parentesco lingüístico*, sobre todo en los estadios iniciales de la ASL.

(iii) El componente cultural.

El vocabulario conlleva los valores y las creencias de la comunidad de la LO (López Pérez, 2013), por tanto, hay que tener en cuenta que se trata de aportar un tipo de conocimiento sociocultural al alumnado, además del lingüístico.

Se aplicaría un *criterio social*, es decir, trabajar con unidades léxicas típicas de un grupo o de comunidad de habla (Pastora Herrero, 1990: 90).

(iv) El tipo de curso.

La selección léxica debe estar sometida al curso que se ofrece: no es lo mismo programar un curso intensivo o un curso extensivo y, además, la programación cambia ulteriormente si dicho curso se da en el país de origen del estudiantado o en contexto de inmersión. De la misma manera, la selección es distinta si el programa está orientado a la enseñanza de la lengua general o a la lengua con fines específicos (lengua de especialidad).

5.3 Criterios del MCER y del PCIC

Los documentos oficiales europeos para la enseñanza-aprendizaje de las LE también consideran algunos de estos criterios durante la selección de las unidades léxicas que incluyen en sus directrices y, es más, sugieren la aplicación de otros patrones, entre los que el dominio del idioma extranjero del aprendiz es el primero.

Se recomienda graduar la selección y la presentación de las piezas léxicas conforme al nivel lingüístico, empezando por un inventario de palabras más cercanas a la realidad cotidiana del individuo y, solo en niveles avanzados, añadir términos menos inmediatos que permitan ampliar el bagaje léxico y mejorar la precisión comunicativa del discente puesto que ya debería dominar un “registro neutro”.

Asimismo, el MCER (2002, apartado 6.4.7.3.) defiende el uso de textos auténticos³⁹ y aboga por la aplicación de normas léxico-estadísticas en la elección de palabras y unidades multipalabra que deben respetar, de un lado, el criterio de frecuencia y, del otro, adaptarse a determinadas áreas temáticas que cubran los intereses y las necesidades de los aprendices.

5.4 La disponibilidad léxica como criterio de selección

En suma, las pautas que dirigen los procesos de selección léxica son muy variadas y la crítica que se les puede hacer es que, en algunos casos, resulta demasiado subjetiva porque, pese a la presencia de muchos criterios objetivos basados en recuentos estadísticos, algunas unidades léxicas pueden ser afectadas por las peculiaridades de cada

³⁹ Aunque Izquierdo Gil (2005) subraya que el uso de documentos auténticos ha disminuido la importancia de los procesos de selección léxica y ha influido, negativamente, en la calidad del *input*.

entorno educativo o comunidad lingüística. Para ello, como ya hemos mencionado, con el objetivo de proponer una solución que supere algunos de los inconvenientes reseñados, se ha planteado otro criterio de selección: la *disponibilidad léxica*.

Aunque abordaremos detenidamente este concepto en el siguiente capítulo, nos parece oportuno introducir brevemente ahora algunas cuestiones ya que se trata de un método empírico de medición del léxico, útil también como criterio de selección para elegir el vocabulario de una manera objetiva y capaz de resolver los problemas que acabamos de señalar:

Sin duda alguna, una de las cuestiones más complejas, una vez decidida la variedad de referencia, es seleccionar el vocabulario que deben aprender nuestros alumnos, debido a que mientras los rasgos fonéticos y morfosintácticos de las distintas variedades forman un catálogo más o menos cerrado, finito, y por lo general bien analizado por disciplinas como la Dialectología y la Sociolingüística, el nivel léxico representa un verdadero quebradero de cabeza por su amplitud y su constante evolución. Dicha selección debe estar basada en ciertos criterios como son la frecuencia (léxico básico) y la disponibilidad (léxico disponible). (Contreras, 2014: 223)

En primer lugar, el léxico disponible es el conjunto de lexemas que se utilizan en una situación comunicativa dada (López Morales, 1990: 62) y presenta una gran carga nocional, frente a las unidades funcionales extraídas mediante el criterio de frecuencia.

En segundo lugar, el carácter semántico de los ítems léxicos “disponibles” implica que la importancia de incluir en las propuestas didácticas aquellas palabras que aparecen en estos listados desde el momento en el que para comunicar se necesitan un tema y un contexto⁴⁰.

En tercer lugar, es posible reducir la subjetividad de la selección gracias al análisis de datos reales: los resultados de las encuestas de disponibilidad léxica proporcionan informaciones objetivas y “una base estadística sólida para efectuar la selección y graduación de las unidades léxicas que debemos enseñar. [...] pueden ayudarnos a reducir

⁴⁰ A estas alturas, destaca la correspondencia entre las áreas temáticas abordadas en muchos de los estudios de disponibilidad léxica y los ámbitos aconsejados tanto por el MCER, como por el PCIC. La propuesta de Gómez Molina (2004: 795) resume las indicaciones generales del Instituto Cervantes en lo que se refiere a los contenidos léxicos adecuados a los objetivos de cada nivel: 1. identificación personal; 2. casa y alojamiento; 3. trabajo, estudios y ocupación; 4. tiempo libre; 5. viajes y transporte; 6. relaciones sociales; 7. salud y estado físico; 8. compras; 9. comidas y bebidas; 10. edificios de servicios públicos. Si se observa bien, coinciden en gran medida con la detallada propuesta, más temprana, dividida en quince temas y nociones de Sánchez Lobato y Aguirre (1998): 1. identificaciones personales; 2. circunstancias personales; 3. profesión, trabajo, creencias; 4. paisaje y fauna; 5. condiciones atmosféricas; 6. entretenimientos y aficiones; 7. actividades artísticas, prensa; 8. aficiones: deportes y toros; 9. viajes, alojamiento y transporte; 10. salud, enfermedad y figura humana; 11. educación y enseñanza; 12. alimentación, comida y bebida; 13. comercio, industria y vestido; 14. moneda, correo, sociedad y estado; 15. pesos, medidas y números.

la eventual arbitrariedad del vocabulario introducido en los manuales o en el aula” (Izquierdo Gil, 2005: 85).

Por último, se potencia el aprendizaje gracias a la agrupación de las unidades léxicas en campos nocionales que, como hemos visto, permiten la creación de redes y asociaciones que ayudan su almacenamiento en el lexicón mental.

6. LA COMPETENCIA LÉXICA

En esta última sección del capítulo, revisamos la noción de *competencia léxica*, aunque de manera muy resumida dada la complejidad de la cuestión y las posibilidades de explicarla más pormenorizadamente, asumiendo que se trata del “conocimiento del vocabulario de una lengua y la capacidad para utilizarlo” (MCER, 2002: 108).

Aceptamos la propuesta de Rufat (2018: 74-75), que parte de las contribuciones de la lingüística, la psicolingüística y la pedagogía para explicar las “tres dimensiones del vocabulario” que, a nuestro entender, se repercuten en la competencia léxica.

Primeramente, la dimensión lingüística se refiere tanto a las características de los ítems léxicos (la forma y el significado) como al desarrollo de aquellas destrezas que permiten la comunicación en la LO (las estrategias de aprendizaje, de retención y de uso).

En efecto, el Consejo de Europa (2001) establece que la competencia léxica permite el desarrollo de una habilidad comunicativa eficiente: no poder, o no saber, decodificar una palabra puede imposibilitar el éxito de la comunicación. Por eso, en esta óptica, la competencia léxica se incluye dentro de la más amplia competencia lingüística (junto a la gramatical, la semántica, la fonológica, la ortográfica y la ortoépica), la cual, por su parte, es una subcompetencia de la competencia comunicativa.

De la misma manera, para el *Diccionario de términos clave de ELE*, determina la capacidad de uso de las palabras y coincide con la interiorización del vocabulario, algo parecido al *lexicón mental*.

Šifrar Kalan (2009: 165) resume estas posturas al afirmar: “La competencia léxica es parte fundamental del conocimiento lingüístico y una herramienta cognitiva que nos permite comunicar en una lengua.”

Todo esto, nos lleva a considerar imprescindible el conocimiento del vocabulario para poder comunicarse y, en consecuencia, se hace necesario para el docente planificar el desarrollo de la competencia léxica de su alumnado durante la ASL.

En lo que atañe a sus componentes, el MCER (2002: 108-109) afirma que está formada por elementos léxicos y gramaticales, a la vez, que deben tenerse en cuenta en los procesos de enseñanza, en las propuestas y en las actividades didácticas.

Hay que presentar todas aquellas unidades que son parte integrante de la competencia léxica y permiten su alcance. En específico se trata de:

(i) elementos léxicos:

- expresiones hechas (fórmulas fijas, modismos, metáforas, estructuras fijas, otras frases hechas, régimen semántico, etc.);
- polisemia.

(ii) elementos gramaticales:

- artículos;
- cuantificadores;
- demostrativos;
- pronombres personales;
- pronombres relativos;
- adverbios interrogativos;
- posesivos;
- preposiciones;
- verbos auxiliares;
- conjunciones.

Asimismo, desde el Enfoque Léxico, por encima de las palabras entendidas como el material gráfico comprendido entre dos espacios en blanco, se considera útil y provechoso para el aprendiente no nativo conocer la combinatoria léxica de la lengua (Lewis, 1993 *apud* Hidalgo, 2017b: 47), a saber:

- unidades multipalabras;
- colocaciones;
- expresiones fijas;
- expresiones semifijas.

De acuerdo con todo ello, “la competencia léxica se distingue por un equilibrio entre el conocimiento generativo (basado en reglas) y el formulaico” (Rufat, 2018: 86) que hacen el léxico recordable.

La siguiente figura resume esquemáticamente esta clasificación (Pérez Serrano, 2017: 16):

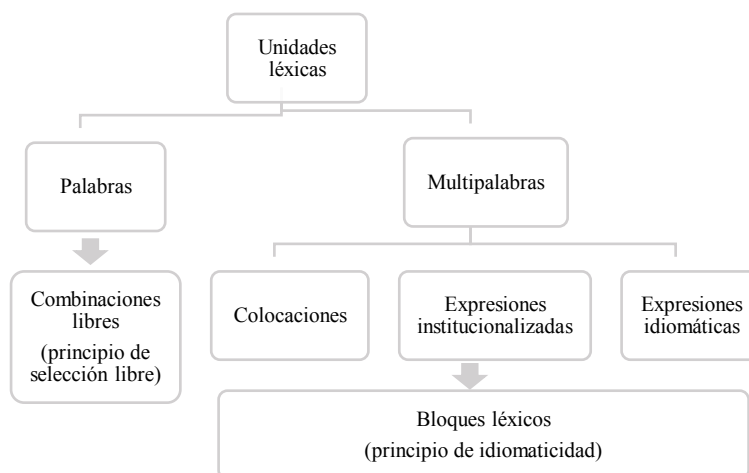


Figura 2. Clasificación de las unidades léxicas.

Claro está que durante el aprendizaje estos ítems se reparten en una escala graduada en función del nivel lingüístico del grupo de discentes, es decir, según una gradación del conocimiento del léxico y de la capacidad de controlar ese conocimiento, definidas *dominio* y *riqueza* del vocabulario. Jiménez Catalán (2002: 152) plantea una doble definición de competencia léxica: por un lado, es “el conocimiento que se debe poseer para poder utilizar la palabra con propiedad”, y del otro, es “la capacidad de reconocer, aprender, recuperar y relacionar las distintas palabras a nivel oral y escrito”. Esto implica que además del conocimiento de la forma oral y escrita de la unidad léxica, es necesario reconocer su significado y saber diferenciarlo de otras unidades:

Es ya un lugar común destacar la importancia de un vocabulario que nos ayude a entender y tener confianza en nosotros mismos, que sea variado, preciso y se ajuste a nuestras necesidades. Así pues, el hablante que tiene un buen vocabulario, rico y amplio sabe muchas palabras, lo que significan y cómo se escriben. Sin embargo, [...] saber una palabra no termina ahí sino que implica conocer un complejo número de aspectos. (López-Mezquita Molina, 2005)

En la misma línea, Marconi (1999) ha contestado a la pregunta *¿en qué consiste la capacidad de utilizar las palabras?* aportando una teoría semántica sobre la competencia léxica, cuyos patrones suponen que “ser capaz de utilizar una palabra” implica otras dos subcompetencias que interactúan⁴¹:

⁴¹ Aunque sean independientes una de la otra (véase la distinción entre *léxico fonológico*, *léxico gráfico* y *léxico semántico* en Marconi, 1999: 81 y ss.).

- (i) la *competencia referencial* hace referencia al desarrollo de acciones perceptivas y motoras: la denominación de los objetos que nos rodean (asignarles una palabra a los objetos) y la aplicación de las palabras en el mundo real (escoger el objeto relacionado con una palabra determinada);
- (ii) la *competencia inferencial* coincide con el conocimiento de la red de conexiones entre palabras y otras expresiones lingüísticas del mismo sistema, es decir la capacidad de definición, inferencia semántica y paráfrasis.

En el frente de la dimensión psicolingüística del léxico se abarca el estudio de las capacidades cognitivas del ser humano en lo que se refiere a los mecanismos de procesamiento del lenguaje, esto es, cómo se aprende el vocabulario y cómo se almacena y se recupera de la mente. Por eso, suele acompañarse a la competencia léxica el concepto de lexicón mental, que coincide con

[...] la parte de la competencia lingüística que contiene las piezas léxicas o formantes, es decir, las raíces y temas, los afijos flexivos y derivativos, y las reglas que regulan su combinación. El concepto es cognitivo, dinámico, procesual, a diferencia del concepto de “léxico”, como sinónimo de “vocabulario”, entendido como un simple listado de palabras, o como una organización de campos semánticos. El lexicón es un constructo que da cuenta de la capacidad creativa del lenguaje, y por lo tanto de las lenguas. Permite comprender y explicar los fenómenos de generación de nuevas palabras, que pueden ser entendidas, procesadas y re-creadas de manera novedosa, sin información explícita, por cualquier hablante nativo e, inclusive, por un hablante no nativo de una lengua. En este sentido, el adjetivo “mental” pue-de ser redundante, pero mantenemos la calificación con el propósito de recalcar su carácter cognitivo. (Baralo, 2001: 12)

Según esto, la adquisición del vocabulario de una LE, que lleva al desarrollo de la competencia léxica, no es un conocimiento estático, sino que es un proceso dinámico en el que las palabras se aprenden, se memorizan y se guardan de manera ordenada en un almacén que está en continua evolución (Richards, 1976). En un estadio inicial, *saber* una palabra significa conocer su forma, pero, sucesivamente, *saber realmente* una palabra abarca otros aspectos: la primera etapa corresponde con un conocimiento superficial de la pieza léxica y la final coincide con la capacidad de usarla correcta y eficazmente en la producción de un acto de habla.

Sin entrar en detalles, la dimensión pedagógica del léxico, además de tratar la gradación de los contenidos, pone de relieve la importancia de la planificación de la didáctica del vocabulario y, según los principios del Enfoque Léxico, plantea su instrucción explícita, pues ha tomado protagonismo como foco de la ASL finalizado a la ejecución de las tareas, las actividades y los procesos lingüísticos que los alumnos llevan

a cabo como agentes sociales con el objetivo de desenvolverse en todo tipo de interacción (MCER, 2002: 128).

En resumidas cuentas, se supone que la intención comunicativa del aprendizaje lingüístico requiere el desarrollo de la competencia léxica, la cual permite la producción del lenguaje, ordenada en tres fases en función de una secuencia de causa-efecto: (i) conceptualización; (ii) formulación; (iii) articulación, que se refleja en el postulado de Nation (2001) sobre el “conocer una palabra” que aborda la forma, el significado y el uso de tal palabra.

Es incuestionable que el desarrollo de la competencia léxica tiene un rol central en los procesos de ASL y debe entenderse de forma transversal: tiene aspectos ligados a la competencia lingüística, a la sociolingüística y a la pragmática y, es más, es un fenómeno extremadamente complejo relacionado con saberes declarativos, procedimentales y estratégicos.

CAPÍTULO II.

LA DISPONIBILIDAD Y LA RIQUEZA LÉXICA: ORÍGENES Y DESARROLLO ACTUAL

1. PREMISA TERMINOLÓGICA

Para procesar adecuadamente la información epistemológica de cualquier disciplina es requisito indispensable precisar los términos utilizados y evitar ambigüedades. Por ello, antes de pasar al estado de la cuestión, que ocupa este capítulo, desglosamos las nociones en las que se basan los análisis de la parte empírica. Definimos una serie de términos clave propios del marco teórico en el que se sitúa la presente Tesis.

1.1 Palabras y vocablos

En el lenguaje común los términos *palabra* y *vocablo* suelen utilizarse indistintamente para referirse a los lexemas de una lengua, o bien a los que un hablante utiliza en sus actos de habla (Casadei, 2003), pero, aunque puedan parecer sinónimos, en léxico-estadística designan dos conceptos bien distintos.

Müller (1968) puntualiza que el término *palabra* tiene dos acepciones: abarca, a la vez, la unidad del texto y la unidad del léxico. En el primer caso, se refiere a los *tokens*, es decir a todas y cada una de las unidades léxicas, repetidas o no, que aparecen en un corpus (o en un texto); en el segundo caso, se alude a los *types*, esto es, a cada una de las piezas léxicas distintas que se registran en una muestra, sin repeticiones. Asumimos que la *palabra* es la unidad básica del léxico, en su doble acepción, que se caracteriza por resultar de fácil identificación al ser el material gráfico comprendido entre dos espacios en blanco.

Si consideramos los procesos empíricos de medición de la disponibilidad y de la riqueza léxica, las *palabras* son todas las unidades aportadas por los informantes en un test, repeticiones incluidas; mientras que los *vocablos* son todos los ítems diferentes arrojados, sin tener en cuenta las repeticiones.

Sin embargo, en esta línea de investigación, no todos los autores emplean los mismos términos: Ávila Muñoz (2006) distingue *tokens* y *lemas*; Ahumada (2006) habla de *ocurrencias* y *registros*; Martínez Olmos (2007) se refiere a *ocurrencias* y *tipos léxicos*

diferentes; López González (2014b) utiliza *palabras totales* y *palabras diferentes*: “[...] Esta variedad, en definitiva, no es más que la demostración del interés en dicha distinción, lo que deja patente la importancia de diferenciar dichas nociones para una correcta interpretación de los resultados en disponibilidad léxica” (Hidalgo, 2017b: 70).

A lo largo de este trabajo, salvo que se indique lo contrario, utilizaremos indistintamente las dicotomías *palabra/vocablo* y *token/type*.

1.2 Tipos de léxico

Igualmente, en este ámbito, el concepto de *léxico* tiene diferentes acepciones. Veamos los significados relacionados con los principios que rigen los métodos de disponibilidad y riqueza léxica que nos ocupan aquí.

(i) El léxico frecuente.

El *léxico frecuente* es el conjunto formado por aquellas palabras que presentan un mayor número de ocurrencias en el recuento de una muestra lingüística. Si se extrae la lista de las unidades más repetidas de un corpus, dichas unidades pertenecen al léxico frecuente. Por esta razón el criterio que se percibía como más provechoso para una selección pedagógica adecuada del léxico han sido, hasta la primera mitad del siglo XX, los índices de frecuencia, lo cuales permitían establecer cuáles eran los vocablos más estables de un idioma y que, sin duda, aparecían en cualquier contexto comunicativo independientemente del tema, por lo que se debían enseñar. Sin embargo, se ha reconsiderado pronto la exclusiva aplicación de este factor ya que resultaba que muchas de estas unidades frecuentes eran palabras funcionales del discurso, de poco, o nulo, contenido semántico.

(ii) El léxico básico.

El *léxico básico*, o *usual*, abarca las lexías de una lengua que se emplean habitualmente en las comunicaciones cotidianas. Se relaciona con el léxico frecuente tanto que a veces se confunde: parece que hay una relación de inclusión o, mejor, de evolución del léxico frecuente hacia el léxico usual. Sin embargo, se trata de dos conceptos distintos ya que el primero es un léxico derivado de un proceso de medición independiente del contexto, mientras que el segundo es un conjunto de vocablos más específico y vinculado con el uso real de las palabras.

Los listados de léxico básico recogen las unidades léxicas más estables de la lengua donde el criterio de selección pasa de la frecuencia al uso en todo tipo de situación comunicativa y de texto.

Frente a la consideración de que un nativo, y aún menos un aprendiz extranjero, no puede conocer todo el léxico de una lengua, el aprendizaje de las unidades básicas lleva a un primer desarrollo de la competencia léxica y, por supuesto, un primer nivel de conocimiento lingüístico.

(iii) El léxico disponible.

El *léxico disponible* se refiere a las palabras que se activan en una situación comunicativa concreta y abarca, en la mayoría de los casos, unidades léxicas de contenido nocional. Este concepto ha sido elaborado por Michéa (1953) a partir de la distinción entre *palabras temáticas* (lexías concretas relacionadas con el tema de la comunicación, que pueden ser más o menos frecuentes, es decir palabras inestables)⁴² y *palabras atemáticas* (lexemas gramaticales que aparecen en cualquier texto independientemente del tema comunicativo, dotadas de un mayor índice de frecuencia, es decir palabras estables). Las palabras temáticamente disponibles para cualquier hablante (por ejemplo, *tenedor*) pueden ser muy infrecuentes en un corpus de textos escritos, incluso de grandes proporciones, destacando así una de las limitaciones del criterio de frecuencia, que no resulta proficuo para la apropiada selección del vocabulario destinada al aprendizaje. Efectivamente, la literatura crítica subraya muy pronto que el léxico frecuente presenta palabras descontextualizadas, procedentes exclusivamente de textos escritos que dejan de lado el vocabulario más común propio de la comunicación hablada espontánea y que, además, se caracteriza por la demasiada abundancia de palabras funcionales (Gougenheim *et al.*, 1964: 31-35).

Se infiere que una palabra definida “disponible” es un ítem que, sin ser necesariamente frecuente, está siempre listo para ser usado y viene a la mente de forma inmediata, y natural, según el tema de la comunicación. Las palabras disponibles son las que acuden más fácil y rápidamente a la mente de un hablante en una situación comunicativa dada y se caracterizan por ser vocablos de alto contenido nocional-informativo (Borrego Nieto, 2004): se trata del vocabulario “que se usa en una situación

⁴² Subraya López Chávez (1994: 69): “Son inestables debido a que es necesario que se toque un tema específico en el discurso para que aparezcan.”

comunicativa específica, actualizándose sólo ante la necesidad de comunicar algo referente a esa situación concreta” (López Morales, 1999: 9).

(iv) El léxico fundamental.

Los avances de las disciplinas léxico-estadísticas han permitido plantear nuevas categorías de léxico, más abarcadoras, como es el caso del *léxico fundamental*, producto de la suma del básico con el disponible.

Se asume que el léxico frecuente, por sí solo, no es el único índice válido para trabajar sobre el vocabulario de una lengua y tampoco lo son el léxico básico o el disponible, así pues, es necesario implementarlos con ítems contextualizados. De ahí que se combinen el vocabulario real de uso (estable) con un vocabulario más concreto, influenciado por el contexto comunicativo (y por eso menos estable), a partir de los cuales se recopilan listados de palabras y unidades más complejas que unen los índices de uso y de disponibilidad.

En lo que atañe al ámbito de la enseñanza-aprendizaje de LE, el propósito es ofrecer un *input* que permita una selección adecuada del vocabulario en la confección de materiales didácticos, orientándolo hacia el desarrollo de un conocimiento “natural” del léxico, próximo al de un nativo (en nuestro caso, el español “común”).

2. LA DISPONIBILIDAD LÉXICA

2.1 Introducción

Los *diccionarios de frecuencia* (Käding, 1897; Thorndike, 1921; Henmon, 1924; Buchanan, 1927; Bakonyi, 1933; Van der Beke, 1935; West, 1935; Rodríguez Bou, 1952; García Hoz, 1953; Juilland y Chang-Rodríguez, 1964), concebidos desde finales del siglo XIX y durante la primera mitad del XX como herramientas lexicográficas construidas con procedimientos estadísticos, han tenido una rápida aplicación en el ámbito del aprendizaje de las LE. Consistían en listas de palabras estadísticamente frecuentes en un corpus textual representativo, por lo que su finalidad didáctico-pedagógica primaria permitía la selección del léxico en la confección de materiales didácticos, orientándola hacia el desarrollo de un conocimiento “natural” de la LO, próximo al de un nativo. A partir, precisamente, de los mencionados diccionarios se elaboraban listas de palabras

frecuentes para la enseñanza del vocabulario ya que entonces no se reconocía que el conocimiento del léxico constituía un factor de éxito del proceso de aprendizaje lingüístico por lo que no era necesario elaborar otras estrategias, más eficaces.

La literatura ha destacado muy pronto (Gougenheim *et al.*, 1964: 31-35) que el mero criterio de frecuencia no era proficuo para el aprendiz extranjero. En efecto, los juicios negativos de la didáctica y la lingüística aplicada del siguiente siglo han llevado a criticar estos recursos porque sustentaban el recuento del léxico de una comunidad solamente en función de la mayor o menor aparición de las palabras, método que, en realidad, no se ha revelado eficaz, pues se ha detectado que lexías que presentaban índices de frecuencia bajos o muy bajos (que, por tanto, no habían sido incluidas en dichos diccionarios) eran, por el contrario, comúnmente usadas por los nativos. Basarse solo en este criterio no resultaba tan provechoso como para ofrecer al discente el caudal y el tipo de lexías necesarias para el adecuado aprendizaje de la LE. Asimismo, los diccionarios de frecuencia han sido tachados, en primer lugar, de estar descontextualizados ya que presentaban meras listas de palabras y, en segundo lugar, de ser limitados por considerar como fuente solo textos escritos (cultos o especializados) descartando así el vocabulario más común. Esta inadecuación, además, se ha visto afectada también por la abundancia de unidades funcionales contenidas y la ausencia de las unidades de contenido nocional, debida al bajo índice de frecuencia que estas presentaban, pese a ser muy conocidas y empleadas.

Así pues, desde los años cincuenta, se ha realizado una revisión metodológica, a partir de la cual han iniciado los análisis de la *disponibilidad léxica* gracias a esta preocupación surgida por una selección léxica adecuada al establecimiento del caudal y del tipo de vocabulario que debía aprenderse de una LE.

2.2 El estudio pionero

El estudio pionero sobre disponibilidad léxica se ha llevado a cabo en países francófonos por parte de un grupo de lingüistas que se ha planteado este método para solucionar las limitaciones de la selección léxica fundada exclusivamente en el criterio de la frecuencia, incluyendo la contextualización de las palabras.

El concepto de disponibilidad léxica se ha desarrollado a partir de la elaboración del *Français Fundamental* (1964), aunque ya en 1953 se había puesto de relieve el hecho

de que algunas palabras, aun siendo perfectamente conocidas por los usuarios nativos de una lengua, solo se emplean en determinadas situaciones comunicativas, por esta razón, no constituyen el léxico frecuente de una lengua que, en consecuencia, no puede ser la única vía para la enseñanza del vocabulario. Se requería la inclusión de otras palabras, definidas *disponibles*, para obtener el cuadro completo de la situación: los estudios sobre el léxico disponible han permitido seleccionar las unidades léxicas del francés concebidas como básicas y, por tanto, más apropiadas para la enseñanza en los niveles iniciales de aprendizaje.

Sin embargo, como veremos, habrá que esperar hasta finales del siglo para apreciar la mayor difusión de los trabajos de esta índole, que abrirán un amplio abanico de posibilidades de aplicación en distintos campos.

A la luz de estos hallazgos, se sentaron las bases para esta nueva metodología que cumpliera con el objetivo didáctico asignado por el *Ministère de l'Éducation Nationale* francés, esto es, facilitar el aprendizaje del vocabulario del francés como LE. El propósito de superar las limitaciones del criterio de frecuencia observado por el comité de especialistas encargados de su edición llevó a un análisis que abordaba también el estudio de aquellas palabras conocidas por los nativos franceses pero que no aparecían en las listas de frecuencia utilizadas hasta la fecha.

Este proyecto nació a partir del presupuesto de que el vocabulario de una lengua es inagotable para los hablantes nativos, y aún más para los aprendices extranjeros, por lo que un grupo de especialistas elaboró una nueva lista de palabras, llamada inicialmente *Français Élémentaire* (1954), que contenía los elementos esenciales para permitir un rápido acceso al léxico de la LO.

Resultaba, por ende, indispensable llevar a cabo una ulterior selección de los contenidos didácticos previamente propuestos⁴³. Tras estas labores, que completaban los datos de la frecuencia contextualizando las palabras de las listas, se definió, el concepto de disponibilidad léxica:

En una situación determinada, las primeras palabras que vienen a la mente son aquellas que están relacionadas con tal situación y que la caracterizan [...]. Una palabra disponible es una

⁴³ Desde una perspectiva histórica, es sabido, se ha modificado el método de enseñanza de las LE ya que las lenguas modernas no se aprendían de la misma manera en la que se estudiaban el griego y el latín. Asimismo, las necesidades del alumnado habían cambiado: tenían objetivos diferentes, más prácticos y comunicativos.

palabra que, sin ser particularmente frecuente, siempre está lista para ser utilizada, y acude inmediatamente a la mente cuando parece necesario. (Michéa, 1953: 342)⁴⁴

De ahí, el *léxico disponible* se ha identificado con el vocabulario activado en una situación comunicativa concreta (López Morales, 1984: 62). Al sumarlo con el *léxico básico* de la lengua se consigue recopilar una lista que comprende el *léxico fundamental* de dicha lengua.

El *Français Élémentaire* pasó a ser *Français Fondamental* y se publicó como la obra de consulta de referencia diseñada para el aprendizaje del léxico francés LE con el objetivo de propiciar el vocabulario apropiado para los estudios de los aprendices extranjeros.

Todo esto supuso un punto de inflexión en el marco de la léxico-estadística y un proceso de reajuste metodológico que culminó en la planificación de un nuevo sistema de recogida de datos para la elaboración del léxico disponible de una comunidad.

Se organizaron pruebas escritas de tipo asociativo (aun vigentes en este ámbito investigativo) suministradas a grupos de informantes, que recopilaban listas de palabras relacionadas con unos estímulos temáticos, definidos *centros de interés* (Michéa, 1950: 189), “dando lugar a listados de vocabulario que permiten conocer el léxico real con el que cuenta un hablante cuando el tema de conversación gira en torno a uno de estos centros” (Sánchez-Saus, 2016: 22). Por su parte, López Morales (1999: 32) ha puesto de manifiesto que se trata de “[...] las únicas [pruebas] que hacen posible en condiciones experimentales, que se realicen en la actuación lingüística las unidades léxicas [aunque] con poca estabilidad estadística”.

La primera prueba de disponibilidad léxica se ha llevado a cabo con un grupo de escolares franceses, elegidos porque en aquella edad los niños ya tienen un vocabulario bien desarrollado, pero libre de todo tipo de especialización sectorial (Gougenheim *et al.*, 1964). Hernández Muñoz (2015) corrobora esta idea al afirmar que la formación previa al acceso al mundo laboral es uno de los factores ligados a la experiencia que modifican el conocimiento léxico de los hablantes de una lengua.

El procedimiento de recogida de datos suponía la realización del test asociativo

⁴⁴ Nuestra traducción de “*En présence d’une situation donnée, les mots que viennent les premiers à l’esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent [...]. Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé, et se présente immédiatement à l’esprit ou moment où l’on en a besoin.*”

en cuadernillos donde se encontraba un cuestionario sociológico basado en preguntas personales (nombre y apellido; sexo; edad; medio rural o urbano; región de procedencia; nivel sociocultural y nivel escolar; fecha de nacimiento; profesión de los padres) que servían para analizar la posible incidencia de algunas variables sociales en el desarrollo de los conocimientos lingüísticos⁴⁵.

La primera página presentaba otra pregunta, sobre el rendimiento académico del alumno que debía ser rellenada por el profesor encargado de administrar la encuesta (con un juicio de 1 a 3). Las siguientes páginas tenían las instrucciones de la prueba: escribir veinte palabras relacionadas con el centro de interés que les venían a la mente, sin un límite de tiempo⁴⁶. Se precisaba además que las faltas de ortografías no habrían sido evaluadas negativamente; era necesario poner un artículo delante de cada sustantivo y se debía escribir una sola palabra para cada línea. Por último, en las páginas a continuación aparecían las fichas por rellenar.

La edición de los datos para la confección del *Français Fondamental* ha sido la etapa sucesiva que se ha conseguido sumando las palabras más frecuentes a las primeras palabras disponibles procedentes de los análisis de las pruebas realizadas por los niños (aunque nunca se han difundido los listados completos).

Cabe señalar que de entre las palabras más frecuentes se han incluido aquellas que tenían un índice igual o mayor de 29 (aunque, después, han sido aceptadas también las que resultaban útiles para la enseñanza no obstante tuviesen menos apariciones) y se han registrado solo aquellas que estaban en al menos en cinco textos (lo cual ha implicado un decrecimiento del léxico frecuente de 1.400 a 805 ítems). Análogamente, se han reducido las listas eliminando muchas unidades que el comité científico no consideraba rentables (por ejemplo, las lexías coloquiales y las vulgares) y se han quitado

todas aquellas palabras que cuentan con un sinónimo más frecuente –a este respecto señalan los autores que, aunque no existen sinónimos exactos, hay palabras que en el uso son prácticamente intercambiables– [...] y todas aquellas que, en general, no consideraron adecuadas para el nivel inicial (Sánchez-Saus, 2011: 69-70).

Todo esto ha implicado que el resultado fuese marcadamente subjetivo. En efecto,

⁴⁵ Al final, el informe publicado por el *Ministère* solo ha proporcionado los resultados de tres, a saber, el sexo, el medio rural o urbano, el departamento de origen.

⁴⁶ El proyecto preveía un máximo de veinte respuestas para cada área semántica, que no consentía un análisis estadístico de los resultados cuantitativos, paso que se ha modificado en estudios posteriores. Igualmente se ha revisado el tiempo a disposición fijando un límite de dos minutos.

al observar el *Français Fondamental* se ve que se han incorporado ciertas lexías consideradas necesarias, aun a sabiendas de que no aparecían en las respuestas de los informantes, según el principio del empirismo racionalista (Gougenheim *et al.*, 1964: 205) que ha sido largamente menospreciado debido a que se había añadido o eliminado elementos según la opinión del comité. En particular, Juilland (1970) critica tanto la subjetividad en la selección de los campos semánticos, como la ausencia de variantes en las listas de disponibilidad; la no distinción de homónimos y el retoque final de la comisión.

De cualquier manera, el *Français Fondamental* ha estrenado una nueva corriente de estudios sobre el léxico, ha determinado un antes y un después en las metodologías léxico-estadísticas, pese a los problemas y las críticas que le han denunciado y a las propuestas de mejora que ha recibido. Al fin y al cabo, ha introducido un notable cambio en el ámbito de la investigación en lingüística y en la didáctica de LE, siendo un punto de partida, pero no necesariamente un punto de llegada.

2.2.1 *Los centros de interés*

Una de las novedades más destacables del *Français Fondamental* fue el establecimiento de los estímulos temáticos que eran, y son, elementos nucleares de los trabajos en este campo: los denominados *centros de interés* (CI).

Se trata del punto inicial para el diseño de la prueba asociativa que permite recoger el léxico disponible del grupo encuestado. Con el objetivo de contextualizar el vocabulario y facilitar el aprendizaje, los autores se preguntaron cuáles fuesen las áreas semánticas más adecuadas para activar aquellos lexemas que no aparecían en los diccionarios de frecuencia. Propusieron los siguientes dieciséis CI que, en su opinión, reunían el vocabulario universal de una lengua (Gougenheim *et al.*, 1964: 152-153):

01. Las partes del cuerpo;
02. La ropa (hombre y mujer);
03. La casa (sin muebles);
04. Los muebles de la casa;
05. Los alimentos y las bebidas;
06. Los objetos colocados en la mesa para la comida;
07. La cocina, sus muebles y utensilios;

08. La escuela, sus muebles y material escolar;
09. La calefacción y la iluminación;
10. La ciudad;
11. El pueblo o el burgo;
12. Los medios de transporte;
13. Los trabajos del campo y del jardín;
14. Los animales,
15. Los juegos y las distracciones;
16. Las diferentes profesiones.⁴⁷

Estos centros todavía tienen validez, pese a que hayan sido cuestionados por la supuesta universalidad de los temas⁴⁸ y la naturaleza de los proyectos sobre disponibilidad léxica que se realizaron a continuación, puesto que resulta de gran importancia incluir aquellos estímulos que cada investigador cree necesarios (Rubio Lastra, 2018).

Los CI estuvieron y están, sujetos a arbitrariedad: es posible reformularlos, reducirlos o sustituirlos con otros conforme a los objetivos propuestos en cada trabajo (Juilland, 1970; Mackey, 1971a; Benítez Pérez, 1994; López Morales, 1999; Bombarelli, 2005; Sánchez-Saus, 2011; Moreno Fernández, 2012; González Fernández, 2014; Paredes García, 2014; Tomé Cornejo, 2015). Estos cambios se detectan en mayor medida en los estudios dedicados a aprendices extranjeros, ya que sus características difieren de los nativos⁴⁹. Afirma Hidalgo (2017b: 164):

[...] de la inclusión de estos nuevos centros, junto con la alteración de los epígrafes o títulos de los mismos, se desprende que la cuestión de los centros de interés no está cerrada, todo lo contrario, y que existe flexibilidad a la hora de escogerlos, pudiendo excluir algunos e incluir otros, agruparlos, reformularlos o realizar otras modificaciones en función de los objetivos del trabajo. No obstante, debemos ser conscientes de que la inclusión o exclusión de campos nocionales, así como la variación en los títulos, puede afectar a la capacidad asociativa y, por tanto, a su productividad y riqueza.

⁴⁷ La versión original francesa sería: 01. *Les parties du corps*; 02. *Les vêtements (homme et femme)*; 03. *La maison (sans les meubles)*; 04. *Les meubles de la maison*; 05. *Les aliments et boissons des repas*; 06. *Les objets placés sur la table pour les repas*; 07. *La cuisine, ses meubles et ses ustensiles*; 08. *L'école, ses meubles et son matériel scolaire*; 09. *Le chauffage et l'éclairage*; 10. *La ville*; 11. *Le village ou le bourg*; 12. *Les moyens du transport*; 13. *Les travaux des champs et du jardinage*; 14. *Les animaux*; 15. *Les jeux et distractions*; 16. *Les différents métiers*.

⁴⁸ Ya en 1999 López Morales se interrogaba sobre si fuera posible circunscribir todos los intereses humanos en esos dieciséis CI y cuántos de ellos debería tener en cuenta un estudio que pretende ser exhaustivo.

⁴⁹ Por ejemplo, contamos con la introducción de campos semánticos finalizados a activar categorías gramaticales diferentes de los sustantivos; algunos relacionados con temáticas más cercanas a la actualidad; otros encaminados a indagar los conocimientos culturales y la opinión de los estudiantes sobre la comunidad nativa de la LO.

Además de las modificaciones de los campos semánticos, también las pautas metodológicas aplicadas para la realización de las encuestas de disponibilidad léxica han evolucionado tal y como han cambiado las finalidades investigativas de los autores, pero, al mismo tiempo, siempre se ha intentado guardar ciertas semejanzas procedimentales ya que, como veremos, lo importante es respetar unos mínimos criterios para no impedir los cotejos entre diferentes estudios.

2.3 Los sucesores del *Français Fondamental*

El valor de la disponibilidad léxica y la elaboración de los vocabularios básicos ha sido ampliamente reconocido y pronto ha tenido frutos en forma de trabajos herederos. Disponemos de un gran número de investigaciones pioneras, consideradas tales desde el momento en el que el método empírico estaba todavía configurándose.

Destacan, en particular, el trabajo dedicado a la lengua inglesa de Dimitrijevic (1969), quien ha trabajado con escolares de Edimburgo y provincia. El investigador ha introducido algunos cambios metodológicos importantes: (i) ha propuesto el uso de listas abiertas, porque el objetivo de este tipo de encuesta es conocer cuáles y cuántas palabras saben los informantes sin restricción; (ii) ha limitado el tiempo a cinco minutos para cada CI (tras un test previo en el que ha observado el tiempo máximo que empleaban los encuestados para seguir actualizando palabras); (iii) ha reducido los CI a once con el propósito de evitar los campos demasiado generales o vagos al trabajar con niños de 14 años.

En segunda instancia, Mackey (1971a) ha planteado un análisis que examinase el léxico disponible de los hablantes franceses de Canadá para compararlo con los de Francia por voluntad *del Centre International de Recherches sur le Bilinguisme de la Université Laval de Québec*. Si bien ha añadido seis nuevos centros, ha mantenido tanto las pautas del proyecto original, como la tipología de informantes y el cuestionario sociológico, fuesen cuales fuesen sus defectos, de manera que los resultados pudiesen cotejarse⁵⁰. Su conclusión más remarcable afecta la divergencia de las repuestas arrojadas en relación con la cultura de los encuestados, por lo que ha demostrado que el entorno incide en la organización y el conocimiento del léxico.

⁵⁰ Dudaba de la presunta universalidad de los estímulos temáticos, sobre todo, en lo referente a la adecuación a las características y diferencias personales, culturales, sociales de los participantes.

A partir de ahí se ha exportado la disponibilidad léxica en otros países y, en particular, se ha fomentado una línea comparativa entre los varios proyectos. Prueba de ello, Njock (1979) ha alargado el análisis del léxico disponible del francés comparándolo con el léxico disponible del basaa en las poblaciones bilingües de Camerún.

En último lugar, Pfeffer (1964) ha trabajado con la disponibilidad léxica para una selección del vocabulario adecuada a la enseñanza del alemán. Ha establecido la lista del léxico fundamental de esta lengua mediante la suministración de la prueba en escuelas de Alemania, Austria y Suiza (a saber, la zona de habla germana de Europa central). Asimismo, este estudio ha supuesto otro cambio metodológico: se han aplicado nuevas directrices, que requerían que los participantes aportasen un número cerrado de respuestas en veintiuno CI en un límite temporal de diez minutos.

2.4 La disponibilidad léxica en el mundo hispánico

La lengua española es la que más ha contribuido el desarrollo y la productividad de la disponibilidad léxica pese a que los estudios iniciadores estaban centrados en otros idiomas. A partir de los años setenta del siglo pasado, de hecho, se ha empezado a dedicársele un gran número de trabajos en el ámbito del español lengua materna.

El promotor de este avance en el mundo hispánico ha sido Humberto López Morales, que ha seguido fielmente la propuesta del *Français Fondamental* (1964) en lo que se refiere a los dieciséis centros de interés y al análisis de las variables sociolingüísticas. Ha sido el primero en analizar el léxico disponible de las comunidades de habla de Centroamérica y del Caribe, publicando en 1973 su obra pionera *Disponibilidad léxica de los escolares de San Juan* (Puerto Rico).

Prontamente ha planteado un proyecto de mayor extensión, que ha culminado en la elaboración de *Léxico disponible de Puerto Rico* (1999), tras la publicación de una gran cantidad de estudios previos (1978; 1979; 1983; 1984; 1995; 1999), cuya finalidad era determinar el vocabulario en toda la comunidad, por lo que ha suministrado la encuesta a estudiantes universitarios por todo el país.

La investigación puertorriqueña se ha alargado todavía más, gracias al interés que le han dedicado muchos investigadores que han empezado a administrar la misma prueba a muchas de las comunidades de habla española en el mundo, hasta la creación del afamado *Proyecto Panhispánico de disponibilidad léxica* (PPHDL) coordinado por el

propio López Morales. El objetivo último es la recopilación de un diccionario del léxico disponible que encaje la totalidad de los hablantes nativos de español y permita cotejar las diferencias lingüísticas y culturales entre áreas distintas⁵¹.

Para cumplir con este fin se requería la aplicación de un *modus operandi* compartido, sustentado en un mismo conjunto de normas.

A finales de abril de 1999, los investigadores interesados en el PPHDL se han reunido en Bilbao para establecer una metodología común. El eje de este comité era establecer un proceso homogéneo de recogida y tratamiento de los datos que garantizase la comparabilidad⁵².

Se ha decidido, en primera instancia, que la muestra debía componerse de hispanohablantes nativos preuniversitarios, cuyo número ideal era 400 (para una mayor fiabilidad del corpus). La encuesta se dividía en dos partes: la primera consistía en el cuestionario sociológico y la segunda era prueba de disponibilidad léxica, basada en los dieciséis CI del *Français Fundamental*.

Se ha propuesto el uso de listas abiertas con un límite de tiempo de dos minutos para cada estímulo. Este cambio ha consentido adquirir más informaciones porque se ha podido contabilizar todo el caudal de respuestas por informante. Asimismo, la limitación temporal ha permitido conseguir las unidades léxicas más disponibles evitando dejar demasiado tiempo para la reflexión (Samper Hernández, 2002).

En lo que atañe a las variables sociolingüísticas se han establecido:

(i) El sexo.

El género de los encuestados se incluye en todo trabajo sobre disponibilidad léxica si bien, en realidad, no suele presentar cambios relevantes entre las variantes.

(ii) La edad.

Los participantes suelen ser alumnos preuniversitarios, de edad que oscila entre los 17 y 19 años, porque el vocabulario aún no es contaminado por tecnicismos y rasgos típicos del lenguaje especializado propio de sujetos universitarios o profesionales de un sector definido (Hernández Muñoz y Borrego Nieto, 2004).

⁵¹ Entre todas las contribuciones con las que contamos, señalamos las primeras en haber sido publicadas: Benítez Pérez (1992) en España (Madrid); López Chávez (1993) en México; Alba (1995) en República Dominicana; Valencia y Echeverría (1999) en Chile. Desde entonces, los estudios se han multiplicado en todas las provincias españolas hasta casi llegar a su completa conclusión.

⁵² Cfr. Samper Padilla, Bellón Fernández y Samper Hernández (2003).

(iii) El nivel sociocultural.

Se trata de una variable que se establece *a posteriori* a partir de informaciones personales sobre los participantes y de la asociación de tres factores, a saber, el nivel de estudios de los padres, su profesión y la renta familiar (López Morales, 1973; 1979), según los cuales se asigna una puntuación y se definen tres niveles: alto (20-24), medio (9-19), bajo (4-8).

(iv) La zona geográfica.

Ya contemplada en el estudio pionero de Gougenheim *et al.* (1964), permite averiguar si la zona de residencia –urbana o rural– influencia el tipo de vocabulario disponible.

(v) El centro de estudios.

Este factor sirve para definir si el centro escolar de los encuestados –público o privado– influencia el léxico disponible de los encuestados.

A continuación, como los datos recogidos siempre requieren ciertas enmiendas de manera que el corpus resulte unitario, se han establecido las directrices de edición del material. Se trata de una fase tanto importante como la selección de los CI, el diseño de la encuesta y su administración. El grupo de trabajo ha optado por armar un proceso de “limpieza” de los datos (Samper Hernández, 2002: 22) y ha decretado las siguientes normas⁵³:

1. eliminación de los términos repetidos;
2. corrección de la ortografía;
3. unificación ortográfica;
4. neutralización de variantes meramente flexivas;
5. unificación bajo un solo rótulo los derivados regulares que no suponen alteración del significado léxico;
6. unión de formas plenas y acortamientos;
7. uso de los paréntesis para aquellas lexías que aparecen con todos sus elementos constituyentes y con reducción de algún componente;
8. aceptación de las marcas comerciales totalmente lexicalizada, o en vía de lexicalizarse.

⁵³ Cfr. Samper Padilla y Samper Hernández (2006).

Para terminar, se ha desarrollado una nueva fórmula que calculase el índice de disponibilidad léxica (ID) que mejorase aquella empleada en el *Français Fondamental*. En efecto, unos de los objetivos de López Morales (1978; 1983; 1984; 1999) era la mejora de dicha fórmula de manera que cuantificase tanto la frecuencia como la posición en la que se actualizaban las palabras (el grado de espontaneidad) para poder detectar aquellas lexías que acudían primero a la mente ante un cierto estímulo (Lorán Santos y López Morales, 1983)⁵⁴.

Para ello, López Chávez y Strassburger Frías (1987; 1991) han presentado una nueva fórmula, que se sigue utilizando actualmente:

$$D(P_j) = \sum_{i=1}^n e^{-23\left(\frac{i-1}{n-1}\right)} \times \frac{f_{ji}}{I_1}$$

Donde:

- D(P_j)= disponibilidad de la palabra j;
- n= máxima posición alcanzada en el CI en la encuesta;
- i= número de posición de que se trata;
- j= índice de la palabra tratada;
- e= número natura e (2,718281828459045);
- f_{ji}= frecuencia absoluta de la palabra j en la posición i;
- I₁= número de informantes que participan en la encuesta.

De ahí que se haya decidido considerar los siguientes elementos para extraer los ID de las encuestas:

- (i) la frecuencia absoluta de cada palabra en cada posición de la lista;
- (ii) la frecuencia absoluta de cada palabra resultante de la suma de las diferentes frecuencias alcanzadas en cada posición;
- (iii) el número de encuestados;
- (iv) el número de posiciones alcanzadas en los centros de interés analizados;
- (v) la posición en que aparece cada palabra.

⁵⁴ En realidad, ya Michéa (1953) y Müller (1973) habían destacado la importancia de la posición de las palabras en las listas de léxico disponible.

Todo esto ha llevado también al desarrollo del *software* de tratamiento informático de los datos creado para el PPHDL, el *Dispolex* (Bartol Hernández *et al.*).

A este punto, pese a que no cabe en nuestro interés revisar toda la literatura científica, mencionamos los trabajos más relevantes en el panorama hispano, apreciando la rapidez con la que esta metodología se ha difundido en los países hispanohablantes del mundo⁵⁵.

Destacamos el papel de López Chávez que en sus trabajos (1992; 1995) ha ofrecido un primer análisis de lo que se ha denominado *léxico fundamental panhispánico*, mediante el cotejo de los resultados extraídos de los estudios de las comunidades de Madrid, Las Palmas, República Dominicana y Puerto Rico.

Nos parece obligado citar al grupo chileno encabezado por Echeverría que ha proporcionado en 1987 la fórmula para la medición del *índice de cohesión*, contribuyendo a profundizar las investigaciones de disponibilidad léxica desde una perspectiva cualitativa. Este indicador calcula el grado de cohesión semántica de los centros de interés y la coincidencia entre las respuestas de los participantes.

Aún más, a partir de este nuevo índice han desarrollado otro programa, el *DispoGrafo* (Echevarría *et al.*, 2008)⁵⁶, que permite el estudio de las redes semánticas que organizan el lexicón mental, conjugando las labores de la disponibilidad léxica con la psicolingüística “[...] basándose en las relaciones de secuencia de las palabras disponibles, permite observar de manera gráfica las relaciones que se establecen entre las unidades, pudiendo analizar, además, la robustez o debilidad de estos vínculos” (Sánchez-Saus, 2011: 84).

En lo que concierne la difusión del PPHDL en España, a finales de los años noventa, han empezado a organizarse estudios transversales por provincias o comunidades autónomas con el fin de examinar toda el área ibérica continental. Entre ellos, sobresale el proyecto de Galloso Camacho (2002), realizado en Salamanca, por haber introducido el análisis estadístico de los datos en el marco de la disponibilidad léxica para comprobar la significación de las diferencias cuantitativas detectadas. Entonces, se desprende la gran evolución de esta metodología finalizada al estudio del

⁵⁵ Cfr. Samper Padilla, Bellón Fernández y Samper Hernández (2003); Gallego Gallego (2014); Samper Padilla y Samper Hernández (2006); Paredes García (2015); López González (2014a); Hidalgo (2017a,); Aabidi (2019).

⁵⁶ Aunque ya había sido empleado en 2006 por Ferreira en su trabajo sobre la lengua inglesa.

español LM que ha implicado notables innovaciones teóricas y prácticas a partir de las múltiples aplicaciones y lecturas posibles de los resultados.

2.5 La disponibilidad léxica en ELE

Gracias al interés que ha despertado en López Morales, la disponibilidad léxica ha llegado al mundo hispano donde una multitud de proyectos ha contribuido al desarrollo de este ámbito investigativo que se ha extendido hasta abarcar la cooperación de la sociolingüística, la dialectología, la etnolingüística, además de la lingüística, la psicolingüística y la didáctica de lenguas. Dentro del marco de esta última disciplina, los trabajos centrados en ELE son la corriente investigativa que más se ha ampliado en los últimos años, de la cual apreciamos una cantidad notable de investigaciones sobre aprendices no nativos de español, idioma más trabajado en los estudios de esta índole.

Los primeros proyectos han adoptado las directrices del PPHDL y han contribuido, sin duda, a alargar su esfera de acción. Pero, en algún momento los investigadores han ajustado las pautas a sus propias exigencias y, en particular, a las de sus informantes que, por ser aprendientes extranjeros, necesitaban un trato diferente, lo que ha llevado a una renovación metodológica centrada en los nuevos encuestados.

El iniciador de esta nueva corriente ha sido Alberto Carcedo, quien ha aplicado los procedimientos originales del PPHDL en el análisis del léxico disponible de un grupo de estudiantes finlandeses de ELE. Se ha acercado al estudio de la disponibilidad léxica asumiendo las ventajas que se podían aportar a la enseñanza del vocabulario del español: entre otras, localizar las carencias de las propuestas didácticas y de los materiales; detectar y solventar los vacíos del alumnado. Para ello, ha planteado numerosos estudios en los que (i) observa el proceso de aprendizaje mediante el análisis del léxico disponible de los estudiantes tras la asistencia a cursos de ELE (1998); (ii) analiza el desarrollo de la competencia léxica en distintas etapas de la ASL (1999a); (iii) indaga sobre la interlengua y los errores más comunes (1999b); (iv) coteja sus resultados con los de los hablantes nativos (1999c, 2000a); (v) abarca la cuestión del componente cultural del vocabulario en la evolución del conocimiento del español (2000b).

Sin embargo, su proyecto más famoso y extenso (2000c) trata una muestra de 350 aprendientes finlandeses de ELE de diferentes niveles, cuya finalidad era ofrecer una imagen exhaustiva del léxico disponible del estudiantado finlandés. Se ha basado en los

dieciséis CI del *Proyecto Panhispánico*, pero ha aportado cambios en lo referente a las variables extralingüísticas tomando en consideración que los estudios sobre aprendices no nativos son distintos con respecto a los del PPHDL, que describen la norma léxica de una comunidad nativa (Carcedo, 2000c: 68-70). Además de los factores sociales clásicos ha introducido algunos inéditos, por lo que ha analizado, al final: el sexo, la lengua materna, el nivel de estudios, el conocimiento de otras lenguas románicas, el libro de texto utilizado. Todo esto le ha permitido corroborar que el léxico disponible de los alumnos no era conforme a sus reales exigencias comunicativas en español y, por ende, ha puesto de manifiesto que esta tipología de análisis y sus resultados debían tenerse en cuenta a la hora de diseñar un curso de ELE y crear materiales adecuados.

La tarea de Carcedo ha impulsado muchísimos proyectos dedicados a aprendices de ELE de diferentes nacionalidades, que estudiaban español en su país de origen, como en el caso del propio Carcedo, o bien en un contexto de inmersión lingüística.

Sacamos a colación, a este propósito, la aportación de Samper Hernández (2002), quien ha confeccionado un trabajo con 45 estudiantes de LM diferentes⁵⁷ procedentes de los *Cursos Internacionales de español* en la Universidad de Salamanca. Han participado informantes ingleses, italianos, japoneses y de otras procedencias (en números menores) que vivían en España. Pese a que este análisis se ha apoyado en la metodología de Carcedo⁵⁸, ha presentado algunas modificaciones. Primeramente, en lo referente a las variables sociolingüísticas: el sexo, la lengua materna, el conocimiento de otras lenguas (no simplemente románicas), el nivel de español (Samper Hernández, 2002: 19-20)⁵⁹. En segunda instancia, tomando conciencia de la falta de criterios específicos para la edición de las respuestas de los “nuevos” informantes, ha planteado algunas modificaciones significativas con respecto a las pautas del PPHDL (Samper Hernández, 2002: 16):

una mayor laxitud en la corrección gráfica y en el tratamiento de los extranjerismos y una mayor amplitud en las relaciones asociativas [ya que ha tenido que resolver] cuestiones

⁵⁷ Definido “de iniciación” debido a la exigüidad de la muestra (Samper Hernández, 2002: 16).

⁵⁸ Ha utilizado los mismos centros y, tal y como ha hecho Carcedo (1999c; 2000a), ha incluido un cotejo con sintopías nativas del que ha extraído unas conclusiones parecidas, esto es, que el léxico disponible de sus informantes era cuantitativamente inferior al vocabulario de los hispanohablantes nativos. Es más, ha comparado los datos de Carcedo (1999a), observando que los grupos poseían una competencia similar, no obstante los estudiantes de Salamanca han presentado un caudal ligeramente mayor de palabras y vocablos.

⁵⁹ Del análisis de las variables ha surgido que, en realidad, el sexo y el conocimiento de otras lenguas no afectan al total de palabras conocidas. Al contrario, el nivel de español ha surtido algún efecto, aunque sin la influencia esperada: la autora ha registrado un rendimiento mejor del nivel avanzado frente al inicial, pero, en contra, un descenso del avanzado al superior.

relativas a sus características lingüísticas como la presencia de palabras de la lengua materna de los hablantes, sus ya citadas peculiaridades ortográficas o los numerosos errores que presentan. (Hidalgo, 2017b: 109)

Hacemos patente que estos dos trabajos han sentado las bases para los estudios dedicados a discentes extranjeros de español, pues, se consideran los pilares en los que se funda la mayoría de los proyectos sobre disponibilidad léxica posteriores que han introducido a su vez cambios y mejoras, pero siempre emulando estos dos modelos⁶⁰.

Para empezar, citamos el trabajo de Frey Pereyra (2006; 2008)⁶¹ centrado en el conocimiento del vocabulario de 31 aprendices de ELE que vivían en la ciudad de Alcalá de Henares. La autora ha examinado dos corpus, ambos recopilados por los mismos informantes, uno de léxico disponible y otro de textos para analizar (cuantitativa y cualitativamente) el léxico pasivo y compararlo con el léxico activo utilizado en los relatos. Su propósito principal era la evaluación del uso del léxico disponible en los textos.

Šifrar Kalan (2009; 2012; 2014a; 2014b; 2016) ha dedicado varios estudios al léxico disponible de muestras de aprendices eslovenos de ELE estableciendo como objetivo privilegiado la aplicación pedagógica de los resultados de sus encuestas a la planificación léxica.

En un estudio más reciente (2018) ha extendido el campo investigativo hacia la sociolingüística proponiendo un análisis sobre la influencia que puede ejercer un periodo de estancia Erasmus en España sobre en el léxico disponible de un grupo de aprendices no nativos, en particular con referencia a la cultura española y a los estereotipos culturales.

Mencionamos también la Tesis doctoral de Sánchez-Saus (2011)⁶², por representar otra novedad en el panorama investigativo y que recordamos por la magnitud de las innovaciones. Este proyecto ha contribuido al desarrollo de los trabajos de disponibilidad léxica realizando un pormenorizado estudio estadístico de la muestra (322 sujetos), dividida según cinco factores extralingüísticos: sexo, lengua materna, nivel de español, conocimiento de otras lenguas extranjeras, frecuencia de uso del español.

⁶⁰ Para una revisión completa de la literatura sugerimos la lectura de López González (2014a), Hidalgo (2017a), Aabidi (2019).

⁶¹ El artículo fechado 2006 se ha publicado como estudio piloto de la Tesis defendida en 2008 en la Universidad de Alcalá de Henares.

⁶² Cuyo título es *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas* y de la que se ha publicado una monografía en 2016 titulada *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*.

Asimismo, ha modificado los estímulos temáticos clásicos adaptándolos a la lista de contenidos que el MCER recomienda para el nivel A1, pero, a la vez, intentando mantener lo más posible las semejanzas con los tradicionales para permitir el cotejo con otros trabajos.

Es más, ha introducido el análisis cualitativo-semántico de las respuestas mediante el uso del programa *Dispografo* gracias al cual ha podido (i) describir las unidades léxicas en función del centro de interés al que pertenecen; (ii) relacionarlas a su núcleo temático; (iii) etiquetarlas según categoría gramatical. Así que ha examinado fenómenos diatópicos y diafásicos como los cambios semánticos, la importación (extranjerismos, préstamos, calcos) y la generación léxica (derivación, composición, acortamientos).

Por último, es destacable la Tesis doctoral de Hidalgo (2017b) en la que se presenta otra innovación: además de los análisis del léxico disponible de los 440 aprendices sinohablantes encuestados, ha realizado un estudio comparativo con dos manuales de ELE utilizados en China para averiguar su adecuación a las necesidades del alumnado y determinar su posible incidencia en el conocimiento del vocabulario.

A estas alturas, nos parece un paso obligatorio comentar la situación actual de los estudios italianos ya que a pesar de los importantes resultados de tales trabajos y de su extensión geográfica, no se le ha dedicado en Italia la misma atención, pues, se ha incorporado recientemente a este tipo de estudios.

Contamos con pocas contribuciones, teniendo en cuenta además las altísimas cifras de estudiantes de ELE en este país, donde esta lengua es hoy la segunda LE más estudiada después del inglés, a todos los efectos y en todos los niveles, tanto en la enseñanza no reglada como en contextos institucionales.

El estudio de Caggiula (2013), de modestas proporciones, es una encuesta sobre el léxico disponible de 50 estudiantes italófonos de español (30 preuniversitarios y 20 adultos, universitarios o alumnos de escuelas privadas), divididos según las variables sexo, nivel de estudios y conocimiento de otras lenguas extranjeras.

Rubio Sánchez ha publicado dos trabajos sobre aprendices italianos preuniversitarios: el primero (2015) se ha realizado en contexto de inmersión en Salamanca con 36 informantes, cuya originalidad reside en la inclusión de cuatro CI (*España; Italia y partes del cuerpo; la ropa* –en italiano–). En el segundo (2017) ha encuestado a 173 estudiantes manteniendo los mismos campos y añadiendo otros:

[...] hemos introducido seis centros de interés novedosos: 4. Acciones y actividades habituales, 5. España y los españoles, 9. Aspecto físico y carácter, 10. La lengua española, 12. Italia y los italianos, 13. Partes del cuerpo (en italiano) y 14. Comidas y bebidas (en italiano). La incorporación de los centros 4. y 9. persigue la producción de palabras de categoría gramatical diferente al sustantivo (verbos y adjetivos, respectivamente). La introducción de 10. tiene como objetivo examinar las relaciones efectuadas por los encuestados y su opinión acerca de la materia de estudio. Por otro lado, con la inclusión de 5. y 12. examinaremos la concepción que tienen los informantes de la cultura española e italiana.³ En cambio, con la incorporación de 13. y 14. pretendemos efectuar una confrontación entre el número de vocablos que los encuestados saben en español y en su LM. (Rubio Sánchez, 2017: 145)

Nombramos también a Paolini (2017), cuyo objetivo principal era estudiar la disponibilidad léxica de un grupo de estudiantes universitarios italianos de ELE. Ha analizado cuantitativa y cualitativamente el léxico disponible de los informantes, limitando la selección a ocho de los campos semánticos clásicos y según una amplia gama de variables extralingüísticas: sexo, edad, carrera de estudios, zona geográfica de procedencia, nivel sociocultural, lengua materna, conocimiento de otras lenguas, años de estudio de español LE.

A continuación, apreciamos el proyecto de Del Barrio que se ha ultimado con la publicación de un diccionario de léxico disponible de aprendices itálofonos en 2018⁶³, tras algunos análisis preliminares (2016; 2017a; 2017b) en los que el autor ha insistido en la importancia de proponer estudios con aprendices italianos enmarcadas en este ámbito ya que

Se trata de un estudio que no se había emprendido hasta ahora y que sin duda alguna aportará una nueva perspectiva para la didáctica del léxico en ámbito itálofonos. [...] La bibliografía ha demostrado la utilidad que los diccionarios de disponibilidad pueden tener en la programación curricular de la enseñanza del léxico, en especial, por lo que concierne a la delimitación y la graduación del vocabulario. (Del Barrio, 2016: 141)

Para ello, ha combinado dos grupos de estudiantes, 49 preuniversitarios y 68 universitarios, que han realizado una prueba de disponibilidad léxica diseñada según los patrones del PPHDL, de la que se ha extraído los listados necesarios para la confección del diccionario. Esta investigación ha sido ampliada con el análisis comparativo del vocabulario disponible de 16 informantes bilingües italiano-español (de origen hispanoamericano).

Es más, ha aportado una novedad proponiendo un cuestionario sobre la disposición con la que los participantes se enfrentaban al español que le ha permitido

⁶³ Realizado en colaboración con Mae Vann.

conocer las actitudes del alumnado.

Incluimos en esta breve reseña Nalesso (2018a) donde hemos analizado el léxico disponible del CI “alimentos y bebidas” en un grupo de 25 aprendices para averiguar la eficacia de algunas actividades explícitas para el aprendizaje del léxico.

Asimismo, en Nalesso (2018b; 2019a) hemos realizado un estudio piloto de la presente Tesis con una muestra de 20 informantes encaminado a establecer cuál es el caudal léxico que se aprende tras la asistencia a un curso de ELE.

2.6 Otras aplicaciones de la disponibilidad léxica

La disponibilidad léxica no está limitada a los estudios de lingüística y didáctica, sino que hay otras disciplinas que benefician de sus resultados. Entre los años noventa del siglo pasado y el principio del siglo XXI, en efecto, su aplicación se ha dividido entre distintos campos donde se ha desarrollado un considerable número de proyectos radicados en esta metodología⁶⁴.

La sociolingüística muy pronto (López Morales, 1973) ha influenciado los estudios hispanos que se han dedicado a medir el influjo de las variables sociales en las respuestas de los encuestados (de ahí la inclusión de los condicionantes extralingüísticos en los análisis de disponibilidad léxica), cuyo objeto es examinar el fenómeno de la variación lingüística⁶⁵. En este marco, se proponen continuamente nuevos estudios desde perspectivas inéditas: se ha abordado, por ejemplo, un análisis sionomástico de los antropónimos (Fernández Juncal y Hernández Muñoz, 2019).

La dialectología, por su parte, ha utilizado estos datos para la comparación entre las variedades diatópicas de un idioma ya a partir del trabajo de Mackey (1971a), el cual ha comparado el léxico disponible francés de Canadá y de Francia. Esto permite comprobar las similitudes y las diferencias entre el léxico disponible de los sujetos que hablan la misma lengua en diferentes partes del mundo.

La psicolingüística ha planteado numerosos análisis de corte cognitivo debido a la naturaleza de las pruebas de tipo asociativo puesto que la producción del léxico disponible es una tarea cognitiva compleja en la que intervienen diversos elementos

⁶⁴ Carcedo (2000c: 47-56) y Samper Hernández (2002: 10-16) ofrecen una pormenorizada reseña de estas distintas aplicaciones.

⁶⁵ A este propósito, Samper Hernández (2002: 11) afirma “se ha comprobado que los hablantes de nivel sociocultural bajo tienen sistemáticamente una disponibilidad léxica menor que los de otros estratos”.

psicológicos, más allá de los puramente lingüísticos (Hernández Muñoz, 2006). Se han analizado en este marco: el acceso y la selección del léxico en L1 y L2; las estrategias de recuperación del léxico; las redes semánticas; la organización del lexicón mental (Hernández Muñoz y López García, 2014; Hernández Muñoz, Izura y Tomé Cornejo, 2014; Del Barrio, 2018; Tomé Cornejo, 2018; Gómez Devís, 2019). En este sentido, la disponibilidad léxica permite examinar las asociaciones mentales que se establecen entre las respuestas actualizadas en la ejecución de las pruebas, por eso Hernández Muñoz (2005) la describe como “una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología”.

Asimismo, la etnolingüística es otro campo donde esta metodología resulta útil ya que las palabras actualizadas en las encuestas reflejan las culturas y las costumbres de una comunidad. Mackey (1971a) ha sido el pionero en abordar esta aplicación; mientras que en el ámbito de ELE sobresale Carcedo (2000a; 2000b) que proporciona un análisis de las respuestas en función de las características culturales y geográficas de sus encuestados. Todo esto es de gran interés en los estudios sobre aprendices no nativos ya que permite observar la influencia de la cultura tanto del país de origen como de la comunidad a la que el alumnado se enfrenta.

Por último, los fines educativo-pedagógicos de la disponibilidad léxica son muchos en lo que se refiere a la didáctica de las lenguas, sea materna sea extranjera.

Tras la fase inicial empezada por el *Français Fundamental* destinado a la enseñanza de la LE, la mayoría de los estudios con nativos ha visto su aplicación en:

- (i) la planificación lingüística;
- (ii) la elaboración de materiales;
- (iii) los análisis de tipo cognitivo;
- (iv) la comparación de sintopías o de comunidades bilingües;
- (v) la averiguación de las carencias de las propuestas didácticas.

De manera similar, en los estudios dedicados a la ASL, en primer lugar, es indudable el provecho que se puede sacar de las listas de léxico disponible para la programación de la enseñanza y una selección del léxico ajustada a las necesidades comunicativas del alumnado⁶⁶. En segundo lugar, se investiga sobre la competencia

⁶⁶ Sin duda, es provechoso analizar los recursos de uso corriente en ELE para averiguar su adecuación a la

léxica de los aprendices –como es nuestro caso– para registrar el vocabulario activo y tomar conciencia de eventuales déficits.

Tal cual ha aseverado Carcedo (1998), la disponibilidad permite (i) conocer el vocabulario disponible de un grupo y cotejarlo con otro; (ii) analizar las fases del proceso de aprendizaje lingüístico; (iii) establecer cuáles son las lexías que se deben enseñar, evitando los términos innecesarios o poco productivos para el desarrollo de la competencia comunicativa. En este mismo contexto, señalamos el análisis de la interlengua, la transferencia intralingüística, la generación léxica, el uso de extranjerismos.

En resumidas cuentas, las posibilidades de estudio abarcan una gran variedad de aplicaciones teóricas y prácticas y, en particular, en el marco de la didáctica de ELE es posible lograr una amplia serie de objetivos que Paredes García (2015) retomando las palabras de (Carcedo, 1998) alista de la siguiente manera:

- (i) analizar el léxico disponible de los aprendices en un determinado momento;
- (ii) definir las etapas de aprendizaje;
- (iii) estudiar las interferencias de la L1 y de otras LE conocidas;
- (iv) cotejar el vocabulario de los participantes con otros grupos de estudiantes y con los nativos;
- (v) establecer cuáles son las unidades poco rentables o los déficits léxicos;
- (vi) proporcionar comparaciones culturales;
- (vii) examinar la adecuación de los manuales y los otros materiales pedagógicos;
- (viii) eliminar el léxico superfluo de los contenidos didácticos y llevar a cabo una selección de los contenidos apropiada;
- (ix) ampliar la cobertura de los análisis hacia la terminología (español para fines específicos).

ASL. Benítez Pérez (1994) ha sido el primero que ha correlacionado los resultados de su investigación sobre el léxico disponible en la zona metropolitana de Madrid y el léxico presentado en los manuales más difundidos en aquel periodo. A raíz de este estudio, otros autores se han planteado el mismo análisis comparativo, entre otros, Hidalgo (2017b: 571-692).

3. LA RIQUEZA LÉXICA

3.1 Introducción

A la hora de examinar la competencia léxica de un sujeto, de un grupo o de una comunidad, es indudable que los análisis del léxico disponible son un método provechoso, sin embargo, pese a las muchas ventajas que conllevan para la investigación y la práctica didáctica, permiten conocer exclusivamente el vocabulario que se activa ante un estímulo temático y no aportan otras informaciones sobre la capacidad de utilización real de las palabras, esto es, la capacidad productiva. Para examinar enteramente el dominio del léxico de nuestros encuestados nos parece oportuno un ulterior análisis a través de la metodología de la *riqueza léxica*. Esta completa la disponibilidad mediante el estudio de la producción textual de una muestra:

[Aunque] el léxico está siendo centro de atención desde todos los puntos problemáticos del estudio de la lengua [sobre todo con] trabajos cualificados de vocabulario disponible (palabras aisladas) en escolares, recogidos por centros de interés, nos falta también un vocabulario recogido en textos producidos por los escolares [...], lo que daría nivel de uso en producción real de los escolares y representaría un conocimiento más exacto que el del vocabulario disponible. (Battaner, 1992 *apud* Cuba Vega, 2016: 81)

Es cierto que un buen uso del vocabulario resulta esencial en la determinación de la calidad de un texto, puesto que suscita una impresión positiva en el receptor (Laufer y Nation, 1995: 307) y, sin duda, de un buen nivel de riqueza léxica se desprende un buen uso de la LO.

La riqueza ha sido definida y utilizada de forma variada en diferentes proyectos, pero, en general, se asume que es una metodología que mide el léxico empleado en un texto centrándose en su variedad. El objetivo principal es el análisis de la habilidad de manejar un repertorio de vocablos en un discurso, ya sea oral o escrito. Se trata de aplicar indicadores de naturaleza cuantitativa y comparativa finalizados a examinar el léxico activo de los encuestados (Reyes Díaz, 2007: 147).

3.2 Estudios pioneros

Los estudios sobre riqueza léxica surgieron a mediados del siglo XX en Francia, en concomitancia con los de disponibilidad, basándose en textos producidos por hablantes nativos. Si bien se hayan planteado múltiples propuestas, los primeros patrones de cálculo

siguen aun vigentes.

Guiraud (1954) teniendo en cuenta la distinción entre *palabra* y *vocablo*, realizó la primera investigación sobre riqueza léxica fijando como elemento básico de análisis la relación entre estos dos ítems en un texto escrito⁶⁷. A raíz de este estudio, se sentaron las bases metodológicas para el cálculo de la riqueza léxica que se han ido refinando cada vez más en los trabajos posteriores.

Tras presentar en la primera parte de su trabajo las medidas estadísticas aplicables a la medición del vocabulario, el autor precisa que su objetivo no consistía en el estudio y la evaluación del léxico extraído de los textos poéticos propuestos en la segunda parte, sino que el análisis le servía como punto de partida para corroborar las hipótesis y los planteamientos definidos “teóricamente y en abstracto” (Guiraud, 1954: 74-75).

Para ello, desarrolló las siguientes fórmulas:

$$R = \frac{V}{N} \qquad R = \frac{V}{(2)N}$$

Donde:

R= riqueza del vocabulario;

V= palabras (nocionales y funcionales) del texto;

N= longitud del texto.

Al considerar simultáneamente las dos categorías de palabras se debía utilizar la primera fórmula; en cambio al tener en cuenta exclusivamente las palabras semánticas (*mots-forts*) se aplicaba la segunda, en la que se duplica la extensión del texto ya que parece que las palabras nocionales cubren su mitad.

Más tarde, Müller (1968) afirmó que cuantificar el vocabulario de un texto supone dos procesos distintos y complementarios, que pueden llevarse a cabo al mismo tiempo o en dos etapas seguidas: el conteo de todas las palabras que forman el texto y de los diferentes vocablos detectados. Se insistió en la relevancia de la repetición de las unidades léxicas dentro de una muestra, esto es, en la distinción entre *tokens* y *types* (en los que se

⁶⁷ A este propósito,

La léxico-estadística ha sido el primer peldaño en la constitución de la estadística lingüística; [...] abarca el conjunto de operaciones, a veces sumamente complejas, que toman como unidades de trabajo las palabras y los vocablos; la palabra, unidad del texto, y el vocablo, unidad del léxico. (López Morales, 2011: 17)

basa la mayoría de las fórmulas empleadas en los trabajos empíricos actuales).

Sucesivamente, Ham (1979) amplió la noción inicial de riqueza léxica, válida hasta la fecha, abogando por incluir el criterio de frecuencia asumiendo que es imprescindible indagar también las ocurrencias de los vocablos en los textos, además de los otros factores, porque no es posible contabilizar el nivel de riqueza sin tener en cuenta la dispersión o la concentración de los *types*.

En lo que se refiere a los estudios en ámbito hispano, consideramos pionero el proyecto de López Morales (1984) que marcó un punto de inflexión en el panorama investigativo. Introdujo notables cambios y reformuló las directrices metodológicas. Aplicó nuevos procedimientos, operaciones matemáticas y estableció nuevos indicadores para medir la riqueza de vocabulario: la *diversidad léxica* y el *intervalo de aparición de palabras de contenido nocional* (IAT).

El método prevé el empleo de dos medidas interrelacionadas. En primer lugar, se cuantifica la riqueza computando el porcentaje de vocablos (PV) con respecto a las palabras totales del texto (N), mediante la fórmula:

$$PV = \frac{V \times 100}{N}$$

Este procedimiento supone la división del total de vocablos (V) por el total de palabras del texto (N) y, luego, la multiplicación por 100 que da lugar a la *diversidad léxica*, como “indicador grueso”.

En segundo lugar, se contabiliza el IAT, es decir el intervalo de aparición de las palabras temáticas (PN):

$$IAT = \frac{N}{PN}$$

El resultado de esta división es consecuencia de la concepción que tenía el autor de la riqueza léxica (PN), esto es, la totalidad de *tokens* del texto divididos por el total de las palabras llenas. Resulta que cuanto mayor es el número de estas lexías, menor es el IAT.

De esta forma, se proporcionaron los pasos a seguir para determinar la riqueza léxica de un texto, según este *iter* que, antes de la aplicación de las dos fórmulas descritas, supone el conteo del total de: (i) las palabras; (ii) los vocablos; (iii) las palabras sin

contenido semántico; (iv) las palabras con contenido semántico (López Morales, 2011).

López Morales trabajó sobre la riqueza léxica de informantes de diferente nivel de educación (un alumno de cuarto grado de primaria; una alumna de segundo curso de escuela secundaria; un periodista profesional) con el objetivo de compararlos para explicar tanto su propuesta innovadora, como las ventajas aportadas por este nuevo método.

Por su parte, Linnarud (1986) ofreció un estudio sobre la riqueza léxica de la producción escrita de aprendices suecos de bachillerato que estudiaban inglés como LE y la comparó con un corpus de redacciones de un grupo de anglohablantes nativos coetáneos, fijando por primera vez un límite de palabras en las muestras porque la diversidad léxica no se puede comparar si la extensión de los textos es distinta.

A finales del siglo, Tesitelová (1992) sugirió otra implementación de la metodología con la inclusión de nuevos elementos en los cálculos: (i) la repetición de las palabras de un texto; (ii) la fuerza de la zona de palabras en baja frecuencia (1-10); (iii) la dispersión frente a la concentración.

3.2.1 *Los índices de medición de la riqueza léxica*

A estas alturas, se desprende que el objetivo de la riqueza léxica es cuantificar la variedad y el tamaño del vocabulario usado en un corpus, de ahí que este concepto se utilice a menudo como sinónimo de diversidad léxica (Arnaud, 1984; Jarvis, 2013) aunque, en realidad, abarca una noción más amplia que deriva de diferentes índices.

En este trabajo seguimos la taxonomía propuesta en Read (2000), quien ha revisado la literatura sobre la evaluación del vocabulario y dedicado un apartado especial a la riqueza léxica, concebida como un conjunto de componentes que concurren a la medición estadística de la producción escrita, es decir aquella gama de características que un texto debe presentar para ser “rico”. Los índices son:

(i) La densidad léxica.

Mide el porcentaje de palabras semánticas en proporción al número total de palabras dentro de un texto, que se considera “denso” si tiene un alto número de unidades de contenido nocional (a saber, sustantivos, verbos predicativos, adjetivos, adverbios).

Laufer y Nation (1995) han criticado este índice porque sostienen que no era adecuado para indagar el uso concreto del vocabulario ya que podía verse afectado por

las propiedades sintácticas y cohesivas de un texto. Una exigua presencia de palabras atemáticas podría deberse al gran empleo de oraciones subordinadas o de elipsis típicas de un texto escrito. Al contrario, Johansson (2008: 61) ha declarado que el lenguaje escrito tiende a presentar mayor densidad con respecto al oral.

(ii) La variación léxica.

Es el índice que más se identifica con la medición de la riqueza: calcula la diversidad del vocabulario empleado en función de la cantidad de los vocablos frente a la cantidad de las palabras utilizadas. Consiste en calcular el porcentaje de los *types* sobre los *tokens* aportados por un autor. Se define también *Type/Token Ratio*, o *riqueza media* (Capsada y Torruella, 2017: 350), ya que sin computar las cifras en porcentajes refleja otra perspectiva de la relación entre palabras repetidas y no repetida.

Este indicador es sensible a la longitud del texto, por tanto, al trabajar con muestras de diferente tamaño es imprescindible aplicar medidas de corrección o normalizar los datos⁶⁸.

(iii) La sofisticación léxica.

Contabiliza el porcentaje de palabras de nivel avanzado (definidas *sofisticadas*) frente al total de palabras léxicas que se encuentran en un texto. De nuevo, Laufer y Nation (1995) han tachado el uso de esta medida (sobre todo en las investigaciones dedicadas a aprendices extranjeros) esencialmente porque es necesario decidir qué significa que una palabra es *sofisticada* ya que, a menudo, pertenece a aquel conjunto de unidades de baja frecuencia que tampoco usuarios nativos de una lengua conocen y utilizan (Arnaud, 1984; Linnarud, 1986).

(iv) La individualidad (o originalidad) léxica.

Definida individualidad por Linnarud (1986) y, después, originalidad por Laufer (1991), este índice establece la dispersión o la homogeneidad léxica dentro de un corpus. Consiste en la comparación del texto de un sujeto con respecto al entero muestreo mediante el cálculo del porcentaje de *tokens* que este actualiza y que nadie más dentro

⁶⁸ Se han propuesto medidas alternativas que extraigan datos fiables. Capsada y Torruella (2017: 351) han tratado la cuestión ya que:

[...] no es un valor constante, sino que va disminuyendo a medida que va aumentando la cantidad de tokens del texto que cogemos para su medida. Esta disminución de la TTR se explica a partir de la relación existente entre el número de tipos que van apareciendo con respecto al número de tokens del texto.

del grupo emplea.

Este indicador no es estable porque depende del comportamiento de cada miembro encuestado: en el caso de que el grupo cambie se produciría también una variación del índice (Laufer y Nation, 1995).

(v) El número de errores.

Read (2000: 204-206) ha defendido el recuento de los errores como manera de indicar el grado de riqueza léxica de un texto, al ser un valioso indicador de la competencia expresiva. Lo que resulta esencial, en este caso, es establecer previamente

1. qué es un error léxico;
2. cómo distinguirlo de otro tipo de error o de una falta menor;
3. cómo categorizarlo y valorarlo en la producción analizada.

Algunos de los índices presentados manifiestan algunas limitaciones que implican que no sean utilizables en todo tipo de análisis, a menos que, por ejemplo, los textos considerados tengan la misma longitud o que las características del muestreo sean iguales⁶⁹.

Claro está que antes de proceder a la aplicación de dichos índices es necesario:

- (i) identificar la unidad básica del análisis;
- (ii) tener en cuenta el tamaño de los textos del corpus;
- (iii) decidir cómo ajustar las medidas para que los resultados sean fiables;
- (iv) establecer patrones para el tratamiento de los errores⁷⁰.

3.3 La riqueza léxica en el mundo hispánico

López Morales ha introducido la riqueza léxica en el panorama investigativo

⁶⁹ Para ello, Laufer y Nation (1995) han planteado otro método de medición basado en el criterio de frecuencia. Se trata del *Perfil de Frecuencia Léxica* (LFP) que mide el porcentaje de las palabras a diferentes niveles de frecuencia, completando las informaciones procedentes de los otros indicadores. Pretendían validar el uso del LFP comparando su corpus de trabajo (producción escrita por aprendices de inglés LE) con un corpus de control (producción escrita por anglófonos nativos) en función de cuatro bandas de frecuencia: (i) la lista de las 1.000 palabras más frecuentes de la lengua inglesa; (ii) la lista de las segundas 1.000 palabras más frecuentes de la lengua inglesa; (iii) la lista académica *University Word List*; (iv) otras palabras, que no formaban parte de ninguno de estos registros. El resultado ha corroborado la rentabilidad de esta medida, pero, como se trata de una propuesta realizada *ad hoc* para el inglés y no aplicable al español, no nos adentramos en ulteriores detalles.

⁷⁰ En nuestro caso, es muy útil Barbasán Ortuño (2016), quien se dedica al error léxico en la interlengua del italiano y español y fija normas que permiten distinguir entre las faltas que no perjudican el conocimiento del vocabulario y los errores léxicos.

hispanico, favoreciendo el desarrollo de proyectos destinados a la descripción del vocabulario activo de hablantes nativos de español, bien en Hispanoamérica bien en España. Tras su trabajo de 1984, ha empezado una corriente de estudios que siguen sus directrices, aunque con cambios teóricos y procedimentales.

En México, Ávila (1986) ha estudiado textos escritos por niños de primaria aplicando tres análisis distintos que han confluído en la medición de la riqueza léxica del corpus. En primera instancia, considerando relatos de la misma longitud (100 palabras) ha contabilizado los vocablos recogidos en el conjunto textual poniendo en relación el total del *types* (V) con la extensión del texto (N); después, ha calculado el índice de densidad léxica (D), dividiendo los *types* (T) por el número total de *tokens*, mediante la siguiente operación matemática:

$$D = \frac{T}{N}$$

En segunda instancia, tras ordenar las palabras diferentes según el criterio de frecuencia, se ha ocupado de los textos de distinta extensión. Ha recopilado una lista de vocablos cuyas frecuencias sumadas permitían cubrir el 10%, el 20%, el 100% del total de la muestra. Este patrón sirve para cotejar el léxico sin que la longitud de cada uno condicione los resultados. En este estudio se han considerado también algunas variables extralingüísticas, como se suele hacer en los análisis de disponibilidad léxica, para averiguar si influyen en el rendimiento de los informantes o no: ha surgido que el nivel sociocultural y el sexo surten algún efecto.

En 2001, el mismo autor focalizándose en la importancia del cotejo en este campo investigativo, ha publicado un estudio comparativo de gran tamaño sobre la medición de la riqueza léxica del español en los medios de comunicación pública del Mundo Hispánico (radio, televisión y prensa).

Haché (1991) ha trabajado con un corpus de textos narrativos de escolares dominicanos estratificados según variable (sexo, edad, centro escolar, nivel sociocultural) para examinar el número de vocablos aportados y el IAT. Los resultados más relevantes han destacado la supremacía de los niños frente a las compañeras y de los informantes de edad mayor; mientras que el nivel sociocultural no ha parecido influir en la variedad léxica, pero si en el uso de palabras de contenido nocional. Asimismo, se ha puesto de relieve que es fundamental analizar textos del mismo tamaño ya que estudiar muestras

variadas incide en los cómputos, favoreciendo los textos más breves: los que tenían un número menor de palabras han alcanzado valores mejores.

A partir de los planteamientos de Haché se ha corroborado que después de las primeras 100 palabras el cálculo de la riqueza léxica se distorsiona. Esto ya había sido demostrado por Linnarud (1986) y explicado por López Morales (2011: 24):

[...] sabemos que los datos se desvirtúan al pasar un texto de las 100 palabras. [...] Necesitamos, pues, para tener metas claras, exámenes de textos escritos que se consideren ilustrativos, de los que se pueden tomar varias calas pero nunca superiores a las 100 palabras.

Por su parte, Echeverría *et al.* (1992) han examinado textos escritos por escolares chilenos, repartidos según las variables sexo, nivel sociocultural, tipo de escuela, zona geográfica de procedencia, tipo de texto. Las conclusiones más relevantes muestran que el tipo de discurso incide en la capacidad productiva: el narrativo permite escribir más vocablos (las opciones eran narrativo, descriptivo, expositivo y argumentativo).

Portela (1992) en su trabajo realizado en República Dominicana ha analizado textos narrativos producidos por los mismos estudiantes universitarios en función de dos condicionantes: el momento de realización de la prueba y las variables sexo, nivel socioeconómico y sociocultural, procedencia geográfica. En las conclusiones no ha destacado diferencias significativas entre las dos sum ministraciones; mientras que, en lo referente a los factores sociales, los datos se han inclinado a favor de los informantes de nivel social alto, procedentes de las zonas urbanas y de la enseñanza privada.

Cintrón Serrano (1992) ha corroborado la relación entre el nivel de riqueza léxica y las variables sociales. Los resultados han revelado que los factores extralingüísticos han influido en la riqueza del vocabulario: las mujeres, los niveles socioculturales más altos y los estudiantes de nivel más avanzado han presentado los índices mejores.

Igualmente, Andrés Pérez (1997) ha analizado un conjunto de textos narrativos para averiguar la incidencia de los factores extralingüísticos sexo, nivel socioeconómico, edad, nivel escolar en la determinación del grado de riqueza léxica descubriendo que todos, en cierta medida, han influido en los datos finales.

En Tenerife, Torres González (1999; 2003a; 2003b) ha trabajado con un corpus de textos narrativos redactados por estudiantes de Primaria, ESO y Bachillerato. Ha examinado relatos de 100 palabras a través de las variables curso, tipo de centro, sexo, nivel sociocultural.

Además de los índices que ya se habían empleado en investigaciones anteriores (número de vocablos, IAT, promedios de palabras nocionales), ha incluido un nuevo indicador: el *índice de hápax*, que se utiliza para detectar aquellas unidades que solo aparecen una vez en un texto y concurre a establecer la variación léxica

Reyes Díaz (2010) ha analizado un corpus de textos narrativos producidos por estudiantes grancanarios de Bachillerato (repartidos en función del tipo de centro escolar, el sexo, el nivel sociocultural) para investigar su destreza escritora, mediante el porcentaje de vocablos, el IAT y el índice de hápax. A la luz de los hallazgos de los estudios precedentes, ha fijado un límite máximo de 100 palabras por cada texto, corroborando que la extensión textual influye en la determinación de la riqueza léxica.

Madrigal y Vargas (2016) se han concentrado en estudiantes universitarios de Costa Rica porque los profesores habían lamentado ciertas deficiencias léxicas, en particular una falta de variación del vocabulario en los textos académicos. Los objetivos eran: (i) calcular la riqueza léxica; (ii) cotejarla con la producción de escritores profesionales; (iii) identificar las causas de los déficits; (iv) por último, sugerir estrategias para un rendimiento mejor de los estudiantes.

Todo esto hace patente que desde finales de los años ochenta, en distintos países hispanos, se han impulsado numerosos estudios sobre riqueza léxica que evaluaran el uso del español por parte de hablantes nativos, cuyos objetivos, a nuestro parecer, se pueden asemejar a los del PPHDL, desde otra perspectiva analítica.

3.4 La riqueza léxica en ELE

En el marco investigativo del español LE, a diferencia del interés por el análisis de disponibilidad, no se ha promovido un número suficientes de proyectos dedicados a aprendices no nativos.

El pionero en este ámbito es García Rosas (1996), quien ha analizado un corpus de redacciones escritas por estudiantes extranjeros en función del tipo de texto (descriptivo, narrativo, argumentativo), del nivel de ELE (intermedio, avanzado, superior), del sexo, de la edad y de la LM. Los resultados más relevantes han mostrado, otra vez, que el modo discursivo que ofrece un rendimiento mejor es el narrativo; la edad y la lengua materna no influyen en la riqueza léxica; y, por último, los informantes que

tienen un mayor dominio lingüístico aportan un número mayor de palabras y vocablos, como era esperable.

Cuba Vega y Cuba (2004) han estudiado el vocabulario de aprendices brasileños de ELE en la Escuela Latinoamericana de Medicina (Universidad Internacional de Cuba), analizando la riqueza léxica según los índices de variación e IAT en un corpus de relatos formados por 100 *tokens*.

Lucas Puerta (2006) se ha dedicado a la competencia léxica en español de aprendices francófonos de ascendencia hispanófono, teniendo en cuenta las variables sexo, formación, contacto con la LE/estancia en España⁷¹. El análisis del corpus, recopilado con redacciones de tipo narrativo-descriptivas de 300 palabras, se ha centrado en la medición de la riqueza léxica y ha abarcado también un estudio paradigmático y sintagmático de los textos.

Contamos también con el trabajo de Berton (2014), que ha analizado un conjunto de textos narrativos de diferente extensión, escritos por alumnos de un instituto sueco que estudiaban español como tercera lengua (L3). La evaluación de la diversidad léxica es el foco del análisis, que pretendía arrojar luz también sobre su relación con el conocimiento léxico-morfosintáctico de los informantes. La prueba de elicitación de datos contenía un test sobre la competencia lingüística de la población investigada (estandarizado según las pruebas del Instituto Cervantes), un cuestionario para la recogida de informaciones lingüísticas adicionales y un apartado dedicado a la redacción de los textos. Los resultados han corroborado, especialmente, que la variedad del vocabulario es sensible al número de *tokens* utilizados en cada texto: a mayor extensión de la muestra se correspondía una menor variación.

Wang (2016) ha trabajado con aprendices sinohablantes en la Facultad de Español para No Hispanohablantes de la Universidad de La Habana, examinando un corpus de textos compuestos por 100 palabras, según las directrices de Cuba Vega y Cuba (2004). Los encuestados, tras la administración de la prueba, se han dividido en dos grupos

⁷¹ Se trata de aprendices de español L2 que

[...] han recibido enseñanzas de español en el marco del programa “Clases Complementarias de lengua y cultura españolas”. Programa dependiente del Ministerio de Educación y Ciencia, y concebido para mantener la lengua y cultura españolas de aquellos alumnos de ascendencia española que, no pudiendo recibir dichas enseñanzas en el marco del sistema educativo francés, desean mantener los vínculos culturales y lingüísticos con España. (Lucas Puerta, 2006: 3)

(pregrado y postgrado) para detectar si se manifiesta una evolución de la riqueza léxica a medida que sube el nivel lingüístico.

En Italia, Basso (2017) propone un estudio sobre la riqueza léxica en la producción escrita de aprendices itálofonos de ELE, cuyas finalidades eran (i) analizar el índice de originalidad léxica para averiguar si el aprendizaje del léxico era homogéneo; (ii) comprobar cómo varían los índices de variedad y densidad en función de la longitud de los textos; (iii) examinar el grado de incidencia de los errores cometidos en la variación del vocabulario. Las conclusiones han confirmado las hipótesis de partida: (i) el grado de homogeneidad de los textos era apreciable; (ii) la riqueza ha sufrido una disminución al considerar una cantidad mayor de palabras, con lo cual los textos se hacen menos variados y “densos” (al extender el tamaño textual es difícil limitar las repeticiones y seguir actualizando unidades semánticas nuevas); (iii) la inclusión o exclusión de los errores en la aplicación de las fórmulas no han implicado modificaciones significativas.

También con estudiantes italianos es el trabajo de Nalesso (2019b) dedicado al análisis de la riqueza léxica en virtud de la originalidad léxica; la variación léxica; la *Type/Token Ratio*; la densidad léxica; el IAT en textos especializados de historia del arte de diferentes longitudes escritos por estudiantes universitarios, que ha corroborado la hipótesis de que cuanto más largo es un texto, menos rico es el vocabulario.

Por último, la reciente contribución de García Marcos (2019) se ocupa de la adquisición del español L2 por parte de un grupo de niños de escuela primaria, inmigrantes en la zona de Almería. El autor ha dedicado un apartado a la medición de la riqueza léxica producida en un corpus de textos narrativos escritos (compuestos por un mínimo de 200 palabras) con los objetivos de evaluar la variación del vocabulario, la ortografía y la madurez sintáctica; comparar estos datos con los textos de los compañeros nativos. Los resultados han puesto de relieve que la zona de residencia no tiene efectos, mientras que la procedencia y el sexo inciden en la capacidad expresiva, favoreciendo, respectivamente, los informantes nativos y las mujeres.

3.5 Otras aplicaciones de la riqueza léxica

Al igual que hemos observado para la disponibilidad léxica, la riqueza también puede ser explotada en ámbitos distintos de la mera investigación lingüística.

En este sentido, destacamos aquí la publicación de contribuciones encaminadas a

una revisión o una modificación de la metodología práctica. De entre estos, Capsada y Torruella (2017) han reexaminado la literatura sobre el asunto y los métodos de medición vigentes para proponer un nuevo protocolo empírico.

Asimismo, como hemos podido apreciar en el marco de los estudios hispanos, la sociolingüística se ha interesado por el rol que ejercen los condicionantes sociales, culturales y económicos en la capacidad expresiva de los individuos. En este sentido, Roberto, Martí y Salamó Llorente (2012) han puesto de manifiesto la utilidad de esta metodología “para predecir atributos demográficos latentes en textos de opinión del español”. Su objetivo era determinar hasta qué punto la riqueza léxica puede predecir el sexo, la edad y la procedencia del autor de un texto.

En la misma línea, Ávila Muñoz (2016) ha cotejado lo datos recogidos en diferentes estudios para averiguar qué patrones extralingüísticos condicionan la variación del vocabulario, y cuánto. Ha destacado que, tanto desde un punto de vista cualitativo como cuantitativo, se utilizan palabras y cantidades diferentes en virtud del nivel social, el grado de educación, el acceso a la cultura y la profesión.

Sin embargo, los buenos resultados obtenidos por la investigación sobre riqueza léxica hacen hincapié en su utilidad para la didáctica de LE ya que, además de medir la capacidad de uso del vocabulario de un grupo, ofrece la posibilidad de utilizar los resultados de los análisis para la creación de actividades, UD y materiales específicos finalizados a incrementar la capacidad productiva y la “calidad de escritura” (López Morales, 2011), ligadas tanto a la fluidez y precisión en el discurso, como a las posibilidades comunicativas de los aprendices. Como la competencia léxica coincide con la suma de las palabras conocidas, las pasivas y las derivables, la tarea del docente es trabajar para el desarrollo de dicha competencia de manera que el conocimiento pasivo se convierta en activo, es decir que cada alumno “vaya pasando unidades de su léxico a su vocabulario” (Lamíquiz, 1974: 401 *apud* Reyes Díaz, 2010: 147).

De la misma manera, todo tipo de análisis de riqueza léxica (cuantitativo, cualitativo, comparativo) se revela una herramienta útil en los procesos de evaluación de los conocimientos y en la programación didáctica puesto que detecta lo que ya conocen los aprendices antes de proponer otro *input*. En otras palabras, permite conocer el dominio del léxico activo de un sujeto o de un grupo y tener constancia del vocabulario manejado para programar la enseñanza de las unidades léxicas desconocidas (Reyes Díaz, 2007).

En resumidas cuentas, las posibilidades y las ventajas ofrecidas son muchas y pueden llevar a resultados fructíferos en distintos campos. Haché (1991 *apud* Reyes Díaz, 2010: 149) presenta una lista de aplicaciones utilizadas en estudios realizados con niños hispanohablantes nativos, exportable a nuestro contexto:

1. Las pruebas son de utilidad para diagnosticar qué vocabulario conocen los alumnos, de manera que se pueda actuar en consecuencia.
2. El vocabulario activado por los niños en las pruebas indicaría qué voces habrían de incluirse en los cuentos de los más pequeños para facilitarles la comprensión del contenido (que tal vez podría bloquearse por el sobreesfuerzo que exigiría el empleo de un lenguaje plagado de voces desconocidas) y el gusto por la lectura.
3. Las listas de los lexemas activados podrían contrastarse con el vocabulario de los textos escolares y de las lecturas infantiles con el fin de comprobar si existe adecuación.
4. Las listas aportarían información para la elaboración de diccionarios infantiles adecuados al entorno lingüístico y cultural circundante.
5. Las listas constituyen un dispositivo recomendable para la actividad docente a la hora de establecer una valoración certera del desarrollo del vocabulario activo de cada uno de los alumnos y de la totalidad de estos en un grupo determinado.

A raíz de estos planteamientos trasladamos las mismas finalidades a ELE:

1. Las pruebas permiten averiguar qué y cuánto vocabulario conocen los aprendices para que se pueda adaptar la enseñanza en consecuencia.
2. El vocabulario activado por los aprendices en las pruebas indicaría qué voces habrían de incluirse en los materiales didácticos para facilitar la comprensión del contenido y el gusto por la lectura en la LO.
3. Las listas de los lexemas activados podrían contrastarse con el vocabulario de los manuales y de las lecturas destinadas a este tipo de aprendiz con el fin de comprobar si está en línea con sus necesidades.
4. Los resultados proporcionarían información útil para la elaboración de diccionarios de aprendizaje para no nativos.
5. Los datos son un recurso provechoso para los docentes durante la evaluación

del desarrollo del vocabulario activo de cada alumno y de su estudiantado en general.

Se trata de ejemplos de la posible explotación de la riqueza léxica ya que, como se ha observado, no hay un suficiente número de investigaciones. En efecto, a pesar del provecho que se puede sacar de estos estudios y de los resultados alcanzados hasta ahora, se ha planteado un número esporádico de trabajos dedicados a hablantes nativos o no nativos de varias lenguas (entre otros: Linnarud, 1986; Engber, 1995; Jarvis, 2002; Dewaele y Pavlenko, 2003; Laufer, 2003; Spreafico, 2003; Johansson, 2008; Malvern y Richards, 2009; Lemmouh, 2010; Lindqvist, 2010; Crossley, Salsbury y McNamara, 2012; Šišková, 2012; Vedder y Benigno, 2013).

Análogamente, se ha abierto el camino de los análisis sobre estudiantes de ELE, pero aun son muchos los pasos por cumplir para que esta metodología se difunda como instrumento habitual en el panorama investigativo. Por eso, es deseable fomentar el planteamiento de nuevos estudios y grupos de investigación que profundicen los conocimientos adquiridos hasta la fecha e incrementen el protocolo de experimentos empíricos.

Todo esto destaca que la riqueza léxica, como la disponibilidad, tiene un doble aprovechamiento, en campo didáctico e investigativo, porque, además de ser una herramienta para la valoración de la competencia léxica, permite otras aplicaciones, ya que como ha aseverado uno de los autores pioneros en este campo, López Morales (2011: 27): “Vale la pena rescatar aquella experiencia, actualizar esos materiales y proseguir con la tarea.”

CAPÍTULO III.

METODOLOGÍA DE LA INVESTIGACIÓN

1. PAUTAS GENERALES

Tras el recorrido historiográfico, la reseña del estado de la cuestión y el establecimiento del marco teórico de los apartados anteriores, exponemos la metodología empleada en la parte empírica de la investigación. Detallamos: (i) la selección y la distribución de la muestra; (ii) las técnicas de elicitación que han estimulado la producción de datos por parte de los informantes, esto es, la prueba de medición de la competencia léxica y los procesos de recogida; (iii) los criterios de edición y el tratamiento, informático y estadístico, del material.

Aunque nos basamos en los patrones expuestos en el marco teórico, algunas de las decisiones tomadas son novedosas con respecto a otras investigaciones. Nuestro objetivo es analizar y evaluar la competencia léxica de los aprendices italianos de ELE: ofrecemos una imagen del procesamiento y del dominio del léxico que sea lo más fidedigna posible por lo que aportamos algunas modificaciones de lo ya practicado en otros estudios de la misma naturaleza.

Nos centramos en dos métodos empíricos, la disponibilidad y la riqueza léxica, encaminados a medir el vocabulario que un estudiante sabe y puede activar en una situación comunicativa ya que:

La riqueza, el alcance y el control del vocabulario son parámetros importantes en la adquisición de la lengua y, por ello, de la evaluación del dominio de la lengua que tiene el alumno, y de la planificación del aprendizaje y de la enseñanza de lenguas. (Consejo de Europa, 2002: 86)

2. LA MUESTRA

2.1 Selección de los informantes

Esta investigación se basa en un experimento realizado a partir de una prueba de medición de la competencia léxica, cuyos resultados consisten fundamentalmente en tres tipos de productos: (i) diccionarios de léxico disponible, elaborados a partir del conjunto

de palabras que el grupo encuestado actualiza; (ii) estudios cualitativos sobre la naturaleza del léxico aportado; (iii) análisis estadísticos realizados a partir de índices que miden la cantidad de unidades léxicas producidas, ya sea en el eje paradigmático, ya sea en el eje sintagmático.

Esta encuesta se ha suministrado a 150 alumnos⁷² de la *Università degli Studi di Padova* durante el año académico 2017/2018. Se trata de aprendices italianos adultos de ELE procedentes de los cursos de *Lingua Spagnola II* y *III* del grado en lenguas extranjeras⁷³. Hemos optado por trabajar con esta muestra, que tiene un nivel B1 y B2 de español (según las indicaciones que nos han proporcionado los profesores conforme al MCER), ya que nuestro doble propósito es indagar (i) la evolución del conocimiento del alumnado de un nivel a otro y (ii) comprobar si dentro del mismo nivel se cumple un efectivo desarrollo de la competencia tras la asistencia a un curso anual.

Hemos contactado con los docentes titulares de los cursos para que nos concediesen unas horas de clases para suministrar las encuestas. Seguidamente, las hemos proporcionado dos veces a los mismos estudiantes, una a comienzo y una al final del año académico para capturar una imagen transversal y longitudinal de la competencia léxica⁷⁴.

En específico, la primera administración se ha llevado a cabo entre los meses de octubre y noviembre de 2017, se han recopilado 100 encuestas que nos han permitido el análisis transversal de los resultados. En el mes de mayo de 2018, hemos suministrado la segunda prueba al grupo más avanzado, recogiendo otras 50 encuestas para el análisis longitudinal⁷⁵.

De esta forma hemos obtenido dos corpus:

⁷² El número de participantes respeta los criterios de representatividad y rentabilidad necesarios para que la muestra sea significativa (Mackey, 1971b; Sinclair, 2005; Torruela, 2017).

⁷³ Se trata de cursos de grado ofrecidos por el Departamento de literaturas, lenguas y lingüística (*DiSLL*) denominados *Lingue, letteratura e culture moderne*; *Mediazione linguistica e culturale*; *Lingue, letteratura e mediazione culturale*.

⁷⁴ De manera parecida, Laufer (1991) ha realizado un estudio sobre riqueza léxica que examina longitudinalmente 47 aprendices de inglés L2 en diferentes etapas de la carrera universitaria. La autora afirma que en intervalos de 14 y 28 semanas es posible averiguar un desarrollo del vocabulario, desde el conocimiento hasta el uso activo. Esto contesta positivamente a la pregunta que se pone Read (2000: 197) sobre si la medición estadística y el cotejo de dos textos escritos por los mismos aprendices en dos etapas temporales distintas permiten conseguir resultados fructíferos para investigar la competencia léxica.

⁷⁵ En el sistema universitario italiano los estudios universitarios se dividen en dos ciclos. Contamos con un primer período de tres años que lleva a la licenciatura de *Laurea Triennale*, y una segunda etapa más especializada, de dos años, que permite la obtención de otro título, la *Laurea Magistrale*, equivalente aproximadamente a un máster (cuarto y quinto año en España), que posibilita el acceso a los programas de posgrado. Hemos elegido este grupo ya que debería haber alcanzado un pleno nivel B2 de ELE para obtener la licenciatura de *Laurea Triennale*.

- (i) el *corpus general* que consta de los datos de la primera prueba realizada por 50 aprendices del II año (nivel B1) y 50 aprendices del III año (nivel B2, que formará el *corpus B2_a* en el análisis longitudinal);
- (ii) el *corpus B2_b* constituido por los datos procedentes de la segunda prueba efectuada por 50 estudiantes del III año.

2.2 Las variables sociolingüísticas

Como es habitual en los estudios de esta índole, hemos trabajado con algunas variables sociolingüísticas con el propósito de comprobar si influyen o no en el conocimiento del vocabulario. Hemos dividido el muestreo en función de tales factores solo tras empezar la fase de análisis de los datos ya que, inicialmente, hemos seguido un proceso aleatorio de selección de los informantes, en el que se ha tenido en cuenta solo la asistencia a los cursos académicos mencionados para la administración de los tests. Por esta razón, las pruebas han sido suministradas durante una clase cualquiera, sin informar previamente al alumnado para eludir el absentismo. De los muchos factores considerables solo aplicamos tres, pese a que teníamos la posibilidad de servirnos de otros: el sexo, el nivel de ELE, el número de lenguas extranjeras conocidas⁷⁶.

2.2.1 Variable “sexo”

El primer condicionante utilizado para la clasificación de los encuestados, presente en todos los proyectos de corte sociolingüístico, es el sexo. Queremos comprobar si en nuestro corpus se detecta la predominancia de un género frente al otro, como ocurre en otros estudios⁷⁷.

⁷⁶ No hemos considerado, por ejemplo, la lengua materna, una de las variables más utilizada en este tipo de investigación porque asumimos que los alumnos fuesen itálofonos de nacimiento o bien que, a pesar de no ser nativos, manejasen el italiano como LM al estudiar en un contexto universitario italiano.

Por otra parte, en el grupo de estudiantes solo había tres estudiantes extranjeros (dos rumanos y uno serbio), con lo cual este condicionante no habría sido significativo.

Al contrario, otros estudios sobre aprendices de ELE examinan grupos variados de informantes, donde las diferencias radican propio en la L1, por lo que la distinción entre uno u otro idioma es imprescindible: Carcedo (2000c) coteja hablantes de finés y de sueco; Samper Hernández (2002), Sánchez-Saus (2011) y Rubio Sánchez (2015) trabajan en contextos de inmersión con estudiantes de diferentes LM.

Tampoco hemos tenido en cuenta la edad de los alumnos (Berton, 2014) o el nivel sociocultural y el tipo de centro (Reyes Díaz, 2010) porque no eran significativos para nuestros objetivos.

⁷⁷ Muchas investigaciones destacan valores más altos en la población femenina que en la masculina,

Sexo	Corpus general	Informantes
	Mujer	87
	Hombre	13
	TOTAL	100
	Corpus B2_a	Informantes
	Mujer	41
	Hombre	9
	TOTAL	50
	Corpus B2_b	Informantes
	Mujer	45
	Hombre	5
	TOTAL	50

Tabla 1. Distribución de los informantes en función de la variable “sexo”.

2.2.2 Variable “nivel de español”

El nivel de ELE es el segundo factor que hemos tenido en cuenta, permite comprobar la hipótesis de que a mayor dominio lingüístico corresponde un mayor caudal de léxico aprendido y conocido. El profesorado nos ha proporcionado la correspondencia entre los niveles en los que subdivide el estudiantado conforme a las indicaciones del MCER. Por eso, hemos optado por investigar los cursos de II y III año en virtud de nuestro propósito de analizar el nivel intermedio (o *independiente*, según el propio Marco, que lo reparte en B1, *umbral*, y B2, *avanzado*).

Nivel de ELE	Corpus general	Informantes
	B1	50
	B2	50
	TOTAL	100
	Corpus B2_a	Informantes
	B2	50
	TOTAL	50

mientras que pocos autores (Pérez Serrano, 2009; González Fernández, 2013) han detectado la situación contraria.

	Corpus B2_b	Informantes
	B2	50
	TOTAL	50

Tabla 2. Distribución de los informantes en función de la variable “nivel de ELE”.

En el cuestionario sociológico de la prueba hemos preguntado a los participantes que valorasen su nivel de español en una escala que recogía los niveles inicial, intermedio y avanzado para sondear la percepción de sus conocimientos y averiguar si había correlación con los estándares establecidos por el profesorado.

Corpus general	Informantes
Inicial	15
Intermedio	74
Avanzado	11
TOTAL	100
Corpus B2_a	Informantes
Inicial	1
Intermedio	43
Avanzado	6
TOTAL	50
Corpus B2_b	Informantes
Inicial	1
Intermedio	39
Avanzado	10
TOTAL	50

Tabla 3. Percepción de los informantes de su nivel de ELE.

Los estudiantes se conforman con la opinión de los docentes y con las indicaciones curriculares de los cursos en los que están matriculados ya que los porcentajes más altos se detectan en el nivel intermedio.

Solo el 15% de los componentes del corpus general contesta “inicial”, mientras que el 20% de los del B2_b se siente más cercano a un nivel avanzado. Esto se ve reflejado en las respuestas a la pregunta 12 del cuestionario, sobre el grado de dificultad del aprendizaje del español: la mayoría de los informantes afirma que este idioma no presenta muchos problemas y tampoco demasiada sencillez, como muestra la escala de Likert a continuación.

		1	2	3	4	5	
Corpus general	difícil	3%	17%	57%	21%	2%	fácil
Corpus B2_a		2%	26%	52%	18%	2%	
Corpus B2_b		0%	14%	68%	18%	0%	

Tabla 4. Opinión de los informantes sobre la dificultad de aprender español.

2.2.3 Variable “conocimiento de otras lenguas extranjeras”

Por último, analizamos la influencia del conocimiento de otras lenguas extranjeras, además del español, en los procesos adquisitivos, con la finalidad de comprobar la validez de la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979), la cual plantea que un individuo que ya ha conocido una LE emplea sus competencias lingüísticas durante el aprendizaje de un nuevo idioma como elemento facilitador.

Otras LE conocidas	Corpus general	Informantes
	Conoce dos LE	19
	Conoce más de dos LE	81
	TOTAL	100
	Corpus B2_a	Informantes
	Conoce dos LE	9
	Conoce más de dos LE	41
	TOTAL	50
	Corpus B2_b	Informantes
	Conoce dos LE	6
	Conoce más de dos LE	44
	TOTAL	50

Tabla 5. Distribución de los informantes en función de la variable “conocimiento de otras LE”.

2.3 Codificación de las variables

Hemos codificado los factores y sus variantes que acabamos de presentar para dividir los grupos y reconocer los informantes, como indica tabla 6, que proporciona la clave de identificación de cada encuestado.

Variable	Variante	Código
Sexo	Mujer	1
	Hombre	2
Nivel de ELE	Nivel B1	1
	Nivel B2	2
Conocimiento de otras LE	LE =2	1
	LE >2	2

Tabla 6. Codificación de las variantes de las variables.

3. LA ENCUESTA

La encuesta finalizada a la medición de la competencia léxica (Anexo I) que hemos diseñado sigue los modelos de las investigaciones previamente citadas: empieza con unas disposiciones introductorias, que presentan los objetivos y las motivaciones del proyecto, y cuenta con tres apartados:

- (i) el cuestionario sociológico;
- (ii) la prueba de disponibilidad léxica;
- (iii) la prueba de riqueza léxica.

Este test ha sido construido conforme a criterios de fiabilidad, validez y practicabilidad ponderados en función de la muestra elegida.

Tratándose de aprendices de ELE universitarios, las instrucciones se han escrito en español, pero, con el fin de garantizar la plena comprensión y evitar malentendidos, las explicaciones orales se han dado en italiano.

En el momento de la administración siempre se ha seguido el mismo procedimiento, aplicando un protocolo estándar que garantizase la repetitividad del experimento:

1. presentación del investigador y del estudio;
2. explicación de los contenidos y de las instrucciones;
3. aclaración de la anonimidad⁷⁸;
4. administración de las hojas;
5. realización de la prueba.

⁷⁸ Para prevenir cualquier ansiedad, se ha hecho hincapié en que no se trataba de un examen y la prueba no habría influido en la nota de la asignatura. Solo en el apartado de riqueza léxica los estudiantes podían poner su nombre para que el texto fuese corregido y entregado si querían enterarse de los eventuales errores cometidos ya que a para aprobar el examen del curso se prevé una redacción escrita.

3.1 El cuestionario sociológico

El primer apartado, dos páginas, consta de dos partes cuyo propósito es la recogida de datos personales, informaciones sobre el bagaje lingüístico de los encuestados y su opinión sobre el aprendizaje de una LE.

Teniendo en cuenta la preocupación constante de los estudiantes frente a las pruebas, se ha insistido en que esto servía para informes estadísticos y que solo nos interesaba su opinión por lo que no había respuestas correctas o erradas (Minera Reyna, 2010).

Este cuestionario contenía más preguntas de las necesarias, sin embargo, no todo lo que se ha pedido ha sido codificado, habiendo fijado los factores sociolingüísticos *a posteriori*, debido a que no sabíamos quién habría participado realmente en la encuesta. Por tanto, como no habríamos sacado partido de ciertas informaciones personales, las hemos descartado.

No se ha establecido un límite de tiempo para rellenar esta parte, sino que nos hemos asegurado de que todos los participantes hubiesen tenido margen suficiente para completarla antes de pasar a la ejecución de las pruebas.

3.1.1 Opiniones de los informantes sobre el aprendizaje de una LE

El cuestionario contenía las siguientes preguntas finalizadas a recoger las opiniones de los informantes en torno al aprendizaje de una LE: (i) *¿Qué cree que es más importante conocer de una LE?* (ii) *¿Cuánto cree que es importante conocer el vocabulario de una LE?* (iii) *¿Cómo cree que aprende más vocabulario?*

Conocer las actitudes de los aprendices es fundamental para el profesorado de modo que pueda diseñar una propuesta didáctica adecuada (cfr. la enseñanza centrada en el alumno) y, en segundo lugar, permite comprobar si los estudiantes son conscientes de cómo funciona la ASL ya que, es importante concienciarlos para que su competencia se desarrolle con éxito.

Corpus general	Gramática	2%
	Vocabulario	7%
	Gramática y vocabulario	91%
	TOTAL	100%

Corpus B2_a	Gramática	2%
	Vocabulario	8%
	Gramática y vocabulario	90%
	TOTAL	100%
Corpus B2_b	Gramática	2%
	Vocabulario	8%
	Gramática y vocabulario	90%
	TOTAL	100%

Tabla 7. Pregunta 9: ¿Qué cree que es más importante conocer de una LE?

Corpus general	Mucho	82%
	Bastante	17%
	Regular	1%
	Poco	0%
	Nada	0%
	TOTAL	100%
Corpus B2_a	Mucho	82%
	Bastante	16%
	Regular	2%
	Poco	0%
	Nada	0%
	TOTAL	100%
Corpus B2_b	Mucho	84%
	Bastante	16%
	Regular	2%
	Poco	0%
	Nada	0%
	TOTAL	100%

Tabla 8. Pregunta 10: ¿Cuánto cree que es importante conocer el vocabulario de una LE?

Corpus general	Ejercicios de léxico	15%
	Lectura en español	40%

	Películas y canciones en español	39%
	Otros	6%
	TOTAL	100%
Corpus B2_a	Ejercicios de léxico	18%
	Lectura en español	56%
	Películas y canciones en español	56%
	Otros	8%
	TOTAL	100%
Corpus B2_b	Ejercicios de léxico	20%
	Lectura en español	46%
	Películas y canciones en español	68%
	Otros	8%
	TOTAL	100%

Tabla 9. Pregunta 11: ¿Cómo cree que aprende más vocabulario?

Las respuestas a la pregunta 9 destacan que los informantes saben que gramática y vocabulario son componentes fundamentales para el conocimiento de una LE (90%). Lógicamente, a la pregunta 10 contestan que aprender léxico es importante. Pero, la pregunta 11 evidencia que creen que se adquiere de forma inconsciente, mediante la lectura de textos, la escucha de canciones o la visión de películas. Les parece que las palabras se aprenden como resultado de otras actividades, no dirigidas específicamente al léxico, ya que no más del 20% indica la respuesta “ejercicios de léxico”. Quizás esto se deba al *modus operandi* al que están acostumbrados: las clases de español se centran más en la gramática, la contrastividad, el análisis del texto y la literatura, descuidando las actividades específicamente diseñadas para el conocimiento del léxico⁷⁹.

3.2 La prueba de disponibilidad léxica

La prueba de disponibilidad léxica coincide con los tests estandarizados

⁷⁹ Merece la pena subrayar que, si bien el Enfoque Léxico ponga de relieve la importancia del aprendizaje incidental del vocabulario, este no es suficiente para el pleno desarrollo de la competencia léxica: hay que insistir en la utilidad de programar tareas que se dirijan a una práctica explícita y a un aprendizaje consciente del alumno.

tradicionales, se compone de cuatro páginas formadas por cuatro columnas y un espacio sobre cada una, numerado de 1 a 16, en el que se escriben los nombres de los centros de interés. Estos se han presentado uno a uno a medida que se completaban para evitar que se pensase antes de tiempo en las palabras vinculadas con el tema y que las respuestas fuesen el resultado de una reflexión (Samper Hernández, 2002).

Los informantes han rellenado las fichas con las unidades léxicas que se les ocurrían en relación con cada estímulo empezando por la primera línea en alto: los ítems más disponibles son los que aparecen en las primeras posiciones de las listas.

Cada columna permite un número máximo de aportaciones, pero para trabajar listas abiertas, hemos explicado que se podía seguir en el reverso del folio si la ficha no era suficiente.

Durante la realización, hemos exhortado a escribir todas las unidades léxicas que les venían a la mente, aunque no estuviesen seguros de la grafía o del significado.

En lo que atañe al margen temporal, nos hemos conformado con los proyectos de esta naturaleza dejando dos minutos para cada CI sin permitir volver atrás una vez terminado el tiempo establecido: “dos minutos era un tiempo suficiente para escribir bastantes unidades e incluso para bucear en su memoria en busca de alguna más” (Samper Hernández, 2002: 21).

3.2.1 *Los centros de interés*

En el ámbito de la investigación en ELE, algunos de los dieciséis campos originales han sido objeto de crítica por la falta de interés y de significatividad para los participantes (Samper Padilla, Bellón y Samper Hernández, 2003) o por no abarcar los temas más comunes utilizados por un aprendiz no nativo, como las acciones de la vida cotidiana y las relaciones familiares (Izquierdo Gil, 2005)⁸⁰.

Se trata de elementos nucleares de la investigación en disponibilidad léxica, pero están sujetos a cierta arbitrariedad debida a los intereses de cada investigador que considera en su estudio los que le resultan necesarios para alcanzar sus objetivos y que

⁸⁰ Tómense de ejemplo las palabras de González Fernández (2014: 52):

[...] se pone de manifiesto como algunos centros de interés clásicos no son idóneos para los trabajos de disponibilidad léxica aplicada a la enseñanza del español como lengua extranjera. En las investigaciones de esta corriente se debería eliminar algunos centros de interés y modificar otros para poder obtener resultados más relevantes.

sean adecuados a sus informantes. Sin embargo, se ha destacado la utilidad de aplicar una metodología empírica común con el eje de facilitar una visión conjunta de las investigaciones, respondiendo a la urgencia de revisar los datos desde una perspectiva sinóptica (Callealta y Gallego Gallego, 2016; Jiménez Catalán, 2017; Paredes García, 2017), por lo que es importante poder cotejar distintos trabajos.

Por ello, en nuestro trabajo utilizamos los estímulos propuestos en el *Français Fondamental* y empleados, después, en el PPHDL, pese a que hayan sido cuestionados, reducidos, reformulados o sustituidos⁸¹. Hemos optado por ellos ya que nos basamos en los patrones del MCER y se ha demostrado su correspondencia con los ámbitos de contenido de las indicaciones europeas (Bombarelli, 2005; Bartol Hernández, 2010)⁸².

Asimismo, a raíz de los resultados de nuestros análisis y del cotejo con otros trabajos sobre distintos grupos de aprendices de ELE, intentamos dar respuesta a la cuestión crítica de la inadecuación de algunos de estos campos semánticos, establecidos como *input* para la elicitación hace más de setenta años.

Esta investigación se plantea como una ulterior reflexión que lleve a una revisión metodológica que permita la introducción de nuevos temas, adaptados a las necesidades y los intereses culturales y comunicativos de un aprendiz del siglo XXI.

Con todo eso, incluimos en la prueba los siguientes centros de interés:

- CI01, “partes del cuerpo” [CUE];
- CI02, “la ropa” [ROP];
- CI03, “partes de la casa (sin muebles)” [CAS];
- CI04, “los muebles de la casa” [MUE];
- CI05, “alimentos y bebidas” [ALI];
- CI06, “objetos colocados en la mesa para la comida” [MES];
- CI07, “la cocina y sus utensilios” [COC];

⁸¹ Samper Hernández (2002: 22), si bien utilice estos mismos centros, afirma:

Esto [la reducción o la reformulación de los campos semánticos] es absolutamente comprensible si tenemos en cuenta que lo que se persigue es la concreción del léxico usado normalmente por los hablantes de una determinada comunidad lingüística, y ese propósito puede conducir a una adaptación de los centros [...].

⁸² Han detectado cierta correlación entre la distribución de los CI tradicionales y los temas propuestos por el Consejo de Europa. De manera parecida, Paredes García (2015) y Ávila Muñoz (2016; 2017) cotejan el léxico disponible extraído de sus trabajos (llevados a cabo, respectivamente, en la Comunidad de Madrid y en la provincia de Málaga) sea con las indicaciones del MCER sea con las nociones específicas del PCIC para medir el grado de rentabilidad de las unidades léxicas propuestas por tales documentos.

- CI08, “la escuela: muebles y materiales” [ESC];
- CI09, “iluminación y calefacción” [ILU];
- CI10, “la ciudad” [CIU];
- CI11, “el campo” [CAM];
- CI12, “medios de transporte” [TRA];
- CI13, “trabajos del campo y del jardín” [TRC];
- CI14, “los animales” [ANI];
- CI15, “juegos y distracciones” [JUE];
- CI16, “profesiones y oficios” [PRO].

3.3 La prueba de riqueza léxica

El tercer apartado, una página, corresponde a la prueba de riqueza léxica en la que se requería que los informantes escribiesen un texto sobre un viaje que han hecho y les ha gustado. Hemos elegido este tema porque queremos indagar la capacidad real de uso del léxico: pretendemos maximizar la posibilidad de expresión de los aprendices de este nivel⁸³ y un argumento más complejo o especializado no habría aportado los datos que necesitamos.

El folio presenta una ficha con líneas, aunque era posible recurrir al reverso al necesitar de más espacio. Lo importante era que los relatos no excediesen las 120 palabras para permitir la recopilación de un corpus homogéneo y porque parece que después de las primeras 100 el cálculo de la riqueza léxica se desvirtúa, tal como afirma López Morales (2011: 24): “Necesitamos, pues, para tener metas claras, exámenes de textos escritos que se consideren ilustrativos, de los que se pueden tomar varias calas pero nunca superiores a las 100 palabras.” Era imprescindible, para el éxito de la prueba, que cada encuestado aportase una cantidad suficiente de unidades léxicas, dejando un cierto margen de error.

Hemos recopilado las 100 primeras palabras de cada redacción para obtener precisión y estabilidad en los cálculos, sensibles a la extensión textual. Asimismo, trabajamos con textos constituidos por un número igual de *tokens* ya que no nos interesa comprobar la variación de los índices según el aumento del tamaño de la muestra.

⁸³ Laufer y Nation (1995: 308) destacan la importancia de medir la competencia de un sujeto según la familiaridad desarrollada sobre un cierto tema.

De todo el material originado durante esta tarea, hemos desechado las narraciones que no cumplieran con los requisitos preestablecidos.

Los encuestados disponían de 30 minutos para redactar su texto, un intervalo de tiempo suficiente considerados el nivel de ELE y la temática sencilla.

4. ANÁLISIS DE LA DISPONIBILIDAD LÉXICA

Para este análisis nos basamos en los patrones del PPHDL y de otros trabajos dedicados a aprendices no nativos de español (Carcedo, 2000c; Samper Hernández, 2002; Sánchez-Saus, 2011; Hidalgo, 2017b). Seguimos estas directrices empíricas ya que, aunque hemos aportado algunas modificaciones, confiamos en la utilidad de cotejar nuestros resultados con los de otros proyectos sobre estudiantes extranjeros, según la hipótesis de que se pueden obtener datos provechosos para una descripción global de la competencia léxica de los aprendices de ELE a partir de la comparación entre estudios.

4.1 Los índices de disponibilidad léxica

Nuestra investigación se basa en dos tipos de estudios del corpus, en los que el análisis cuantitativo coincide con los primeros pasos del apartado analítico.

Los principales indicadores que utilizamos para determinar el léxico disponible de nuestros informantes son el total de palabras y el total de vocablos. Medimos también dos índices que informan sobre el grado de concreción semántica de un área léxica y su constitución interna, mediante la aplicación de las fórmulas matemáticas de

- (i) el índice de cohesión (Echeverría *et al.*, 1987) que estudia la coincidencia de las respuestas dentro del conjunto muestral por CI, esto es, si el uso del léxico es homogéneo:

$$\frac{\text{media de palabras por informante}}{\text{número total de vocablos}}$$

Los valores pueden oscilar entre 0 y 1: cuanto más se aproxima a 1 el índice, mayor cohesión se detecta porque las respuestas coinciden en gran medida y, por tanto, el centro es más compacto y cerrado. Al contrario, cuanto más se aproxima a 0 el índice, resulta una menor coincidencia en las aportaciones, una menor cohesión, con lo cual el centro es más difuso y abierto.

(ii) la densidad léxica, que contabiliza la media de repeticiones de las respuestas:

$$\frac{\text{número de palabras}}{\text{número total de vocablos}}$$

A partir de estos datos desarrollamos tres tipos de análisis: el primero es el análisis transversal que examina los resultados generales del corpus y los resultados por variable (sexo, nivel de ELE, conocimiento de otras LE). El segundo es el análisis longitudinal finalizado a corroborar una de las hipótesis: averiguar si se detecta un desarrollo de la competencia léxica entre el comienzo y el final del año académico basándonos en los corpus B2_a y B2_b. El tercer paso, el análisis comparativo, coteja nuestro corpus general con diferentes trabajos de la misma índole con el propósito de evaluar el bagaje léxico de distintos grupos de estudiantes de ELE de diferentes LM, entornos y contextos de aprendizaje.

Además, ofrecemos el análisis descriptivo estadístico de los datos, que permite examinar el comportamiento intragrupal e intergrupala de los participantes. Estudiamos la simetría o la dispersión mediante diagramas de caja y bigotes que muestran la varianza en el número de unidades aportadas⁸⁴.

En segundo lugar, el análisis cualitativo de los resultados arroja nueva luz en torno a las respuestas.

Examinamos el tipo de léxico disponible activado: averiguamos cuáles son las temáticas predominantes y las categorías gramaticales más difundidas en cada campo mediante la extracción de los vocablos que presentan un índice de disponibilidad (ID) igual o mayor a 0,1 en las listas del análisis cuantitativo. Fijamos esta medida de corte sin aportar solamente las primeras unidades de los listados (Carcedo, 2000c; Samper Hernández, 2002) ya que evita la interferencia de las unidades actualizadas

⁸⁴ Los diagramas se leen como sigue:

- (i) el límite inferior de la caja representa el primer cuartil (Q1), valor por debajo del cual se produce el 25 % de los casos;
- (ii) la línea interior expresa el segundo cuartil (Q2), o mediana, de modo que los valores que se encuentran por debajo informan sobre el 50% de los datos y los por encima, el 50 % restante;
- (iii) la “x” que se puede colocar por encima o por debajo de la mediana o coincidir con ella, indica la media matemática;
- (iv) el límite superior corresponde al tercer cuartil (Q3), por debajo de este valor se encuentra el 75 % de los casos.
- (v) los brazos, inferior y superior, indican los valores máximos y mínimos alcanzados y, al mismo tiempo, los límites a partir de los cuales se detectan casos atípicos (puntos) y extremos (asteriscos).

individualmente o por pocos sujetos⁸⁵.

Para llegar a esto, compilamos otras listas que contienen los vocablos con ID $\geq 0,1$ (que presentan una frecuencia acumulada de $\geq 30\%$ y de aparición $\leq 75\%$) a partir de las cuales hemos calculado la *cardinalidad del conjunto*, es decir el porcentaje de compatibilidad entre las distintas agrupaciones de los informantes según la propuesta de Ávila Muñoz y Sánchez Sáez (2010; 2011):

[...]a partir del número de vocablos más disponibles de cada muestra (denominado también cardinalidad), consideraremos la suma de las unidades comunes o convergentes en todos los subconjuntos (intersección), de las lexías que no están presentes en todos ellos –aunque sí en varios– (unión) y de las exclusivas de cada subconjunto (complemento) –la suma de la unión y de los complementos se denomina suma disyuntiva–. La compatibilidad será, pues, el peso de la intersección con respecto al total de vocablos (intersección + suma disyuntiva), es decir, el porcentaje de unidades léxicas compartidas. (Hidalgo, 2017b: 348)

Como colofón, cotejamos estos vocablos con el *Corpus de Referencia del Español Actual* (2008)⁸⁶, el corpus de referencia para el español, y el *Corpus de aprendices de español* (2018)⁸⁷, que recoge un gran caudal de textos producidos por aprendices de ELE

⁸⁵ Ruiz Basto (1987) es el primero que plantea que el vocabulario activo está formado por aquellas palabras que presentan un ID superior a 0,1 y actualizado por al menos un 20% de los informantes. Siguen Carcedo (2000c; 2001), Bartol Hernández (2001), Samper Padilla, Bellón Fernández y Samper Hernández (2003), Gómez Devís (2004), López González (2010), Fernández-Merino (2011), Šifrar Kalan (2014b), Hidalgo (2017b).

⁸⁶

[...] Se compone de una amplia variedad de textos escritos y orales, producidos en todos los países de habla hispana desde 1975 hasta 2004. Los textos escritos, seleccionados tanto de libros como de periódicos y revistas, abarcan más de cien materias distintas. La lengua hablada está representada por transcripciones de documentos sonoros, obtenidos, en su mayor parte, de la radio y la televisión.

[...] ha sido el punto de partida forzoso para investigaciones sobre el español actual, principalmente lingüísticas, pero también de campos tan dispares como el de la publicidad, la terminología o la sociología, así como para la elaboración de numerosos productos derivados: gramáticas, diccionarios, tesauros, correctores ortográficos, métodos de didáctica del español y aplicaciones de traducción automática, entre otros. (RAE, CREA)

⁸⁷

[...] es un conjunto de textos escritos producidos por estudiantes de español con diferentes grados de dominio lingüístico (niveles A1 a C1 del Marco común europeo de referencia, aplicado al español en el Plan curricular del Instituto Cervantes. Niveles de referencia para el español) y procedentes de seis L1: árabe, chino mandarín, francés, inglés, portugués y ruso. Se trata de una herramienta que permite a los profesionales del campo de ELE (profesores, investigadores, evaluadores, autores de materiales didácticos, responsables y equipos de centros e instituciones lingüísticas, etc.) llevar a cabo investigaciones aplicadas sobre la base de datos sólidos y objetivos, ya que puede proporcionar información sobre dificultades de aprendizaje, errores más comunes, vocabulario más o menos empleado, etc. que se podrá aplicar con facilidad en las aulas o integrar en los textos. [...]

[...] Ha sido construido mediante la recogida de muestras en distintos centros del Instituto Cervantes y universidades de un gran número de países, en un período que va desde octubre de 2011 hasta septiembre de 2013. [...] (Instituto Cervantes, CAES)

de diferentes LM y de distintos niveles de competencia (de los cuales hemos extraído los datos relativos al nivel intermedio). Nuestro propósito es averiguar si el léxico más disponible activado por nuestros informantes coincide con los vocablos de estos listados para comprobar si los encuestados conocen, y tienen a su disposición, aquellos lexemas que son de uso común en español y que se encuentran en las producciones de otros discentes del mismo nivel.

4.2 Tratamiento informático de los datos

La digitalización del material es una de las fases más importantes de la investigación. Nos servimos del programa *Dispogen II* (Echeverría *et al.*, 2005), pese a que en primera instancia hemos pensado en emplear *Dispolex* (Bartol Hernández *et al.*). Sin embargo, este, al haber sido ideado para estudios sobre hablantes nativos de español (PPHDL), presenta ciertas características que no satisfacen nuestras necesidades. *Dispogen* se basa en la fórmula de López Chávez y Strassburger Frías (1987) y permite calcular el total de palabras, el total de vocablos, el promedio por informante y el índice de cohesión. Además, extrae el ID de cada vocablo, junto a su frecuencia, porcentaje de aparición y frecuencia acumulada. Asimismo, examina y cruza los datos por variable sociolingüística, lo que resulta indudablemente provechoso para este proyecto. Es posible computar cinco variables a la vez. Todo ello permite cotejar las distintas informaciones con facilidad ya que trabajamos con tres.

Hemos registrado las respuestas de cada informante con un código que se divide en tres segmentos a fin de permitir el correcto funcionamiento del *software*, como sigue:

12200 001 01 pie, dedo, rodilla, brazo, pierna, cuello, barriga, mano, codos, espalda, cara, boca, nariz, cabeza, orejas, ojos
12200 002 01 mano, boca, brazo, barriga, hígado, dedo, uña, pierna, pie, pelo, mejilla, lengua, rodilla, ojos, estómago, corazón, cuello, cara
21200 003 01 mano, cabeza, pelo, ojos, boca, nariz, espalda, corazón, orejas, pie, rodilla, pierna, dedo, piel
21100 004 01 rodilla, pierna, cuerpo-humano, pie, pelo, ojos, mano, espalda, bigote
21200 005 01 cabeza, pie, ojos, lengua, mano
11200 006 01 barriga, ojos, nariz, pierna, cabeza, pelo, boca, mano, rodilla, dedo, pecho, cintura, garganta
11200 007 01 naso, nariz, boca, ojos, mano, espalda, pelo
11200 008 01 ojos, pierna, espalda, nariz, boca, brazo, cintura, mano, cabeza, estómago, frente
11200 009 01 cabeza, boca, pelo, rubio, pelirrojo, oír, ojos, marrón, azul, labios
21100 010 01 cabeza, mano, pie, rodilla, tobillo, ojos, cara, pelo, nariz, boca, espalda

Los cinco dígitos iniciales coinciden con las variables contempladas en el estudio⁸⁸; en este caso, los tres primeros expresan las variantes del género, del nivel de ELE y del conocimiento de otras lenguas; la serie siguiente indica el número asignado al informante; las dos cifras finales se refieren al centro de interés. Para poner un ejemplo, si tenemos un sujeto identificado con **12200 001 01**, se trata de las respuestas relativas al CI01, “las partes del cuerpo”, de una mujer de nivel B2 que conoce más de dos lenguas extranjeras.

4.3 La edición de los datos

El establecimiento de las “normas de despojo” no es una cuestión baladí, al contrario, se trata de una compleja búsqueda de soluciones que no siempre los investigadores comparten, pero es necesaria ya que de estas dependen los resultados de la encuesta. Antes del procesamiento informático es necesario editar las respuestas de los informantes para que el corpus resulte homogéneo, por eso, hemos procedido a una fase de edición y lematización del material mediante la aplicación de las medidas que ilustramos a continuación.

Las primeras fases de revisión y corrección han sido relativamente sencillas, mientras que la lematización ha implicado un mayor esfuerzo debido a las tomas de decisiones que conlleva.

Para solucionar ciertas dificultades de comprensión hemos tomado como referencia las siguientes obras de la RAE⁸⁹: el *Diccionario de la lengua española* (2017); el *Diccionario panhispánico de dudas* (2005); la *Nueva gramática de la lengua española* (2009); la *Ortografía de la lengua española* (2010); el *Corpus de referencia del español actual* (2008; 2015).

4.3.1 Criterios comunes de la investigación en disponibilidad léxica

Los pasos de edición que siguen se conforman con los patrones de Samper Padilla (1998) planteados inicialmente para el PPHDL que aquí hemos puesto en práctica “con

⁸⁸ Es necesario poner cinco dígitos porque el programa no funciona si se ponen menos cifras, por ello hemos introducido los requeridos, pese a que los últimos dos (00) no estén relacionados con ningún condicionante sociolingüístico.

⁸⁹ En cuanto posible, hemos consultado las versiones electrónicas, con lo cual no haremos referencia a los números de páginas.

cierta laxitud” debido a que trabajamos con informantes no nativos para enmendar lo que ha resultado erróneo sin quitarle naturaleza a las respuestas.

4.3.1.1 *Eliminación de los términos repetidos*

Cuando se ha repetido la misma palabra dentro de un centro de interés hemos descartado la escrita en la posición más baja de la lista.

A veces hemos encontrado la palabra exactamente igual, mientras que en otras ocasiones presentaba un diferente morfema flexivo, una variación ortográfica, una variante plena o acortada, pero en todo caso hemos recogido solo la primera.

4.3.1.2 *Corrección de la ortografía*

Hemos corregido los errores ortográficos. En general, se trata de errores gráficos, generalmente de poca relevancia. Incluimos en el corpus todas las entradas que hemos podido documentar.

Es oportuno señalar que antes de la administración de la prueba hemos explicado a los participantes que la ortografía no habría tenido importancia en el análisis:

Eso tiene una justificación bien clara, sobre todo cuando los encuestados son extranjeros: si comienzan a plantearse dudas de cómo se escribe correctamente una palabra determinada puede pasar bien que decidan no incluirla en el listado, bien que pierdan demasiado tiempo en esta tarea y no añadan otras que conocen. (Samper Hernández, 2002: 24)

Pretendíamos evitar el miedo al error y al bloqueo mental que suele activarse en los momentos de examen, cuando un sujeto a menudo prefiere callarse en vez de atreverse a decir algo (Dulay y Burt, 1977; Krashen, 1982; Torijano Pérez, 2004)⁹⁰.

⁹⁰ Resultaría interesante partir de esta tipología de error para realizar un estudio de la situación ortográfica de los estudiantes:

Este análisis pormenorizado de los errores ortográficos resultará de utilidad en el ámbito de la enseñanza de la lengua a la hora de la planificación y elaboración de materiales que ayuden a solventar los problemas de ortografía que presentan los usuarios y aprendices del español, tanto como lengua materna como lengua extranjera. [...] Para la enseñanza de la lengua resulta interesante contar con un corpus cacográfico que incluya las palabras disponibles con mayor dificultad así como los errores más frecuentes. Esto favorece una respuesta eficaz a las necesidades estudiantiles para mejorar su competencia ortográfica. [...] Aunque el objetivo de la disponibilidad léxica no persiga estudiar la disortografía, los datos aportados por las encuestas son de gran utilidad para la enseñanza de lenguas puesto que nos permiten conocer la tipología de errores, su frecuencia y su ámbito de aparición para tomar decisiones y avanzar en el terreno de la didáctica de la ortografía [...]. Asimismo, resulta muy interesante atender a los aspectos psicológicos que se activan al escribir para determinar la causa mayoritaria de error y tratar de establecer soluciones desde la psicodidáctica. (Trigo Ibáñez, Romero Oliva y Santos Díaz, 2018: 119, 127-128)

4.3.1.3 Unificación ortográfica

Es frecuente encontrar variaciones ortográficas de la misma palabra, como en el caso de *hierba* y *yerba*. Hemos registrado la forma más repetida entre los informantes, que en este caso es *hierba*.

4.3.1.4 Unificación de variantes meramente flexivas

Entre las respuestas aparecen palabras en el género y el número que acude más instintivamente a la mente de los informantes, por lo que, como regla general a fin de mantener la homogeneidad del corpus se reduce cada entrada a la forma no marcada del paradigma (el masculino singular de sustantivos y adjetivos, el infinitivo de verbos).

Sin embargo, es preciso señalar algunas excepciones con respecto a:

1. el número

- (i) los *pluralia tantum* (*tijeras*);
- (ii) las palabras compuestas cuyo segundo elemento es en plural (*rascacielos*, *videojuegos*);
- (iii) las palabras que se usan mayoritariamente en plural que designan una realidad múltiple (*vacaciones*, *orejas*), aunque tienen su entrada singular en el diccionario (*vacación*, *oreja*);
- (iv) los lexemas que implican una acepción diferente según el número (*padres*, *deberes*, *apuntes*).

2. el género

- (i) los significantes que se refieren a dos entidades distintas según el sexo (*ternero* –animal– *versus* *ternera* –alimento–);
- (ii) los heterónimos (*hombre/mujer*, *gallo/gallina*, *actor/actriz*).

4.3.1.5 Unificación de derivados regulares

Las variantes morfológicas se unifican bajo una sola forma, que no supone alteración de significado a la respuesta del encuestado. Aparecen en el corpus las unidades en su lexema primitivo, sin formas derivadas (*casita* > *casa*).

4.3.1.6 Unificación de formas plenas y acortamientos

Hemos unido las formas plenas, los acortamientos o los abreviamentos bajo una

sola entrada mediante el uso de paréntesis. Contamos, por ejemplo, con *bici* o *bicicleta* > *bici(cleta)*, *tele* o *televisión* > *tele(visión)*, *bus* o *autobús* > *(auto)bús*, que significa que entre las respuestas han aparecido ambas variantes.

4.3.1.7 Tratamiento de los sintagmas

Constatamos la presencia de algunas unidades multipalabra, tanto nominales como verbales, que deben ser tratadas como una unidad única según la teoría de los *chunks* de Lewis (1993), desarrollada por otros investigadores partidarios del Enfoque Léxico que abogan por la enseñanza de bloques léxicos (entre otros, Higuera, 2006).

Hemos registrado los sintagmas nominales con el uso de guiones (*ama-de-casa*), así como las construcciones idiomáticas del español introducidas por *ir de* (se han aceptado *ir de compras*, *ir de vacaciones*, mientras que *ir a la compra*, *ir a una fiesta*, *ir a la universidad* se han reducido a *compra*, *fiesta*, *universidad*).

Entre los sintagmas verbales actualizados hay algunos que no aparecen en los diccionarios al no ser locuciones fijadas en la norma, sino combinaciones donde “el verbo describe la acción directamente relacionada con el sustantivo que le sigue” (Sánchez-Saus, 2011: 233-234), en este caso hemos registrado los sustantivos de referencia (*ir al cine* > *cine*).

4.3.1.8 Tratamiento de las marcas comerciales

Hemos aceptado solo las marcas consideradas apelativos comunes por la mayoría de los hispanohablantes (*Coca-Cola*, *PlayStation*, *Facebook*), como propone Samper Hernández (2002), alejándonos de investigaciones más recientes que admiten una cantidad mayor (Sánchez Saus, 2011: 235; Hidalgo, 2017b: 212).

4.3.2 Criterios propios de la investigación en español como lengua extranjera

En segunda instancia, aplicamos criterios específicos para la edición de los datos aportados por informantes extranjeros (Samper Hernández, 2002; Sánchez-Saus, 2011).

4.3.2.1 Tratamiento de los errores

Se aceptan todas las entradas que se entienden con facilidad, salvo que se trate de palabras españolas. Parece este un criterio de poco rigor científico, pero:

cualquier otro criterio más restrictivo, [...] elimina gran parte de las palabras escritas por los informantes de niveles más bajos, que son los que presentan un mayor número de errores – los cuales, aun así, han registrado significativamente muchas menos palabras que los informantes de niveles medio y alto–. (Sánchez-Saus, 2011: 236)

4.3.2.2 *Tratamiento de los préstamos*

Hemos solucionado la presencia de préstamos de otras lenguas como sigue:

- (i) aceptación de los “neologismos universalizados” presentes en el DRAE, tanto en la grafía adaptada fonológica y ortográficamente al español (ej. *hámster*) cuanto en la grafía original (ej. *pasta*);
- (ii) inclusión de los extranjerismos utilizados con frecuencia por los hispanohablantes, aunque no se encuentren en el diccionario, por no haber otra forma de referirse a tales objetos o realidades;
- (iii) exclusión de palabras de la LM o de otras LE.

4.3.2.3 *Tratamiento de las interferencias de otras lenguas*

En relación con el apartado previo, aparecen en las respuestas extranjerismos adaptados semántica o estructuralmente al español, es decir palabras a las que los informantes han añadido una desinencia que concebían como hispana: las hemos eliminado por ser el resultado de interferencias lingüísticas⁹¹.

4.3.2.4 *Unificación en la presentación de las unidades*

Algún encuestado ha acompañado sus respuestas con un artículo, un adjetivo o un adverbio: se descartan todos, a excepción de los que forman unidades multipalabra.

4.3.2.5 *Amplitud de las relaciones asociativas*

Esta etapa es muy llamativa ya que se trata de un dictamen que lleva a resultados distintos, el investigador determina cuáles respuestas admitir y cuales no. Es un paso común de todos los estudios sobre disponibilidad léxica, pero al trabajar con aprendices no nativos, procedentes de una cultura diferente y con un limitado conocimiento del español, cobra una dimensión particular (Samper Hernández, 2002: 29).

⁹¹ Aunque el análisis de esas palabras sería una muestra interesante de los procesos de creación léxica durante la ASL (Samper Hernández, 2002: 29), aquí no pretendemos analizar tales mecanismos, por lo que no las hemos tenido en cuenta.

Por eso, aceptamos algunas asociaciones secundarias (sobre todo en los CI considerados más complejos para un estudiante extranjero), o sea palabras un poco fuera de contexto, y eliminamos solo las asociaciones inexplicables en relación con el estímulo temático.

4.3.3 *Criterios particulares por centros de interés*

Reseñamos ahora las decisiones tomadas durante la edición de cada uno de los centros de interés, con algunos ejemplos representativos.

4.3.3.1 *CI01, “partes del cuerpo”*

Estamos ante un centro muy productivo y poco problemático: hemos corregido solo algunos errores de acentuación o de elección falsa de consonante (**oido*, **musculo*, **naríz*, **tovillo*). La decisión más importante que hemos tomado afecta a la gran alternancia de aportaciones en singular y plural del mismo vocablo. Registramos las formas plurales más frecuentes como *ojos*, *codos*, *nalgas*, *uñas*, *labios*, *dientes*, *muelas*.

Se han actualizado vocablos relativos a la anatomía del cuerpo humano, al aspecto físico (*pelirrojo*, *rubio*, *marrón*, *azul*), verbos (*oír*) y sentidos (*oído*).

Hemos eliminado todos los extranjerismos (*mustache*), las palabras inventadas (**cavil*, **piedes*) o adaptadas al español (**orillas*, posiblemente del francés *oreille*, **gamba*, **unghia*, del italiano).

4.3.3.2 *CI02, “la ropa”*

Este CI presenta un alto número de respuestas y escasos problemas. Por eso, la corrección se ha concentrado en los pocos casos de errores ortográficos (**cinturon*, **calzetines*, **camiceta*, **jersei*, **falta*).

Se han incluido en las listas, además de la prenda propiamente dicha, también los accesorios (*anillo*, *collar*, *pendientes*, *pajarita*, *gemelos*), algunos verbos (*ponerse*, *quitarse*, *vestirse*, *estirar*) y adjetivos (*estampado*, *colorado*, *elegante*, *deportivo*).

Sin embargo, hemos descartado los extranjerismos *boxer*, *t-shirt*, *slip*, *collant*, *sweater* (habríamos mantenido el vocablo españolizado *suéter*) a excepción de *bikini* ya que la forma es admitida en español, pese a la posibilidad de utilizar *biquini*; y algunas formas inventadas (**jaqueta*, **escarfa*, posiblemente del inglés *scarf*).

En el caso de palabras escritas en singular y en plural a la vez, como antes, hemos conservado las más frecuente: *zapatos, pantalones, vaqueros, tacones, gafas, calcetines, chanclas*.

Destaca la presencia de algunas colocaciones léxicas que cuentan con varias ocurrencias en el corpus: *traje de baño, gafas de sol, ropa interior, lente de contacto, zapatilla de deporte*.

4.3.3.3 CI03, “partes de la casa (sin muebles)”

En este centro hemos excluido muchas respuestas ya que detectamos un gran número de muebles o electrodomésticos que hemos preferido no admitir al ser objetos del campo siguiente. Se han eliminado los falsos amigos *tienda* y *taberna*, pero se ha mantenido *elevador* desde el momento en el que se utiliza en algunos países de América Latina (Cuba, El Salvador, Guatemala y México), el DRAE lo indica como sinónimo de *ascensor*.

Señalamos la presencia de unidades complejas a partir del término *cuarto*: *cuarto de baño, cuarto de estar, cuarto trasero, cuarto matrimonial*.

4.3.3.4 CI04, “los muebles de la casa”

Tampoco en este tema la edición ha sido complicada. Los encuestados han actualizado numerosos nombres de muebles, decoraciones y electrodomésticos (*ducha, fregadero*) que hemos recopilado sin encontrar graves errores de ortografía, además de las reiteradas faltas de acentuación (**estanteria, *lampara, *frigorifico*).

Encontramos otra palabra procedente de la variedad hispanoamericana del español, *hielera* (*nevera*).

Hemos unificado las entradas *mesa de noche, mesita de noche* y *mesilla de noche* bajo la forma *mesilla de noche*, la más utilizada.

Por primera vez, hemos utilizado los paréntesis para indicar la elisión de partes de una unidad: *(horno)microondas, tele(visión)*.

4.3.3.5 CI05, “alimentos y bebidas”

Se trata de uno de los campos semánticos más productivos, donde el porcentaje mayor de aportaciones es de sustantivos y en menor medida de verbos y adjetivos (*beber,*

asado, gaseoso).

Hemos mantenido los plurales de los vocablos que, aunque tienen su forma singular, han sido actualizado más frecuentemente en plural (*tapas, mejillones, churros, lentejas, cereales*).

Los errores más frecuentes son ortográficos y no perjudican el conocimiento del vocabulario: **lemón, *gaspacho, *jamon, *salsicha*.

Merece la pena subrayar la fuerte presencia de extranjerismos que hemos quitado de las listas (*riz, sushi, hamburger, cocktail, hot-dog, gin, spaghetti*), en particular porque en algunos casos tienen su adaptación al español (*hamburguesa, cóctel, perrito-caliente, ginebra, espagueti*). En lo que se refiere a los préstamos italianos *pizza* y *pasta*, hemos descartado el primero por aparecer en cursiva en el DRAE, lo que denota su naturaleza de barbarismo, mientras que hemos mantenido el segundo al estar totalmente incorporado en castellano. Asimismo, si bien la palabra *mojito* no sea incluida en el diccionario, la admitimos debido a la gran cantidad de ocurrencias en el corpus.

Constatamos ciertas palabras del registro coloquial, *bocata* y *montadito*, y de otras léxicas poco comunes en estudiantes de ELE como *aguacate, albahaca, apio, chontaduro, hierbabuena*, cuyo conocimiento resulta notable.

Contamos con la aparición de algunas locuciones, recopiladas con guiones: *agua con gas, agua sin gas, agua tónica, café solo, café con leche, jamón cocido, zumo de fruta, zumo de naranja*.

Finalmente, hemos encontrado y mantenido la marca comercial *Coca-Cola*.

4.3.3.6 CI06, “objetos colocados en la mesa para la comida”

Este centro resulta de escasa productividad de palabras, por lo que hemos decidido ampliar las posibilidades asociativas, aceptando acciones secundarias relacionadas con este tema (*aceite, sal, pimienta, vinagre, desayuno, cambiar, poner, tomar, pulir*).

4.3.3.7 CI07, “la cocina y sus utensilios”

Han sido actualizados nombres de utensilios y electrodomésticos que se pueden utilizar en masculino o en femenino, nos hemos inclinado por registrar la forma masculina al ser mayoritaria en los encuestados (*batidor, refrigerador, tostador*) a excepción de la femenina *licuadora*. Hemos excluido de las listas las palabras de otras lenguas (*freezer*,

moka, pentola, mixer) y las inventadas (**cucino*).

Hemos registrado con guiones las combinaciones *olla a presión* y *pala de freír*.

4.3.3.8 CI08, “*la escuela: muebles y materiales*”

La edición de este campo no ha supuesto grandes problemas. Hemos mantenido todas las respuestas, tanto las palabras referidas a materiales y muebles, como los significantes relacionados con personas y acciones perteneciente al ámbito escolar (*estudiante, secretaria, aprender, revisar, repetir*).

De nuevo, hemos recurrido a los paréntesis para indicar la elisión de partes de una unidad, *bolí(grafo)* y *profe(sor)*, aceptado las entradas que repetidamente aparecían en plural (*tijeras, apuntes, deberes, sacapuntas*) y registrado en la forma masculina los utensilios que figuraban sea en masculino que en femenino (*computador, calculador*).

Los extranjerismos y las interferencias han sido descartados (*post-it, *cantina* y **lavaña*, posiblemente del francés *cantine* y del italiano *lavagna*).

Contamos con otro vocablo de la variante hispanoamericana, *marcador* (*rotulador*), que hemos aceptado.

Los encuestados han arrojado algunas colocaciones y locuciones léxicas, que hemos registrado con guiones: *tomar apuntes, prestar atención, goma de borrar, aula magna, tabla periódica*.

4.3.3.9 CI09, “*iluminación y calefacción*”

Este CI, siendo poco productivo, ha llevado a la activación de múltiples asociaciones secundarias, que hemos mantenido, entre las cuales se han detectado varias clases de palabras: sustantivos (*temperatura, calor, mechero, calentamiento*), adjetivos (*caliente, frío, iluminado, oscuro, luminoso*) y verbos (*apagar, encender, calentar*). Es más, hemos encontrado vocablos referentes la meteorología y al clima (*sol, insolación, verano, estrella, noche, hielo*).

Constatamos las abundantes ocurrencias de dos sintagmas cuyo núcleo nominal es *energía* (*energía eléctrica* y *energía renovable*) al igual que otras unidades complejas como *panel solar, hilo eléctrico, corriente eléctrica* y *aire acondicionado*, que hemos recopilado con guiones. Las formas *hacer calor* y *hacer frío* se han dejado tal cual, sin recurrir a la restricción del elemento nominal como se hace en otras investigaciones.

Aquí también hemos empleado los paréntesis para agrupar bajo un solo lema la palabra *termo(sifón)*, debido a la presencia de ambas formas.

Otros elementos cuya presencia conviene resaltar son *radiador, ventilador, calentador, acondicionador, calefactor, refrescador* posiblemente creados al saber que el uso del sufijo *-or* permite formar sustantivos en español.

4.3.3.10 CII0, “la ciudad”

Se trata de un centro de gran amplitud semántica, puesto que abarca varios subcampos como las partes de la ciudad (*ayuntamiento, estadio, puerto, universidad, calle, carretera, avenida*), los tipos de vivienda (*casa, piso, apartamento*), los servicios (*tienda, supermercado, policía, estanco, aparcamiento*), los monumentos (*estatua, fuente, monumento*), los medios de transporte (*tren, vehículo, coche, metro*), las acciones que se pueden cumplir en una ciudad (*vivir, estudiar, trabajar, explorar, pasear*).

Hemos mantenido las formas plurales de *correos* y *rascacielos*, al ser las más utilizadas por los encuestados, mientras que el vocablo *metrópolis* se presentaba tanto en singular como en plural por lo que lo hemos registrado con paréntesis, *metrópoli(s)*.

De igual manera tratamos con paréntesis *disco(teca), auto(móvil), bici(cleta), (auto)bús*.

Se han suprimido las formas inventadas y los extranjerismos, pero se ha aceptado el anglicismo españolizado *esmog* (en cambio, el barbarismo *smog* ha sido descartado).

No ha sido registrada la multitud de topónimos, monumentos o calles, tanto si estaban escritos en español o en otras lenguas, por tratarse de nombres propios.

La aportación de locuciones ha sido copiosa, por lo que hemos optado por incluirlas sin quitarles ningún elemento nominal, verbal o adjetival: *casco antiguo, centro comercial, zona peatonal, centro histórico, vida nocturna, dar una vuelta, ir de tapas*.

4.3.3.11 CIII, “el campo”

Se trata de una de las áreas léxicas que revela un carácter asociativo muy disperso, por ello hemos mantenido casi todos los vocablos aportados.

Constatamos, de nuevo, el americanismo *chancho (cerdo)*.

En consonancia con los criterios ya expuestos, descartamos los extranjerismos *contadino, prato* y *arbre*.

Hemos observado la presencia de una combinación, *al aire libre*, y una palabra compuesta en plural, *espantapájaros*.

4.3.3.12 CII2, “medios de transporte”

El proceso de edición de este centro ha sido rápido y sencillo. Una de las pocas decisiones que hemos tomado se refiere a la inclusión de *patinetes* en plural debido a su frecuente aparición frente a la forma singular. Destacamos el frecuente uso de paréntesis que pone de manifiesto la aparición de ambas formas de la pieza léxica: *bici(cleta)*, *moto(cicleta)*, *(auto)bús*, *auto(móvil)*. Además de corregir los errores ortográficos (**pasejero*, **caravela*, **tandem*, **pié*), hemos procedido con la exclusión de los extranjerismos *skateboard*, *camper*, *roulotte*, *scooter* (tendría su versión adaptada, *escúter*) o *jeep* (sería *todoterreno*) y del nombre propio *Über*.

Nuevamente recopilamos los americanismos: *carro*, *colectivo*, *guagua*, *máquina*.

Se nota la presencia de las combinaciones: *a pie*, *transporte público*, *globo aerostático*.

4.3.3.13 CII3, “trabajos del campo y del jardín”

En este campo hemos admitido casi todas las respuestas.

Sobresale la gran cantidad de verbos (*cortar*, *plantar*, *regar*, *cultivar*, *arar*, *sembrar*, *cavar*) entre los cuales no hemos aceptado los falsos amigos procedentes del italiano **potar*, **potación*, **bonificar* o las formas inventadas **anafiar*, **seminar*.

4.3.3.14 CII4, “los animales”

Tras una simple revisión ortográfica, la edición de los datos no ha supuesto grandes problemas. Hemos eliminado los extranjerismos (*aquila*, *pesce*, *elephant*) y respetado los heterónimos: *cabrón/cabra*, *toro/vaca*, *gallo/gallina*, *caballo/yegua*.

Constatamos los siguientes americanismos: *tigrillo*, *hicotéa*, *chanchó*, *cochino*.

Han aparecido dos unidades compuestas, registradas con guiones: *caballito de mar* y *estrella marina*.

4.3.3.15 CII5, “juegos y distracciones”

Hemos eliminado numerosos extranjerismos (*baseball*, *football*, *volleyball*, *ping*

pong, ballet, rugby, computer, sport, social network, sortir), marcas comerciales (*Cluedo, Taboo*), invenciones y transferencias (**escondido, *mosca ciega, *luna parque*).

Hemos aceptado los vocablos adaptados al español *beisbol, voleibol* (en la variante sin acento siendo lo más frecuente en el corpus) y los nombres propios *Facebook, Play Station, Monopoly*.

Hemos insertado la forma plural de algunos vocablos siendo la más utilizada: *videojuegos, rompecabezas, cartas, damas, construcciones, dibujos animados*.

De nuevo hemos recurrido al uso de paréntesis al registrar los ítems *disco(teca), pelí(cula), tele(visión), bici(cleta)*.

Antes de terminar, conviene reseñar la edición de los sintagmas en vista de la gran variedad aportada y, aún más, porque muchos aparecen por duplicado, la locución por una parte y el sustantivo aislado por otra: *leer un libro y libro, tocar un instrumento e instrumento musical, jugar al fútbol y fútbol*. Hemos decidido registrar los sustantivos y los verbos como dos entradas distintas: *leer, libro, tocar, instrumento musical, jugar, fútbol*. No es el caso de *escuchar música* ya que hemos encontrado esta forma fija muchas veces, así que la hemos mantenido.

Hemos mantenido las locuciones donde el verbo implica un cambio de significado (*ir de tapas, ir de compras*) o precisa el significado del sustantivo (*salir con amigos*).

Constatamos la presencia de las combinaciones *dibujos animados, parque de diversiones, redes sociales, pasárselo bien, juego de mesa*.

4.3.3.16 CII6, “profesiones y trabajos”

La mayor complejidad que hemos tratado atañe a la cuestión del género: a menudo los informantes han escrito tanto la forma femenina como la masculina de ciertas profesiones. Hemos registrado la masculina, salvo en aquellos casos donde solo se haya actualizado la otra, como *ama de casa* o *señora de la limpieza*. Los heterónimos se han mantenido: *actor, actriz, azafata, preste, cura, monja*.

En lo que se refiere a la corrección de las respuestas, hemos detectado solo faltas ortográficas: **idrúlico, *hefe, *avocado*. Hemos suprimido los extranjerismos, los falsos amigos y las invenciones (*hostess, influencer, autista, *judice, *regista, *impiegado, *infermer*).

Hemos constatado, de nuevo, americanismos como *plomero, mucamo* y el término

coloquial *botón* y, por último, contamos con la aparición de locuciones como *cuidador de perro* y *guía de turismo*.

5. ANÁLISIS DE LA RIQUEZA LÉXICA

5.1 Los índices de riqueza léxica

La riqueza léxica es un método que permite estudiar las “palabras en contexto” midiendo el número de palabras utilizadas en una muestra textual, así como la cantidad de vocablos y su densidad (el empleo de unidades funcionales del discurso frente a las nocionales). Asumimos, pues, la riqueza léxica como un conjunto de aspectos: la variación léxica, la sofisticación léxica, la densidad léxica y el número de errores cometidos (Read, 2000). Por esta razón, aplicamos diferentes índices de medición para obtener datos lo más completos posible (Jarvis, 2013). De la misma manera, Capsada y Torruella (2017: 397) plantean la aplicación de un “eclecticismo como una forma de síntesis” al sugerir de

no escoger un único índice como el paladín de la riqueza léxica, sino, por el contrario, tratar de utilizar la información que nos pueden aportar todos estos cinco índices que han demostrado que presentan un buen comportamiento en sus medidas. De esta manera trabajaremos con más información y, además, el posible mal comportamiento de los índices en algunos textos, como ya hemos indicado que sucede, podrá quedar compensado.

Tras la elicitación, la edición y el tratamiento informático de los datos, el análisis se lleva a cabo en dos momentos distintos y complementarios.

Primero, el análisis cuantitativo establece el grado de riqueza léxica producido por los participantes en los relatos según los indicadores de la variedad y la densidad léxica. La selección de dichos índices no ha sido compleja ya que hemos trabajado con alumnos procedentes del mismo entorno educativo, con lo cual hemos podido aplicar sin dificultad las medidas que nos resultaban más útiles para la investigación⁹².

Hemos descartado los índices que implican errores por ser demasiado subjetivos y dependientes del juicio del investigador (Laufer y Nation, 1995: 310-313), como los que calculan la variación semántica o la calidad léxica: habría sido difícil encontrar

⁹² Laufer y Nation (1995) sugieren el empleo de medidas distintas según se trabaje con informantes pertenecientes a grupos heterogéneos, especialmente de un contexto educacional diferente, pero esto no ha sido nuestro caso.

palabras sofisticadas dado el nivel de ELE de los informantes. Igualmente, no hemos contemplado el índice de la originalidad léxica porque, pese a que permitiría comparar una redacción textual con respecto al conjunto textual, se basa en la cantidad de *tokens* de cada texto y no habría tenido sentido en nuestro análisis ya que los relatos tienen el mismo tamaño. A este propósito, tampoco aplicamos técnicas o medidas de normalización de las fórmulas como las de Zipf (1935) o Yule (1944).

Con todo eso y en base a la bibliografía revisada, hemos seleccionado los siguientes índices, medidos mediante las fórmulas a continuación:

(i) la variación léxica contabiliza el grado de diversidad del vocabulario, indica en tantos por ciento la capacidad de expresarse del autor de una redacción:

$$\frac{\text{número de } types \times 100}{\text{número total de } tokens}$$

(ii) la *Type/Token Ratio* (TTR) es otra forma de computar la variación léxica, describe la relación entre *types* y *tokens* contenidos en un texto:

$$\frac{\text{número de } types}{\text{número } tokens}$$

El resultado de este cálculo puede variar de 0 a 1. El valor máximo al que puede llegar es 1, en el caso en que una palabra nunca se repita, pero es algo imposible en un texto que no sea una oración. De igual manera, es improbable alcanzar el 0 ya que siempre se puede contar con un *type* como mínimo en una muestra (al contrario, no habría ningún texto). De ahí que el valor ideal de TTR resulte el que más se acerca a 1.

Este indicador se define también *Índice de proporción de vocablos*, es decir el porcentaje de vocablos calculado en relación con el total de palabras de cada texto:

[...] La fórmula para su obtención es $PorcV = V/N$. Con este índice se logra establecer qué porcentaje de las palabras del texto son diferentes, así como precisar en qué medida el estudiante repite o no las mismas palabras, con lo que mostrará su madurez léxica. A medida que aumenta el porcentaje, aumenta la riqueza léxica. (Cuba Vega, 2016: 84)

(iii) el índice de hápax detecta las palabras que solo se emplean una vez en un texto (los denominados *hápax*) y concurre a establecer la variación léxica:

$$\frac{\text{número total de } types}{\text{suma de palabras } hápax}$$

Cuanto más pequeño es el valor obtenido en un texto, más rico es su vocabulario ya que la riqueza léxica es inversamente proporcional a este índice: “conforme [...] va aumentando, la variedad léxica va disminuyendo.” (Reyes Díaz, 2010: 148).

(iv) la densidad léxica calcula el porcentaje de unidades nocionales de un texto en proporción al número total de palabras que lo componen:

$$\frac{\text{número de } \textit{tokens} \text{ léxicos} \times 100}{\text{número total de } \textit{tokens}}$$

(v) el intervalo de aparición de palabras nocionales (IAT), relacionado con la densidad léxica, determina el intervalo en el que aparecen las palabras de contenido semántico con respecto a las unidades funcionales:

$$\frac{\text{número total de } \textit{tokens}}{\text{número de } \textit{tokens} \text{ léxicos}}$$

Cuanto más se acerca a 1 su valor, mayor resulta el uso de palabras temáticas y, por consiguiente, la riqueza léxica (claro está que el 1 no es alcanzable porque significaría que todos los *tokens* de un texto fuesen palabras nocionales).

A partir de estos indicadores ejecutamos el estudio cuantitativo de los datos desde las perspectivas transversal (general y por variable), la longitudinal y el análisis comparativo con otros trabajos dedicados a la evaluación y medición de la producción escrita de aprendices de ELE de distintas procedencias.

Incluimos el estudio de los descriptivos estadísticos que examine el comportamiento intragrupal e intergrupar de los participantes en función de cada variante de las variables analizadas y de la fecha de suministración de la prueba y averigüe si estas ejercen algún efecto en el desarrollo de la competencia léxico-comunicativa de los informantes.

Es preciso tener en cuenta que, si bien estos datos sean numéricos, el análisis cuantitativo aporta también informaciones cualitativas sobre la competencia léxica:

La llamada ‘calidad de la escritura’ está integrada por una serie de factores, entre los que destacan sin duda: la riqueza léxica, la madurez sintáctica, los esquemas de cohesión y la coherencia discursiva. La amplitud y variedad del vocabulario está muy apoyado en la disponibilidad léxica del hablante, la madurez sintáctica, en su grado de entrenamiento combinatorial de oraciones simples en el discurso, los esquemas de cohesión y la coherencia

discusiva dependen esencialmente del ‘orden’ que quiera dársele conscientemente a los elementos constitutivos del discurso. (López Morales, 2011: 15)

A partir de ahí, para profundizar el estudio de los datos, como colofón, proponemos un acercamiento a un análisis de tipo cualitativo del corpus general organizado en dos fases.

En primer lugar, recopilamos las listas de los vocablos más frecuentes con el propósito de averiguar qué unidades se emplean de manera más consistente e indagar sobre la alternancia del uso de ítems gramaticales y léxicos en los textos.

En segundo lugar, buscamos cuáles son las palabras clave del corpus, que permiten definir y comprender el tema principal de un conjunto textual cotejando las unidades que lo componen con un corpus de control, formado por la “lengua general”.

Se entienden por palabra clave los lexemas típicos de un corpus porque aparecen allí con más frecuencia con respecto a lo que hacen en el discurso común. Para ello, se utilizan corpus de referencia, como –en este caso– el *Spanish Web corpus 2018 (esTenTen18)*⁹³ que representa el lenguaje estándar no especializado como puede ser, al contrario, el que se emplea en una muestra, cuya temática es específica o delimitada.

5.2 Tratamiento informático de los datos

La medición de la riqueza léxica requiere la edición y la digitalización del material antes de pasar a la fase analítica.

Hemos elegido *AntConc* (Anthony, 2018) como programa de ayuda para nuestro análisis cuantitativo. Este realiza una gran cantidad de cálculos, cuyas funciones principales permiten extraer datos de amplias muestras textuales.

Para obtener los indicadores que acabamos de reseñar, utilizamos los recursos que permiten la extracción de los índices de frecuencia, los rangos de las palabras, el conteo de *types* y *tokens*. Asimismo, con el fin de observar la densidad léxica, el uso de palabras funcionales frente a las nocionales, hemos utilizado la función *Stoplist* para que el programa filtrase una lista de elementos que queríamos eliminar de los cálculos, basándonos en las pautas de Hallebeek (1986).

⁹³ Hace parte del programa informático *Sketch Engine* (Kilgarriff *et al.*, 2004) y de la familia *TenTen corpus*, un conjunto de corpus de más de 10.000 millones de palabras creado a partir de textos recogidos en el Internet. En el caso del español, contiene 17.500 millones de entradas pertenecientes a las variedades europea y americana.

Los textos, en formato *.txt*, que constituyen el corpus de *AntConc* han mantenido el mismo código identificativo para cada informante que se ha aplicado para la disponibilidad léxica.

Durante el análisis cualitativo nos hemos servido de *Sketch Engine*, se trata de un recurso informático que combina estadísticas con criterios lingüísticos que permiten realizar ulteriores búsquedas.

Tras haber cargado nuestro corpus general, hemos utilizado las herramientas *Wordlist* para la recopilación de las listas de frecuencias (la lista general y la lista de las palabras nocionales, divididas según categoría morfológica) y *Keywords* para la extracción de las palabras clave.

La función *Word Sketch* ha permitido averiguar el uso de algunas palabras clave, mediante un sistema de interrogación de corpus que permite obtener de manera automática el comportamiento gramatical y colocacional de un lexema específico. El programa crea tablas e imágenes que ilustran con claridad las relaciones gramaticales, tales como: palabras que sirven como objeto del verbo, palabras que sirven como sujeto del verbo, adjetivos que modifican la palabra, expresiones pluriverbales, etc.

5.3 La edición de los datos

Antes de la informatización hemos dedicado una importante parte del proceso analítico a una minuciosa edición de los datos que prevé la normalización de las variantes flexivas y el tratamiento de los errores. Establecimos algunos criterios siguiendo las prácticas convencionales que se emplean en la bibliografía sobre el análisis de la producción escrita en lengua extranjera. Hemos tomado como referencia, otra vez, las obras mencionadas en el apartado de disponibilidad para solucionar cuestiones de comprensión léxica, gramatical, ortográfica y de uso.

5.3.1 *Tratamiento de los errores*

Según el marco teórico al que nos conformamos, hemos encontrado dos tipologías de error:

- (i) las faltas (ortográficas, fonológicas, morfosintácticas), que no perjudican el conocimiento del vocabulario español, que han sido toleradas y corregidas;
- (ii) los errores léxicos propiamente dichos, que se han descartado.

Hemos corregido los errores, en la medida de lo posible, dejando para futuras investigaciones el análisis de su incidencia en los valores de riqueza léxica entre el texto original y la versión corregida ya que se podría comprobar si la inclusión o exclusión en los cómputos contribuyen a mejorar los datos de manera sustancial o no implican cambios significativos.

5.3.2 *Corrección de la ortografía*

Considerado el nivel de los aprendices, se han ignorado los errores de ortografía (Wolfe-Quintero, Inagaki y Kim, 1998): hemos corregido la acentuación errata (**azúl*); la incorrecta duplicación de consonantes (**professor*); la selección falsa de consonantes (**pizo*) con la excepción de las unidades que, alejándose demasiado de la forma correcta, denotan un claro desconocimiento de la palabra.

5.3.3 *Unificación de variantes meramente flexivas*

Hemos lematizado todos los *tokens* bajo su forma no marcada. En todos los cómputos se han unificado las variantes flexivas, mediante la recopilación de un solo lema por cada palabra.

En lo que atañe a los verbos, los participios se han tratado como lemas independientes de manera que su uso con función de adjetivo no fuera incluido entre los predicados.

5.3.4 *Tratamiento de la morfología derivativa, de formas plenas y acortamientos*

Hemos dejado invariadas tanto las formas léxicas derivadas sustantivas y adjetivas –apreciativas, diminutivas, aumentativas, superlativas, despectivas– (*famosísimo, muchísimo, mercadillo, poquito, pueblecito*) como los acortamientos léxicos (*bici, moto*), puesto que contribuyen a cuantificar el caudal activo del vocabulario, además de testimoniar la capacidad de formación de las palabras.

5.3.5 *Tratamiento de los préstamos y de las interferencias de otras lenguas*

Hemos eliminado los extranjerismos y los préstamos de la LM y de otras LE (*band, B&B, camper, check-in, cous-cous, jeep, pullman, quad, souvenir, tour*); las interferencias, formas híbridas o inventadas, aunque basadas en vocablos existentes en

otras lenguas, al no constituir léxico útil para la investigación (*gita, *dipinto, *posto, *luego).

5.3.6 Tratamiento de los nombres propios

Los nombres propios se han descartado por no formar parte del vocabulario de una lengua, mientras que hemos incluido los topónimos en español ya que muchos de los que están presentes en el muestreo se suelen enseñar, al contrario, excluimos los escritos en italiano o en otro idioma (entre otros, *Firenze, Ginevra, New York*).

5.3.7 Tratamiento de palabras nocionales y palabras funcionales

En cuanto a la distinción entre palabras nocionales y funcionales, hemos recopilado una lista de parada (*Stoplist*) basada en la propuesta de Hallebeek (1986: 210-215) que distingue entre las categorías de: preposición; conjunción; verbo (seudo)copulativo; adverbio conjuntivo; determinante; pro-forma; verbo auxiliar; adverbio aspectual o modal; adverbio preposicional.

Sin embargo, en virtud del nivel de nuestros informantes, hemos elaborado una lista que concediese más libertad de respuestas y una mayor aceptación de palabras. La *Stoplist* que utilizamos, pues, está formada por los siguientes ítems: preposiciones; conjunciones; artículos; pronombres personales y posesivos; demostrativos; indefinidos; relativos; interrogativos y exclamativos; numerales; interjecciones (Anexo II).

De ahí que no se hayan descartado ni los adverbios ni los verbos auxiliares, copulativos y pseudocopulativos.

Por la misma razón, hemos eliminado de la lista de Hallebeek las unidades léxicas desuetas o no pertenecientes a la norma estándar (*aquende, enante, esotro, ogaño, otre, sece, seceno, vusco, vusted*, etc.), como muchos otros adjetivos que contribuyen a medir la competencia de los encuestados, según los criterios que acabamos de presentar.

CAPÍTULO IV.

ANÁLISIS CUANTITATIVO DE LA DISPONIBILIDAD LÉXICA

1. INTRODUCCIÓN

En la primera sección proponemos el análisis cuantitativo descriptivo de los datos procedentes de las primeras pruebas suministradas a los alumnos-informantes (cien sujetos) a comienzo del año académico.

Discutimos los resultados generales en función del número de palabras y vocablos actualizados, los promedios de respuestas, los índices de cohesión y de la densidad léxica. Seguidamente, nos centramos en la incidencia de las variables consideradas (sexo, nivel de ELE, conocimiento de otras LE) según el mismo proceso analítico, basado en el número total de *tokens* y *types* aportados, la media de palabras por informante y por centros de interés. Complementamos el apartado con el análisis de los descriptivos estadístico de cada condicionante a partir de los promedios de palabras por sujeto mediante un diagrama de cajas y bigotes que informa sobre la simetría o la dispersión del comportamiento intragrupal e intergrupala de los encuestados.

En la siguiente sección, introducimos una novedad con respecto a los otros estudios sobre disponibilidad léxica: el análisis longitudinal del muestreo, que llevamos a cabo gracias a la segunda aplicación de las encuestas a los participantes de nivel B2 (otras cincuenta pruebas recogidas) encaminadas a examinar la evolución de los conocimientos tras la asistencia a un curso académico de un año⁹⁴. Para ello, ofrecemos los mismos tipos de análisis, descriptivo y estadístico.

En la última sección, cotejamos nuestro corpus con los de otras investigaciones realizadas con aprendices de ELE de distinta procedencia y LM para averiguar si hay puntos comunes y cuáles son las diferencias en el desarrollo del bagaje léxico de los estudiantes.

⁹⁴ Hidalgo (2017b) realiza un estudio diacrónico sobre el desarrollo del léxico disponible de estudiantes sinohablantes, pero no se trata del mismo análisis que proponemos aquí ya que coteja los resultados de su grupo de informantes con los de otra investigación, realizada con otros sujetos (Lin, 2006).

2. ANÁLISIS TRANSVERSAL

2.1 Resultados generales

Empezamos por los resultados generales obtenidos al considerar el entero corpus: número total de palabras arrojadas, número total vocablos actualizados, promedios por informante, índice de cohesión y densidad léxica.

Partimos de los siguientes valores repartidos en los dieciséis centros de interés.

Informantes	Palabras	Vocablos	Promedio de palabras por informante	Promedio de vocablos por informante
100	12.579	1.490	125,79	14,90

Tabla 10. Índices generales de DL.

2.1.1 Número de palabras

La tabla 11 presenta los índices generales computados: el rango, o posición relativa, de cada campo según el total de palabras y el promedio por informante.

Rango	CI	Palabras	Promedio por informante
2	CI01.CUE	1.341	13,41
8	CI02.ROP	791	7,91
12	CI03.CAS	610	6,10
11	CI04.MUE	660	6,60
1	CI05.ALI	1.463	14,63
13	CI06.MES	530	5,30
14	CI07.COC	365	3,65
7	CI08.ESC	803	8,03
15	CI09.ILU	342	3,42
3	CI10.CIU	1.171	11,71
10	CI11.CAM	685	6,85
6	CI12.TRA	808	8,08
16	CI13.TRC	252	2,52
4	CI14.ANI	1.080	10,80
9	CI15.JUE	784	7,84

5	CI16.PRO	894	8,94
TOTAL		12.579	7,86

Tabla 11. Número total de palabras y media por informante.

El gráfico 1 ilustra la productividad por CI.

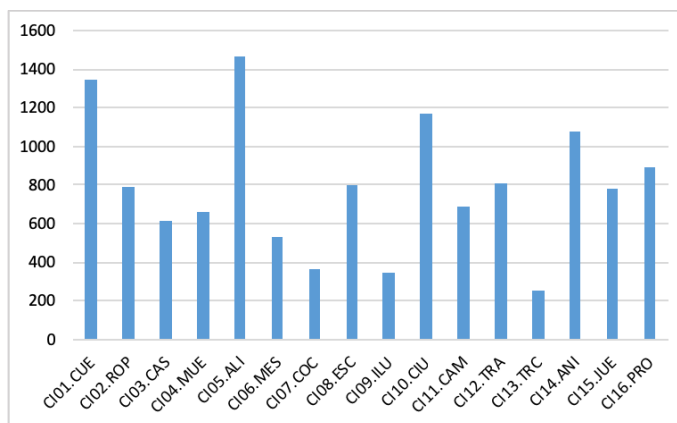


Gráfico 1. Número total de palabras.

Destacan cuatro franjas entre las cuales es posible dividir las áreas temáticas: contamos con un grupo que no alcanza las 400 unidades léxicas (CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”); un conjunto que llega hasta 800 ítems (CI02, “la ropa”; CI15, juegos y distracciones”; CI11, “el campo”; CI04, “los muebles de la casa”; CI03, “partes de la casa (sin muebles)”); CI06, “objetos colocados en la mesa para la comida”); otros que sobrepasan el promedio (CI16, “profesiones y oficios”; CI12, “medios de transporte”; CI08, “la escuela: muebles y materiales”); cuatro campos que superan los 1.000 lexemas (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI10, “la ciudad”; CI14, “los animales”). Estos, por tanto, presentan más *tokens* en absoluto.

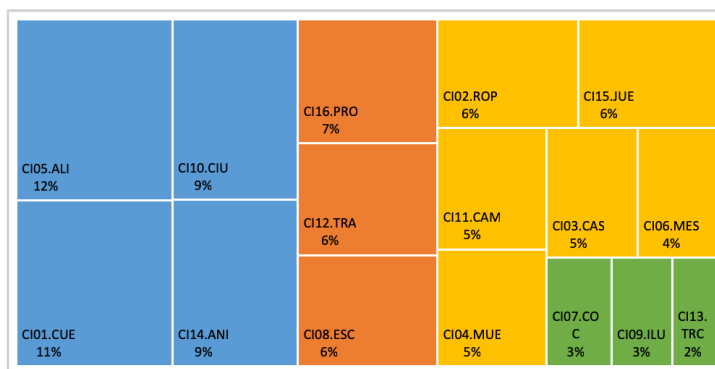


Gráfico 2. Porcentaje de palabras actualizadas por CI.

En lo que atañe al promedio de palabras por encuestado, la tabla 12 muestra los centros ordenados en función de su mayor o menor productividad (por rango). Si tomamos como referencia la media general, es decir 7,86 palabras por informante, vemos cuáles se sitúan por encima y por debajo de este valor.

Rango	CI	Palabras	Promedio por informante
1	CI05.ALI	1.463	14,63
2	CI01.CUE	1.341	13,41
3	CI10.CIU	1.171	11,71
4	CI14.ANI	1.080	10,80
5	CI16.PRO	894	8,94
6	CI12.TRA	808	8,08
7	CI08.ESC	803	8,03
8	CI02.ROP	791	7,91
9	CI15.JUE	784	7,84
10	CI11.CAM	685	6,85
11	CI04.MUE	660	6,60
12	CI03.CAS	610	6,10
13	CI06.MES	530	5,30
14	CI07.COC	365	3,65
15	CI09.ILU	342	3,42
16	CI13.TRC	252	2,52

Tabla 12. CI ordenados en función de su productividad.

Ocho de los dieciséis centros de interés rebasan las medias computadas, casi duplicándolas, mientras que los que se sitúan por debajo de este límite presentan un rendimiento muy inferior, en línea con lo observado por Samper Padilla, Bellón Fernández y Samper Hernández (2003) que insisten en la poca rentabilidad de ciertas temáticas lejanas de los intereses de los jóvenes y a menudo no trabajados ni en el aula ni en los materiales de ELE.

El centro más productivo es el CI05, “alimentos y bebidas”, que ofrece 14,63 palabras de media por informante, seguido por el CI01, “partes del cuerpo”; el CI10, “la ciudad”; el CI14, “los animales”, que alcanzan promedios que superan las diez unidades,

respectivamente. Al contrario, los campos menos rentables, que no consiguen superar las cuatro palabras por encuestado, son: CI07, “la cocina y sus utensilios”; CI09 “iluminación y calefacción; CI13, “trabajos del campo y del jardín”.

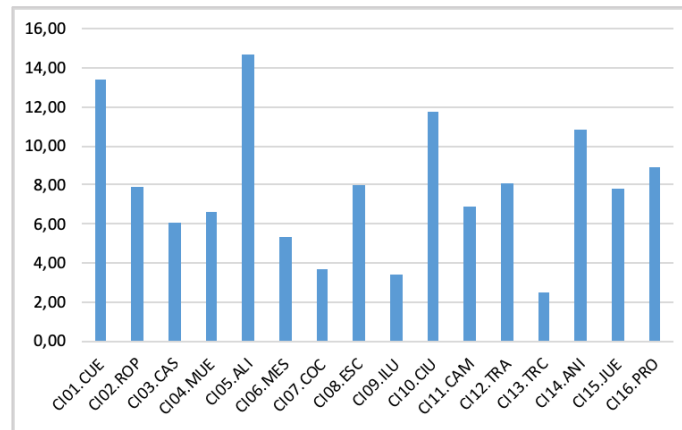


Gráfico 3. Promedio de palabras por informante.

2.1.2 Número de vocablos

Si en el apartado anterior analizamos la capacidad de producción de los informantes en los distintos campos semánticos, ahora hacemos lo mismo con la variedad léxica, esto es, el número de vocablos que han aportado, sin contar las repeticiones. Contestamos a la pregunta *¿existe una distinción entre el número de palabras y el número de vocablos aportados?* para averiguar qué tipo de relación se establece entre estos dos índices y, en consecuencia, si se repiten los mismos datos⁹⁵.

Contamos con 1.490 palabras diferentes en total que se reparten como sigue.

Rango	CI	Vocablos	Promedio por informante
7	CI01.CUE	91	0,91
8	CI02.ROP	72	0,72
15	CI03.CAS	47	0,47
14	CI04.MUE	49	0,49
1	CI05.ALI	175	1,75
16	CI06.MES	43	0,43

⁹⁵ Claro está que las cifras no son completamente reales puesto que habría que quitar los vocablos compartidos en varios centros de interés, pero *Dispogen* no permite realizar esta distinción, con lo cual trabajamos con todos los datos.

11	CI07.COC	66	0,66
9	CI08.ESC	72	0,72
13	CI09.ILU	57	0,57
3	CI10.CIU	157	1,57
4	CI11.CAM	141	1,41
12	CI12.TRA	58	0,58
10	CI13.TRC	67	0,67
6	CI14.ANI	102	1,02
5	CI15.JUE	131	1,31
2	CI16.PRO	162	1,62
TOTAL		1.490	0,93

Tabla 13. Número total de vocablos.

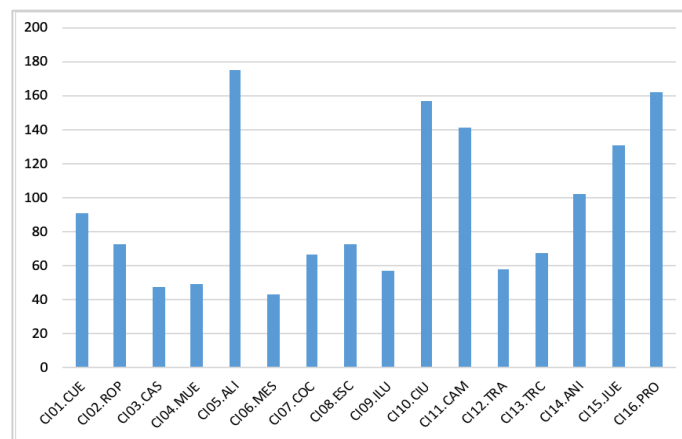


Gráfico 4. Número total de vocablos.

Seis campos presentan más variedad léxica arrojando más de 100 vocablos: CI05, “alimentos y bebidas”; CI16, “profesiones y oficios”, CI10, “la ciudad”; CI11, “el campo”; CI15, “juegos y distracciones”; CI14, “los animales”.

La reproducción gráfica desglosa la partición de los vocablos y muestra que es menos equilibrada con respecto a la del total de palabras que discutimos antes: podemos afirmar que parece no existir una relación entre productividad y riqueza léxica, los centros más productivos no son necesariamente lo más ricos.

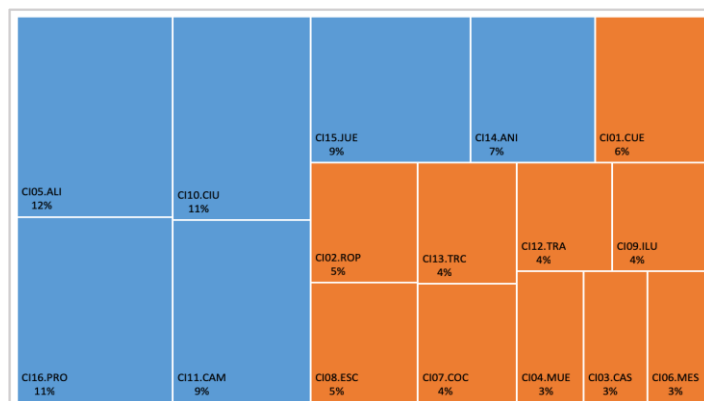


Gráfico 5. Porcentaje de vocablos actualizados por CI.

Presentamos los campos semánticos según el rango, señalando cuáles superan el valor promedio por informante, que es de 0,93.

Rango	CI	Vocablos	Promedio por informante
1	CI05.ALI	175	1,75
2	CI16.PRO	162	1,62
3	CI10.CIU	157	1,57
4	CI11.CAM	141	1,41
5	CI15.JUE	131	1,31
6	CI14.ANI	102	1,02
7	CI01.CUE	91	0,91
8	CI02.ROP	72	0,72
9	CI08.ESC	72	0,72
10	CI13.TRC	67	0,67
11	CI07.COC	66	0,66
12	CI12.TRA	58	0,58
13	CI09.ILU	57	0,57
14	CI04.MUE	49	0,49
15	CI03.CAS	47	0,47
16	CI06.MES	43	0,43

Tabla 14. CI ordenados en función de su variedad.

Destaca, de nuevo, la supremacía de seis campos, en los que notamos una diferencia de más de cien *types* con respecto al último, el CI06, “objetos colocados en la mesa para la comida”, que tiene una media de 0,43 vocablos.

2.1.3 Índice de cohesión y densidad léxica

El estudio de la relación entre palabras y vocablos permite examinar el grado de cohesión semántica al interior de los centros de interés y obtener ulterior información sobre la variedad de las respuestas⁹⁶.

CI	Índice de cohesión	Densidad léxica
CI01.CUE	0,147	14,74
CI02.ROP	0,110	10,99
CI03.CAS	0,130	12,98
CI04.MUE	0,135	13,47
CI05.ALI	0,084	8,36
CI06.MES	0,123	12,33
CI07.COC	0,055	5,53
CI08.ESC	0,112	11,15
CI09.ILU	0,060	6,00
CI10.CIU	0,075	7,46
CI11.CAM	0,049	4,86
CI12.TRA	0,139	13,93
CI13.TRC	0,038	3,76
CI14.ANI	0,106	10,59
CI15.JUE	0,060	5,98
CI16.PRO	0,055	5,52
PROMEDIO	0,092	9,23

Tabla 15. Índice de cohesión y densidad léxica por CI.

La tabla 16 muestra los índices de cohesión y la densidad léxica según el rango, que permite averiguar cuáles CI superan los promedios, que son 0,092 y 9,23 respectivamente.

⁹⁶ Idealmente la escala de los valores del índice de cohesión empieza en 0, el cual indica que el centro de interés es abierto y difuso ya que todos los ítems aportados difieren significativamente, y termina en 1, si el campo es cerrado y compacto porque las respuestas coinciden en gran medida. Del mismo modo, a mayor densidad léxica corresponde una mayor media de repeticiones de las lexías arrojadas, lo que implica una mayor homogeneidad del bagaje léxico.

Rango	CI	Índice de cohesión	Densidad léxica
1	CI01.CUE	0,147	14,74
2	CI12.TRA	0,139	13,93
3	CI04.MUE	0,135	13,47
4	CI03.CAS	0,13	12,98
5	CI06.MES	0,123	12,33
6	CI08.ESC	0,112	11,15
7	CI02.ROP	0,11	10,99
8	CI14.ANI	0,106	10,59
9	CI05.ALI	0,084	8,36
10	CI10.CIU	0,075	7,46
11	CI09.ILU	0,06	6,00
12	CI15.JUE	0,06	5,98
13	CI07.COC	0,055	5,53
14	CI16.PRO	0,055	5,52
15	CI11.CAM	0,049	4,86
16	CI13.TRC	0,038	3,76

Tabla 16. CI ordenados en función de su cohesión y densidad.

Ocho áreas superan el promedio y resultan las más cerradas y compactas, en particular el CI01, “partes del cuerpo”, alcanza la mayor concreción semántica del corpus y, a la vez, es uno de los temas más rentables en absoluto, lo cual destaca la relación que a menudo se establece entre productividad, riqueza y sus respectivas cohesión y densidad (0,147 y 14,74).

Al contrario, el CI11, “el campo”, y el CI13, “trabajos del campo y del jardín”, son los más abiertos (respectivamente 0,049 y 4,86; 0,038 y 3,76).

Según la clasificación de Gómez Molina y Gómez Devís (2004: 83)

- (i) doce CI son iguales o superan el 0,06, por lo que son semánticamente más compactos, con una mayor asociación conceptual;
- (ii) tres CI se acercan al 0,05 presentando un grado medio de coincidencia de las respuestas;
- (iii) un CI solo es el más difuso ya que tiene un índice menor de 0,04.

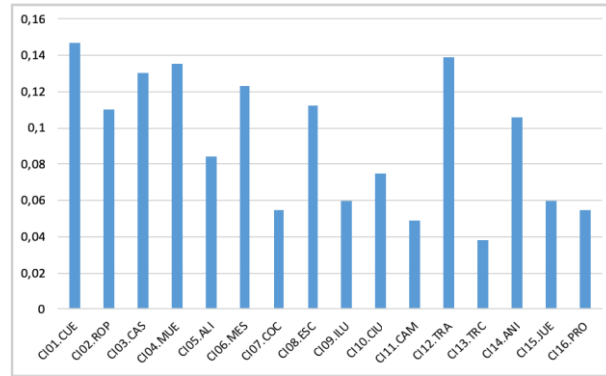


Gráfico 6. Índice de cohesión por CI.

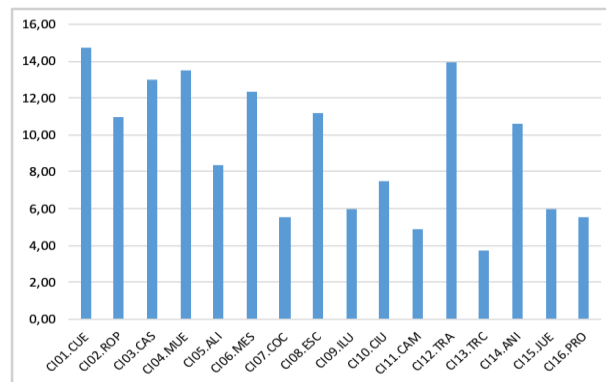


Gráfico 7. Densidad léxica por CI.

Antes de terminar, es interesante comparar estos índices con los demás valores analizados hasta ahora. La tabla muestra el modo en el que el índice de cohesión y la densidad léxica influyen en la relación entre productividad y riqueza léxica de los centros de interés, ya que ponen de relieve la coincidencia o la divergencia de las respuestas de los informantes.

CI	Palabras	Rango	Vocablos	Rango	Densidad léxica	Índice de cohesión	Rango
CI01.CUE	91	7	1.341	2	14,74	0,147	1
CI02.ROP	72	8	791	8	10,99	0,110	7
CI03.CAS	47	15	610	12	12,98	0,130	4
CI04.MUE	49	14	660	11	13,47	0,135	3
CI05.ALI	175	1	1.463	1	8,36	0,084	9
CI06.MES	43	16	530	13	12,33	0,123	5
CI07.COC	66	11	365	14	5,53	0,055	13
CI08.ESC	72	9	803	7	11,15	0,112	6
CI09.ILU	57	13	342	15	6,00	0,060	11

CI10.CIU	157	3	1.171	3	7,46	0,075	10
CI11.CAM	141	4	685	10	4,86	0,049	15
CI12.TRA	58	12	808	6	13,93	0,139	2
CI13.TRC	67	10	252	16	3,76	0,038	16
CI14.ANI	102	6	1.080	4	10,59	0,106	8
CI15.JUE	131	5	784	9	5,98	0,060	12
CI16.PRO	162	2	894	5	5,52	0,055	14

Tabla 17. Índices de DL (recapitulación).

El CI11, “el campo”, es uno de los más ricos y no resulta muy productivo (rango 10) al situarse en el rango 15 de cohesión, el cual indica que hay pocas repeticiones y, en consecuencia, es un campo abierto. El CI16, “profesiones y oficios”, pese a ofrecer una gran cantidad de palabras, es otro de los centros más variados (rango 2) que alcanza un bajo grado de cohesión. En el caso opuesto, el CI01, “partes del cuerpo”, es uno de los más productivos (rango 2) y compactos a la vez ya que su densidad presenta el valor superior de todo el corpus (rango 1), pero no arroja una gran cantidad de vocablos (rango 7). Lo mismo ocurre en el CI12, “los medios de transporte”, que pasa del rango 6 de productividad al 2 de riqueza: cohesión y densidad son elevados (rango 2).

Al mismo tiempo, es esencial tener en cuenta el hecho de que estos datos dependen de la cantidad de *tokens* y *types*. El CI05, “alimentos y bebidas”, y el CI10, “la ciudad”, son simultáneamente muy productivos y ricos: el total de vocablos no incide en los respectivos índices de cohesión y densidad. De manera parecida, el CI13, “trabajos del campo y del jardín”, se coloca en posiciones muy bajas en todos los índices analizados, parece que los informantes aportan pocas palabras que difieren notablemente y no hay una media alta de repeticiones, posiblemente porque no es un tema trabajado en clase. Además, algunos centros (CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI06, “objetos colocados en la mesa para la comida”) no entran ni entre los más productivos ni entre los más ricos, pero la escasez de vocablos permite un alto grado de densidad, quizás debido a un aprendizaje sistemático, que lleva a los estudiantes a escribir las mismas lexías:

Se corrobora así lo que han observado otras investigaciones, sea en ELE sea en ELM, esto es, que “[...] la mayor o menor riqueza de vocablos no se corresponde con la productividad de unidades léxicas en las diferentes áreas temáticas; es decir, que no se cumple la máxima: a mayor número de respuestas, mayor número de palabras diferentes.” (Gómez Molina y Gómez Devís 2004: 76-77)

2.1.4 Índice de disponibilidad, frecuencia de aparición y frecuencia acumulada

Antes de pasar a la discusión del análisis de los resultados por variable, presentamos los vocablos más disponibles en cada CI, datos que nos servirán para realizar el análisis cualitativo que ocupará el próximo capítulo. Examinamos exclusivamente los *types* compartidos por el conjunto de la muestra que alcanzan el valor de $ID \geq 0,1$ tomado como medida de corte. Dejamos fuera las unidades que, aunque hayan sido actualizadas, no se han revelado tan disponibles. Complementamos los datos con los porcentajes de la frecuencia de aparición $\geq 30\%$ y la frecuencia acumulada $\leq 75\%$.

CI	ID $\geq 0,1$	Frecuencia de aparición $\geq 30\%$	Frecuencia acumulada $\leq 75\%$
CI01.CUE	19	16	17
CI02.ROP	12	9	15
CI03.CAS	7	7	11
CI04.MUE	10	5	12
CI05.ALI	25	13	39
CI06.MES	6	6	6
CI07.COC	8	1	15
CI08.ESC	11	9	13
CI09.ILU	3	3	13
CI10.CIU	17	11	36
CI11.CAM	9	3	35
CI12.TRA	10	9	8
CI13.TRC	4	1	18
CI14.ANI	12	8	22
CI15.JUE	13	2	31
CI16.PRO	10	3	41
TOTAL	176	106	332

Tabla 18. Número de vocablos más disponibles.

El total de *types* con $ID \geq 0,1$ son 176, cuyo número por cada campo semántico oscila de 3 en el CI09, “iluminación y calefacción”, a 25 en el CI05, “alimentos y bebidas”. No siempre hay una relación proporcional entre la riqueza léxica y la cantidad de vocablos más disponibles ya que, por ejemplo, el CI09, “iluminación y calefacción”, tiene el menor número de vocablos, pero no es el más pobre. El CI01, “partes del cuerpo”,

aporta 19 *types* que igualan o superan el 0,1 (es el segundo en orden de rango) pero no es uno de los centros más rico (rango 7). Al contrario, destacamos que el CI16, “profesiones y oficios”, presenta solo 10 vocablos con $ID \geq 0,1$, pero se encuentra en la posición 2 conforme a su riqueza.

Al observar los datos relacionados con la frecuencia de aparición, la tendencia cambia: ahora el estímulo temático que llega a valores más altos es el CI01, “partes del cuerpo”, mientras que el CI07, “la cocina y sus utensilios”, y el CI13, “trabajos del campo y del jardín”, aportan las cantidades menores de palabras diferentes.

Los vocablos más frecuentes disminuyen del 66% con respecto a los más disponibles, pasando a ser 106 en total.

La frecuencia acumulada modifica, otra vez, la distribución de los resultados: el CI16, “profesiones y oficios”, ofrece el número más elevado de *types* y el CI06, “objetos colocados en la mesa para la comida”, presenta la cantidad menor. Por lo general, este parámetro registra un incremento de los vocablos en todos los centros, hasta llegar al +89% en relación con el total de los ID y triplicar el total de la frecuencia de aparición.

2.2 Conclusiones parciales

Nuestros informantes han actualizado un total de 12.579 palabras, de las cuales 1.490 son vocablos.

En lo que se refiere a la producción léxica detectamos cuatro franjas en las que repartimos los dieciséis centros de interés, en particular dentro de estos conjuntos notamos cuatro campos que superan las 1.000 unidades (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI10, “la ciudad”; CI14, “los animales”). Es más, ocho centros superan el promedio de palabras por sujeto mientras que la otra mitad logra valores muy inferiores, quizás debido al poco interés suscitado por las temáticas y la escasa, o nula, atención que les otorgan los programas curriculares y los materiales de ELE.

La variedad léxica, por otro lado, presenta una distribución distinta ya que predominan seis áreas que coinciden en parte con las que acabamos de exponer (CI05, “alimentos y bebidas”; CI16, “profesiones y oficios”; CI10, “la ciudad”; CI11, “el campo”; CI15, “juegos y distracciones”; CI14, “los animales”) y arrojan más de 100 vocablos. Sin embargo, esto permite afirmar que no existe una relación directa entre rentabilidad y riqueza ya que ser el centro más productivo no implica ser a la vez el más

rico, o viceversa, si excluimos el CI05, “alimentos y bebidas”, que alcanza los resultados mejores en casi todos los índices analizados.

Lo mismo ocurre al examinar el grado de cohesión semántica interna de cada centro porque hemos observado datos diferentes que cambian la tendencia de nuevo: ocho campos resultan los más cerrados, en particular el CI01, “partes del cuerpo”; al contrario, el CI11, “el campo”, y el CI13, “trabajos del campo y del jardín”, son los más abiertos.

Hemos propuesto el conteo relativo al índice de disponibilidad ($\geq 0,1$), la frecuencia de aparición ($\geq 30\%$) y la frecuencia acumulada ($\leq 75\%$) los cuales confirman que ser el centro más rico no equivale a tener vocablos más frecuentes y más disponibles.

2.3 Resultados por variable

En este apartado indagamos cómo varían los índices en función de las variables sociolingüística contempladas.

Como en el apartado anterior, nos centramos tanto en el número de palabras totales y palabras diferentes actualizadas, como en la media de unidades por informante, el índice de cohesión y la densidad léxica.

Cabe señalar que si bien esta última dependa del número de sujetos que componen un grupo hemos optado, en todo caso, por incluirla en nuestra comparación ya que estudiamos su influencia basándonos en los centros de interés ordenados en función del rango, abstrayéndonos de los valores en sí mismos.

2.3.1 Variable “sexo”

Contamos con una mayor presencia de mujeres y una menor participación de varones que componen un grupo más restringido⁹⁷ que presentan los siguientes índices de léxico disponible para el conjunto de los dieciséis centros de interés.

⁹⁷ La misma distribución se detecta en otros trabajos de esta índole:

En Carcedo (2000c) la distribución es de 295 mujeres y 55 hombres; en Samper Hernández (2002) es de 28 mujeres y 17 hombres; en Lin (2006), 200 mujeres y 63 hombres; en Pérez González (2006), 202 mujeres y 39 hombres; en López Rivero (2008), 28 mujeres y 15 hombres; en Pérez Serrano (2009), 31 mujeres y 12 hombres; en López González (2010), 302 mujeres y 61 hombres; en Sánchez-Saus Laserna (2011), 216 mujeres y 106 hombres; en González Fernández (2013), 46 mujeres y 32 hombres; en Sandu (2014), 204 mujeres y 76 hombres; o en Gallego Gallego (2014b), 123 hombres y 177 mujeres. (Hidalgo, 2017b: 259)

Informantes	Palabras	Vocablos	Promedio de palabras por informante	Promedio de vocablos por informante
Mujer (87)	11.070	1.389	127,24	15,97
Hombre (13)	1.509	617	116,08	47,46

Tabla 19. Índices generales de DL según la variable “sexo”.

Las informantes actualizan una cantidad mayor de palabras (+9,61%), pero los hombres son superiores en el número de vocablos, indudablemente más elevado (+66,35%).

Como tenemos en cuenta la gran diferencia en la constitución de las dos agrupaciones, trabajamos con el rango de palabras, vocablos y los promedios de respuestas para presentar un análisis más fiable.

De todos modos, las tablas 20 y 21 muestran los datos en bruto por cada variante, aunque no podamos utilizarlos en la fase de cotejo, y las medias computadas que sí, nos permiten verificar la influencia del factor género.

Mujer						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	1.167	2	13,41	83	7	0,95
CI02.ROP	699	8	8,03	67	9	0,77
CI03.CAS	532	12	6,11	46	14	0,53
CI04.MUE	590	11	6,78	46	15	0,53
CI05.ALI	1.273	1	14,63	162	1	1,86
CI06.MES	456	13	5,24	39	16	0,45
CI07.COC	317	14	3,64	65	10	0,75
CI08.ESC	722	6	8,30	71	8	0,82
CI09.ILU	307	15	3,53	56	12	0,64
CI10.CIU	1.056	3	12,14	152	2	1,75
CI11.CAM	600	10	6,90	123	4	1,41
CI12.TRA	700	7	8,05	53	13	0,61
CI13.TRC	229	16	2,63	65	11	0,75
CI14.ANI	954	4	10,97	92	6	1,06
CI15.JUE	685	9	7,87	120	5	1,38
CI16.PRO	783	5	9,00	149	3	1,71
TOTAL	11.07		7,95	1.389		1,00

Tabla 20. Índices de DL según la variante “mujer”.

Hombre						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	174	2	13,38	58	3	4,46
CI02.ROP	92	8	7,08	33	8	2,54
CI03.CAS	78	11	6,00	26	10	2,00
CI04.MUE	70	13	5,38	27	9	2,08
CI05.ALI	190	1	14,62	86	1	6,62
CI06.MES	74	12	5,69	25	11	1,92
CI07.COC	48	14	3,69	20	14	1,54
CI08.ESC	81	10	6,23	24	12	1,85
CI09.ILU	35	15	2,69	17	15	1,31
CI10.CIU	115	4	8,85	53	4	4,08
CI11.CAM	85	9	6,54	49	7	3,77
CI12.TRA	108	6	8,31	21	13	1,62
CI13.TRC	23	16	1,77	11	16	0,85
CI14.ANI	126	3	9,69	50	6	3,85
CI15.JUE	99	7	7,62	51	5	3,92
CI16.PRO	111	5	8,54	66	2	5,08
TOTAL	1.509		7,25	617		2,97

Tabla 21. Índices de DL según la variante “hombre”.

2.3.1.1 Análisis descriptivo

Si tomamos como referencia la productividad de los CI en la correlación de los datos, en ambos conjuntos coinciden las posiciones de siete campos (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI16, “profesiones y oficios”; CI02, “la ropa”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”), entre los cuales los más productivos son el CI05, “alimentos y bebidas”, y el CI01, “partes del cuerpo”, como ocurre en los resultados generales que analizamos antes. Por el contrario, los que presentan un menor número de actualizaciones son el CI07, “la cocina y sus utensilios”; el CI09, “iluminación y calefacción”; el CI13, “trabajos del campo y del jardín”.

El único campo que ocupa rangos claramente diferentes según el sexo de los encuestados es el CI08, “la escuela: muebles y materiales”, donde los hombres son menos productivos (rango 10 frente al 6).

Pasando a la variedad, el centro de interés más rico en ambos grupos es el CI05, “alimentación y bebidas”, siguen en las mismas posiciones el CI15, “juegos y distracciones” (rango 5); el CI14, “los animales” (rango 6); el CI12, “los medios de transporte” (rango 13). Las diferencias más evidentes atañen a los centros CI01, “partes del cuerpo”; CI08, “la escuela: muebles y materiales”; CI07, “la cocina y sus utensilios”; CI13, “trabajos del campo y del jardín”; CI03, “partes de la casa (sin muebles)”, que cambian su colocación de 4 a 6 puestos según cada opción de la variable.

Fijémonos en el gráfico que ilustra las divergencias en cuanto a la productividad.

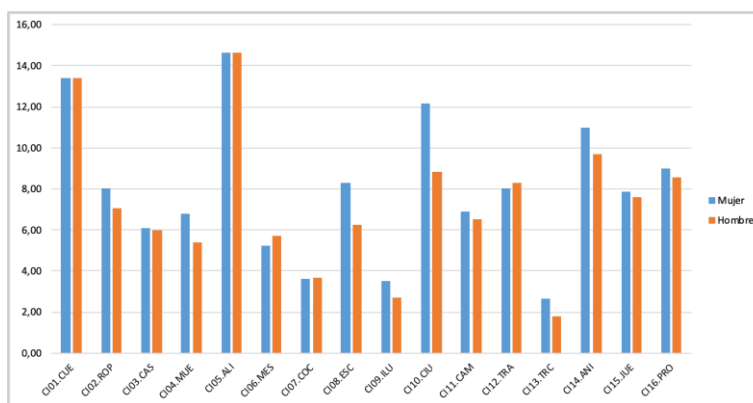


Gráfico 8. Promedio de palabras según la variable “sexo”.

No se notan valores muy distintos: en general, el desnivel es del 9,62%, pero si desglosamos los promedios, averiguamos dónde la desigualdad es más alta: los campos más homogéneos son el CI05, “alimentos y bebidas”, y el CI01, “partes del cuerpo”.

El rendimiento cambia a favor del conjunto femenino, entre otros, en CI10, “la ciudad” (+37,21%); CI08, “la escuela: muebles y materiales” (+33,19%); CI04, “los muebles de la casa” (+25,94%); CI14, “los animales” (+13,14%). En cambio, los hombres resultan ligeramente más productivos en CI07, “la cocina y sus utensilios” (+1,37%); CI12, “los medios de transporte” (+3,23%); CI06, “objetos colocados en la mesa para la comida” (+8,59%).

La tabla 22 contiene los descriptivos estadísticos que se representan gráficamente en el diagrama de cajas a continuación.

Descriptivos	Variante	
	Mujer	Hombre
Media	7,95	7,25

Mediana	7,95	6,81
Mínimo	2,63	1,77
Máximo	14,63	14,62

Tabla 22. Estadísticos descriptivos para la variable “sexo”.

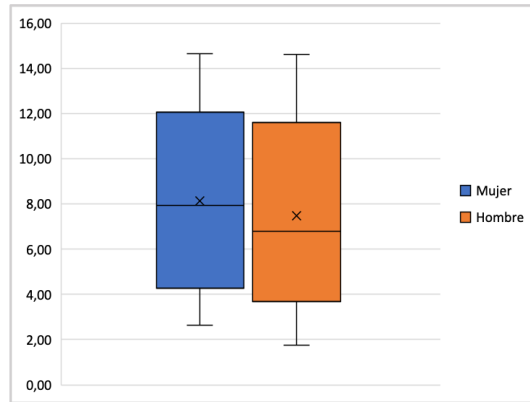


Gráfico 9. Diagrama de cajas para la variable “sexo”.

El gráfico refleja la superioridad de las mujeres y al mismo tiempo pone de relieve una desviación superior en las respuestas de los hombres: la mediana (6,81) se coloca por debajo de la media matemática (7,25) y el intervalo de datos entre el Q2 y el Q3 resulta más amplio (de ahí que la caja sea más alargada en la parte superior), con lo cual se supone una mayor variedad en el número de unidades aportadas.

Si analizamos los promedios de palabras diferentes se hace patente un cambio notable de la tendencia, sobresale una gran diferencia entre los dos grupos. Los hombres sobrepasan la media de las mujeres, casi triplicándolas, ofreciendo una riqueza mayor.

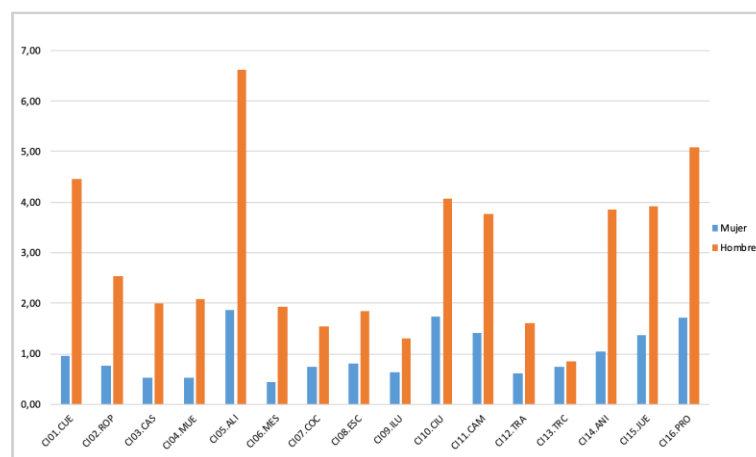


Gráfico 10. Promedio de vocablos según la variable “sexo”.

En efecto, los valores más altos correspondientes a los hombres son patentes en todas las áreas temáticas propuestas. Señalamos, sobre todo, el CI05, “alimentos y bebidas”; el CI01, “partes del cuerpo”; el CI16, “profesiones y oficios”, donde los promedios alcanzan valores más elevados. Mientras que las diferencias se reducen de manera notable en CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos de campo y del jardín”, ya que los cómputos revelan cifras más similares.

Por último, cotejamos los índices de cohesión.

CI	Índice de cohesión	
	Mujer	Hombre
CI01.CUE	0,162	0,231
CI02.ROP	0,120	0,214
CI03.CAS	0,133	0,231
CI04.MUE	0,147	0,199
CI05.ALI	0,090	0,170
CI06.MES	0,134	0,228
CI07.COC	0,056	0,185
CI08.ESC	0,117	0,260
CI09.ILU	0,063	0,158
CI10.CIU	0,080	0,167
CI11.CAM	0,056	0,133
CI12.TRA	0,152	0,396
CI13.TRC	0,040	0,161
CI14.ANI	0,119	0,194
CI15.JUE	0,066	0,149
CI16.PRO	0,060	0,129
PROMEDIO	0,100	0,200

Tabla 23. Índice de cohesión por CI según la variable “sexo”.

Dentro de las mujeres el CI01, “partes del cuerpo”, tiene un índice de cohesión y una densidad léxica entre los más altos, con lo cual la repetición de las mismas unidades es considerable, y, simultáneamente, es uno de los campos más productivos en absoluto. Consideremos ahora que, en el conjunto masculino, el CI16, “profesiones y oficios”, se manifiesta la situación opuesta, cohesión y densidad son bajos, por lo que no es uno de los estímulos más productivos, sino que aparece como uno de los más ricos (rango 2) ya

que el grado de concreción semántica es bajo. A la hora de correlacionar los datos, las respuestas de las mujeres resultan más heterogéneas, por ende, sus índices son superiores y el conjunto de lexías es más disperso con respecto a los hombres que ofrecen un muestreo de palabras más compacto, es decir con una mayor asociación conceptual.

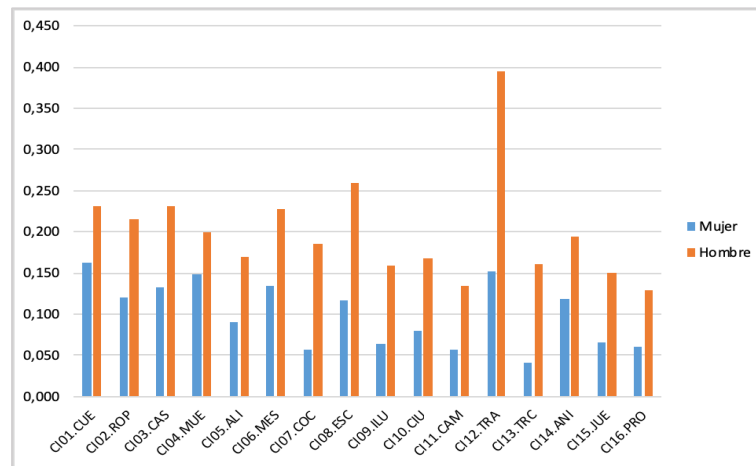


Gráfico 11. Índice de cohesión por CI según la variable "sexo".

Por último, abordamos el análisis de la densidad léxica.

CI	Densidad léxica			
	Mujer	Rango	Hombre	Rango
CI01. CUE	14,06	1	3,00	3
CI02. ROP	10,43	6	2,79	6
CI03. CAS	11,57	5	3,00	4
CI04. MUE	12,83	3	2,59	7
CI05. ALI	7,86	9	2,21	10
CI06. MES	11,69	4	2,96	5
CI07. COC	4,88	15	2,40	9
CI08. ESC	10,17	8	3,38	2
CI09. ILU	5,48	12	2,06	13
CI10. CIU	6,95	10	2,17	11
CI11. CAM	4,88	14	1,73	15
CI12. TRA	13,21	2	5,14	1
CI13. TRC	3,52	16	2,09	12
CI14. ANI	10,37	7	2,52	8
CI15. JUE	5,71	11	1,94	14
CI16. PRO	5,26	13	1,68	16
PROMEDIO	8,68		2,60	

Tabla 24. Densidad léxica por CI según la variable "sexo".

Advertimos ya en las premisas del presente apartado que a pesar de que este índice no sea directamente comparable entre los dos grupos ya que depende del número de los componentes, podemos examinar en cuáles campos las mujeres o los hombres tienen una media más alta de repeticiones mediante el análisis de los rangos, alejándonos de los valores en sí mismos.

Las chicas presentan una mayor densidad en el CI01, “las partes del cuerpo”, que es uno de los más productivos, seguido por el CI12, “los medios de transporte”, que se ubica en el rango 13 debido a la poca cantidad de *types* actualizada. Lo mismo ocurre en el muestreo de hombres, donde este centro alcanza su densidad máxima. Con respecto a la situación contraria, las mujeres presentan un valor bajo en CI13, “los trabajos del campo y del jardín”, posicionado en el rango 16 según el total de *tokens*, a continuación, hallamos con el CI11, “el campo”, en virtud de la gran aportación de vocablos que limitan la media de repeticiones. Los chicos, por su parte, escriben una gran cantidad de *types* en el CI16, “profesiones y oficios”, que aparece como el menos denso, seguido por el CI11, “el campo”, al igual que en el otro grupo.

2.3.1.2. Conclusiones parciales

Hemos visto que las mujeres arrojan más palabras (+9,61%) mientras que los hombres aportan una cantidad mayor de vocablos (+66,35%), pero comparando los promedios no destacan valores en realidad muy distintos: el desnivel entre las informantes femeninas y sus compañeros es del 9,62% y si desglosamos los datos por cada centro de interés observamos diferencias que no alcanzan el 40%. Al contrario, analizando las medias de vocablos, se pone de manifiesto un cambio de tendencia: hay una gran disparidad entre las variantes, los varones superan a las mujeres, casi triplicándolas, proporcionando una riqueza mayor, sobre todo en el CI05, “alimentos y bebidas”; el CI01, “partes del cuerpo”; el CI16 “profesiones y oficios”.

Al relacionar la producción de *tokens* por centro de interés, ambas muestras ofrecen más respuestas en el CI05, “alimentos y bebidas”, y el CI01, “partes del cuerpo”, y menos en el CI07, “la cocina y sus utensilios”; el CI09, “iluminación y calefacción”; el CI13, “trabajos del campo y del jardín”. En la producción de *types*, el CI05, “alimentación y bebidas”, sigue colocándose en el rango 1 en los dos grupos, mientras que las diferencias más evidentes atañen a varios centros que cambian su rango de 4 a 6 puestos

según variante.

El grado de cohesión semántica por cada área temática indica que las respuestas de las mujeres son más heterogéneas y dispersas con respecto a los hombres que aportan un conjunto de léxias más compacto y cerrado. La densidad léxica revela que el grupo femenino alcanza índices mayores.

2.3.2 Variable “nivel de ELE”

Los encuestados se dividen homogéneamente en dos grupos: 50 alumnos de nivel B1 y 50 alumnos de nivel B2⁹⁸.

Como era esperable, se nota un incremento del rendimiento del grupo B2, que ofrece cantidades de *tokens* y de *types* superiores con respecto a los aprendices del B1.

Informantes	Palabras	Vocablos	Promedio de palabras por informante	Promedio de vocablos por informante
B1 (50)	6.062	1.018	121,24	20,36
B2 (50)	6.517	1.161	130,34	23,22

Tabla 25. Índices generales de DL según la variable “nivel de ELE”.

Las cifras detalladas que los informantes han conseguido según el propio nivel de conocimiento lingüístico son las que siguen y nos llevarán a comprobar el peso del factor.

Nivel B1						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	594	3	11,88	58	7	1,16
CI02.ROP	364	9	7,28	54	9	1,08
CI03.CAS	300	12	6,00	35	13	0,70
CI04.MUE	316	11	6,32	38	10	0,76
CI05.ALI	669	1	13,38	121	1	2,42
CI06.MES	249	13	4,98	30	16	0,60
CI07.COC	161	15	3,22	35	14	0,70
CI08.ESC	392	8	7,84	57	8	1,14
CI09.ILU	173	14	3,46	36	12	0,72
CI10.CIU	596	2	11,92	118	2	2,36

⁹⁸ Por eso no es necesario normalizar los datos como en el caso del factor sexo y, como veremos, en el próximo apartado dedicado al conocimiento de otras lenguas extranjeras, donde el desnivel entre una variante y otra es alto.

CI11.CAM	349	10	6,98	92	4	1,84
CI12.TRA	396	7	7,92	38	11	0,76
CI13.TRC	107	16	2,14	35	15	0,70
CI14.ANI	518	4	10,36	71	6	1,42
CI15.JUE	417	6	8,34	90	5	1,80
CI16.PRO	461	5	9,22	110	3	2,20
TOTAL	6.062		7,58		1.01	1,27

Tabla 26. Índices de DL según la variante “nivel B1”.

Nivel B2						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	747	2	14,94	72	7	1,44
CI02.ROP	427	6	8,54	60	8	1,20
CI03.CAS	310	12	6,20	40	14	0,80
CI04.MUE	344	10	6,88	41	13	0,82
CI05.ALI	794	1	15,88	148	1	2,96
CI06.MES	281	13	5,62	33	16	0,66
CI07.COC	204	14	4,08	51	11	1,02
CI08.ESC	411	8	8,22	54	10	1,08
CI09.ILU	169	15	3,38	46	12	0,92
CI10.CIU	575	3	11,50	116	3	2,32
CI11.CAM	336	11	6,72	102	4	2,04
CI12.TRA	412	7	8,24	40	15	0,80
CI13.TRC	145	16	2,90	55	9	1,10
CI14.ANI	562	4	11,24	90	6	1,80
CI15.JUE	367	9	7,34	96	5	1,92
CI16.PRO	433	5	8,66	117	2	2,34
TOTAL	6.517		8,15		1.161	1,45

Tabla 27. Índices de DL según la variante “nivel B2”.

2.3.2.1 Análisis descriptivo

Los estudiantes que se encuentran en una etapa de estudio más avanzado aportan un caudal mayor de respuestas en todos los indicadores, en particular las medias por informante de palabras y vocablos, que traducidas en porcentajes han incrementado respectivamente del 7,51% y 14,05%, a pesar de que los alumnos de nivel inferior sean

más productivos en CI09, “iluminación y calefacción” (+2,37%); CI11, “el campo” (+3,87%); CI10, “la ciudad” (+3,65%); CI16, “profesiones y oficios” (+6,47%); CI15, “juegos y distracciones” (+13,62%) y un poco más variado en CI08, “la escuela: muebles y materiales” (+5,55%) y en CI10, “la ciudad” (+1,72%).

Si abarcamos el análisis de los rangos de productividad, siete centros se encuentran en las mismas posiciones, sin distinción de variante (CI03, “partes de la casa (sin muebles)”); CI05, “alimentos y bebidas”; CI08, “la escuela: muebles y materiales”; CI12, “medios de transporte”; CI13, “trabajos del campo y del jardín”; CI14, “los animales”; CI16, “profesiones y oficios”). Entre ellos el más productivo es el CI05, “alimentos y bebidas”, que se aleja del último en más de quinientas palabras. Este, el CI13, “trabajos del campo y del jardín”, está en el puesto más bajo y, por tanto, es menos rentable en ambos muestreos.

Fijándonos en el número de vocablos, los campos que coinciden según el rango son CI01, “partes del cuerpo”; CI05, “alimentos y bebidas”; CI06, “objetos colocados en la mesa para la comida”; CI09, “iluminación y calefacción”; CI11, “el campo”; CI14, “los animales”; CI15, “juegos y distracciones”. De nuevo, el CI05, “alimentos y bebidas”, ocupa la primera posición y resulta el más rico en ambos niveles.

Por el contrario, el CI06, “objetos colocados en la mesa para la comida”, es el menos variado.

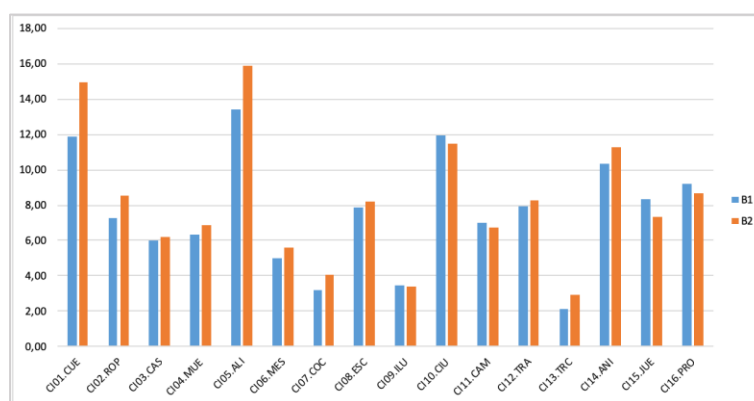


Gráfico 12. Promedio de palabras según la variable “nivel de ELE”.

Los promedios también favorecen los informantes de nivel B2: el desnivel oscila, como máximo, hasta el +25,76% en el CI01, “partes del cuerpo”, y la disparidad en general va de 7,58 en la variante B1 a 8,15 en la B2 (+7,52%).

La tabla muestra la distribución de los datos estadísticos y el gráfico ayuda a organizar la información.

Descriptivos	Variante	
	B1	B2
Media	7,58	8,15
Mediana	7,56	7,78
Mínimo	2,14	2,90
Máximo	13,38	15,88

Tabla 28. Estadísticos descriptivos para la variable “nivel de ELE”.

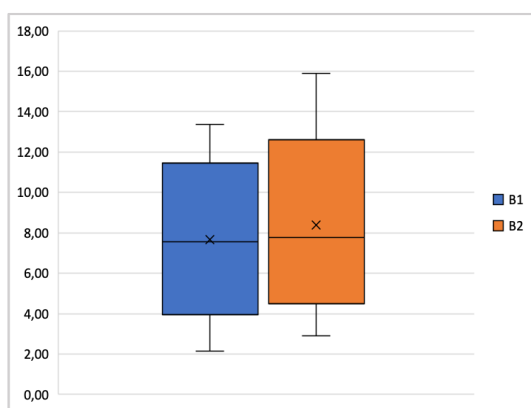


Gráfico 13. Diagrama de cajas para la variable “nivel de ELE”.

Se nota la capacidad del grupo avanzado para actualizar más palabras: el máximo y mínimo absolutos corresponden, respectivamente, a 15,88 y 2,90 mientras que en el conjunto de nivel umbral las cifras son menores: 13,38 y 2,14.

Los alumnos de nivel B2 presentan una desviación mayor de las respuestas entre el segundo y el tercer cuartil, que incluye la media matemática (8,15), la cual se coloca por encima de la mediana (7,78).

En cambio, media y mediana de los compañeros se solapan y, sin embargo, los valores que están por debajo o iguales a ellas representan el 50 % de los datos, y los mayores, el 50 % restante.

De esta manera, la producción intragrupal es más simétrica.

Asimismo, analizamos los promedios de vocablos.

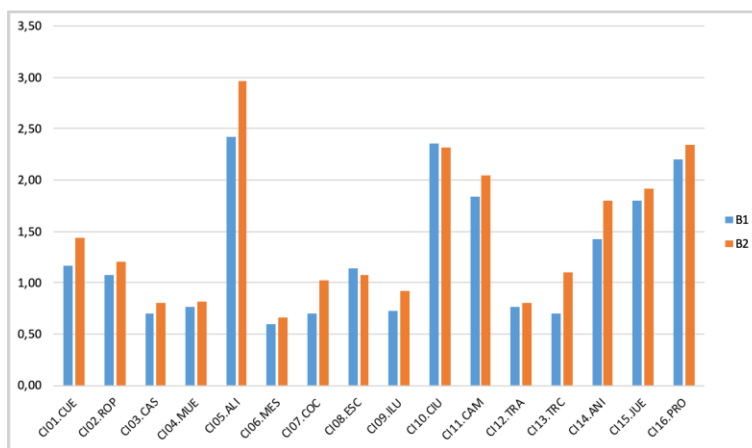


Gráfico 14. Promedio de vocablos según la variable “nivel de ELE”.

El grupo avanzado prevalece del +14,17% y mantiene su superioridad en todos los CI, llegando a responder con una media de *types* mayor del +57% en CI13, “trabajos del campo y del jardín”, y +46% en CI07, “la cocina y sus utensilios”, a excepción de los CI08, “la escuela: muebles y materiales”, (-5%) y CI10, “la ciudad”, (-2%), en los que domina, ligeramente, la otra variante.

En lo que atañe a los índices de cohesión y densidad, la tabla muestra los valores de las dos agrupaciones, desglosados por cada centro de interés.

CI	Índice de cohesión		Densidad léxica	
	B1	B2	B1	B2
CI01.CUE	0,205	0,208	10,24	10,38
CI02.ROP	0,135	0,137	6,74	7,12
CI03.CAS	0,171	0,149	8,57	7,75
CI04.MUE	0,166	0,161	8,32	8,39
CI05.ALI	0,111	0,103	5,53	5,36
CI06.MES	0,166	0,164	8,30	8,52
CI07.COC	0,092	0,077	4,60	4,00
CI08.ESC	0,138	0,146	6,88	7,61
CI09.ILU	0,096	0,071	4,81	3,67
CI10.CIU	0,101	0,095	5,05	4,96
CI11.CAM	0,076	0,063	3,79	3,29
CI12.TRA	0,208	0,198	10,42	10,30
CI13.TRC	0,061	0,051	3,06	2,64
CI14.ANI	0,146	0,120	7,30	6,24

CI15.JUE	0,093	0,074	4,63	3,82
CI16.PRO	0,084	0,071	4,19	3,70
PROMEDIO	0,128	0,118	6,40	6,11

Tabla 29. Índice de cohesión y densidad léxica por CI según la variable “nivel de ELE”.

Si partimos de los promedios, el desnivel entre las dos opciones no es sobresaliente: en B1 computamos un +8,47% en el índice de cohesión y un +4,25% en la densidad léxica, con lo cual parece que este factor no tiene influencia. Al contrario, los resultados por centro manifiestan que el B2 alcanza una cohesión más alta solo en tres (CI08, “la escuela: muebles y materiales”; CI01, “partes del cuerpo”; CI02, “la ropa”), pero en menor grado. Como consecuencia, el otro grupo es ligeramente más compacto en los demás campos semánticos. Lógicamente, se detecta la misma tendencia en la densidad: los avanzados presentan cifras un poco más bajas.

Con respecto a la incidencia de estos índices sobre la relación entre productividad y riqueza léxica, ponemos de relieve lo que ocurre en ambas variantes en el CI11, “el campo”, y el CI12, “los medios de transporte”: en el primero calculamos un bajo nivel de cohesión y densidad que corresponde a una gran variedad de *types* (rango 4) mientras que en el segundo los índices se colocan entre los mayores, por lo que el centro es más productivo que rico. Es más, esta dependencia se hace más manifiesta en B2, donde el CI01, “partes del cuerpo”, pasa de un rango 7 de riqueza a un 2 de rentabilidad y, al revés, el CI13, “trabajos del campo y del jardín”, de la posición 9 a la 16, conforme al cambio de estos índices.

Los gráficos muestran una imagen de la situación descrita.

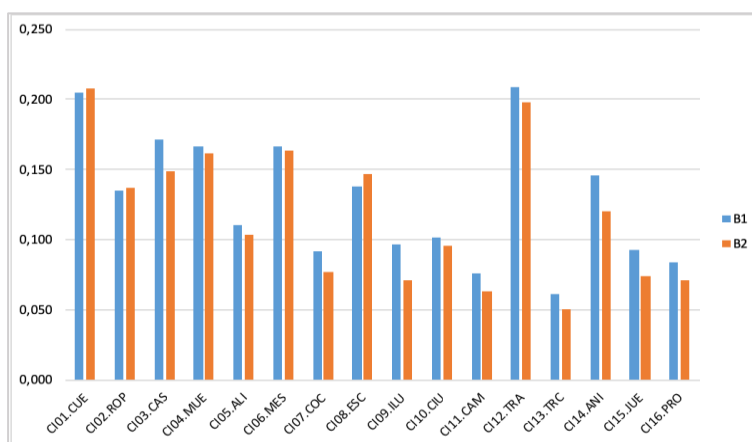


Gráfico 15. Índice de cohesión según la variable “nivel de ELE”.

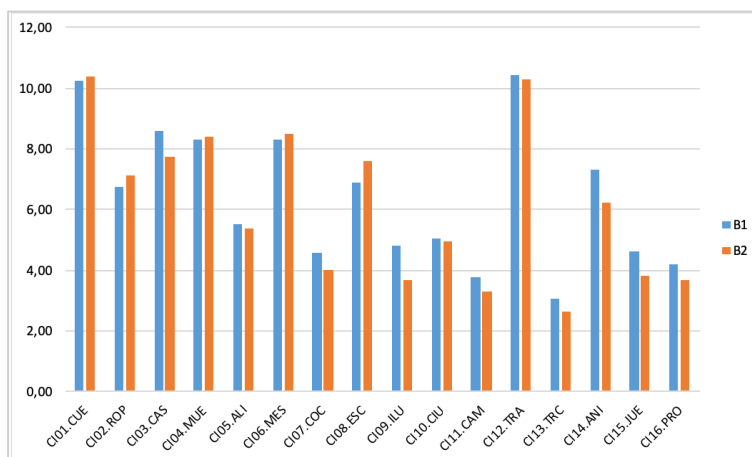


Gráfico 16. Densidad léxica según la variable “nivel de ELE”.

2.3.2.2 Conclusiones parciales

Notamos un aumento de los resultados del grupo B2, que produce el 7,51% más de palabras y el 14,05% de vocablos, con respecto al conjunto B1. Ofrecen un número mayor de respuestas en todos los índices, pese a que los alumnos de nivel inferior sean más productivos y variados en algunos centros, pero sin alcanzar valores notablemente más elevados: en el CI15, “juegos y distracciones”, incrementan su productividad de 50 *tokens* (3,62%) y en el CI08, “la escuela: muebles y materiales”, su riqueza de 3 *types* (5,55%).

Centrándonos en las palabras, siete centros se encuentran en los mismos rangos, en ambos muestreos (CI03, “partes de la casa (sin muebles)”); CI05, “alimentos y bebidas”; CI08, “la escuela: muebles y materiales”; CI12, “medios de transporte”; CI13, “trabajos del campo y del jardín”; CI14, “los animales”; CI16, “profesiones y oficios”).

Al observar los vocablos, son ocho los campos semánticos que coinciden sin distinción de grupo (CI01, “partes del cuerpo”; CI05, “alimentos y bebidas”; CI06, “objetos colocados en la mesa para la comida”; CI09, “iluminación y calefacción”; CI11, “el campo”; CI14, “los animales”; CI15, “juegos y distracciones”).

Al examinar la cohesión interna de cada sección, computamos en el nivel umbral un +8,47% en el índice de cohesión y un +4,25% en la densidad léxica, con lo cual es levemente más compacto en su producción de respuestas que el avanzado; lo mismo ocurre cuando se trata la densidad léxica. Estos índices inciden en la relación entre productividad y riqueza de los centros ya que, a menudo, donde calculamos un bajo nivel de cohesión y densidad detectamos una gran variedad léxica mientras que cuando los

valores aumentan los centros son más productivos que ricos.

2.3.3 Variable “conocimiento de otras LE”

Disponemos de dos agrupaciones dispares, los informantes que conocen dos lenguas extranjeras (LE =2) son 19 y los que conocen más de dos (LE >2) constituyen un conjunto más numeroso.

Informantes	Palabras	Vocablos	Promedio de palabras por informante	Promedio de vocablos por informante
LE =2 (19)	2.202	659	115,89	34,68
LE >2 (81)	10.377	1.381	128,11	17,05

Tabla 30. Índices generales de DL según la variable “conocimiento de otras LE”.

De nuevo, examinamos los rangos por palabras y por vocablos, y los promedios que nos permiten comprobar la incidencia de la variable. Sin embargo, señalamos ya que los alumnos que conocen más de dos LE aportan el 10,54% más de palabras en media, si bien no logran la misma supremacía en el promedio de vocablos, ya que los compañeros que solo conocen dos los superan del 50,84%.

Antes de empezar el análisis, veamos los datos brutos por cada muestreo.

Conoce 2 LE						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	242	2	12,74	53	6	2,79
CI02.ROP	135	8	7,11	34	9	1,79
CI03.CAS	111	11	5,84	25	10	1,32
CI04.MUE	100	13	5,26	24	11	1,26
CI05.ALI	264	1	13,89	92	1	4,84
CI06.MES	102	12	5,37	23	12	1,21
CI07.COC	60	14	3,16	22	13	1,16
CI08.ESC	131	9	6,89	35	8	1,84
CI09.ILU	50	15	2,63	16	15	0,84
CI10.CIU	193	4	10,16	62	4	3,26
CI11.CAM	119	10	6,26	53	7	2,79
CI12.TRA	158	5	8,32	18	14	0,95
CI13.TRC	36	16	1,89	15	16	0,79

CI14.ANI	208	3	10,95	66	3	3,47
CI15.JUE	150	6	7,89	54	5	2,84
CI16.PRO	143	7	7,53	67	2	3,53
TOTAL	2.202		7,24	659		2,17

Tabla 31. Índices de DL según la variante “LE =2”.

Conoce más de 2 LE						
CI	Palabras	Rango	Promedio por informante	Vocablos	Rango	Promedio por informante
CI01.CUE	1.099	2	13,57	82	7	1,01
CI02.ROP	656	7	8,10	70	8	0,86
CI03.CAS	499	12	6,16	47	15	0,58
CI04.MUE	560	11	6,91	48	14	0,59
CI05.ALI	1.199	1	14,80	160	1	1,98
CI06.MES	428	13	5,28	39	16	0,48
CI07.COC	305	14	3,77	63	10	0,78
CI08.ESC	672	6	8,30	67	9	0,83
CI09.ILU	292	15	3,60	52	13	0,64
CI10.CIU	978	3	12,07	149	3	1,84
CI11.CAM	566	10	6,99	126	4	1,56
CI12.TRA	650	8	8,02	56	12	0,69
CI13.TRC	216	16	2,67	62	11	0,77
CI14.ANI	872	4	10,77	88	6	1,09
CI15.JUE	634	9	7,83	121	5	1,49
CI16.PRO	751	5	9,27	151	2	1,86
TOTAL	10.377		8,01	1.381		1,07

Tabla 32. Índices de DL según la variante “LE >2”.

2.3.3.1 Análisis descriptivo

Los dos grupos coinciden en el grado de productividad en seis centros de interés (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI11, “el campo”; CI13, “trabajos del campo y del jardín”), de los que los más rentables son, otra vez, el CI05, “alimentos y bebidas”, y el CI01, “partes del cuerpo”.

Por otro lado, el CI13, “trabajos del campo y del jardín”, arroja menos *tokens* en absoluto (rango 16). Los demás campos se colocan en los mismos rangos, mientras que

los que se sitúan en posiciones distintas son CI08, “la escuela: muebles y materiales”; CI12, “medios de transporte”; CI15, “juegos y distracciones”, que cambian de tres puestos.

La variedad léxica pone de manifiesto lo mismo en lo referente a los CI menos ricos: el CI05, “alimentos y bebidas” (rango 1); el CI16, “profesiones y oficios” (rango 2); el CI15, “juegos y distracciones” (rango 5). Mientras que los campos con una mayor cantidad de vocablos son: CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI06, “objetos colocados en la mesa para la comida”; CI07, “la cocina y sus utensilios”; CI11, “el campo”, que cambian hasta 5 posiciones.

Los promedios por informante aparecen gráficamente como sigue.

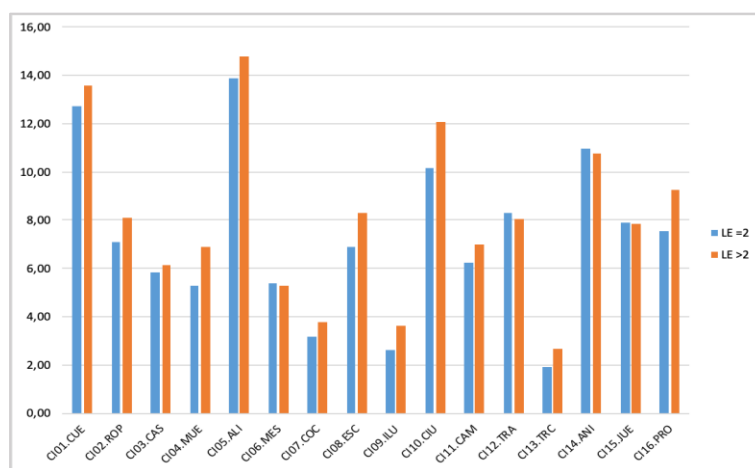


Gráfico 17. Promedio de palabras según la variable “conocimiento de otras LE”.

En general, la disparidad es del 10,63%, a favor de los encuestados que conocen más de dos LE.

Al desglosar los valores según centro de interés, se revela un desnivel más alto, de casi dos palabras, en el CI10, “la ciudad” (+18,86%), seguido por el CI08, “la escuela: muebles y materiales” (+20,33%); el CI16, “profesiones y oficios” (+23,19%); el CI04, “los muebles de la casa” (+31,36%). En los otros temas las diferencias son menores, pero siempre a favor de este conjunto.

El rendimiento aumenta ligeramente en el grupo LE =2 en CI12, “los medios de transporte” (+3,61%); CI14, “los animales” (+1,67%); CI06, “objetos colocados en la mesa para la comida” (+1,51%); CI15, “juegos y distracciones” (+0,89%).

La tabla 33 y el gráfico 18 ilustran la simetría o la dispersión detectadas.

Descriptivos	Variante	
	LE =2	LE >2
Media	7,24	8,01
Mediana	7,00	7,93
Mínimo	1,89	2,67
Máximo	13,89	14,80

Tabla 33. Estadísticos descriptivos para la variable “conocimiento de otras LE”.

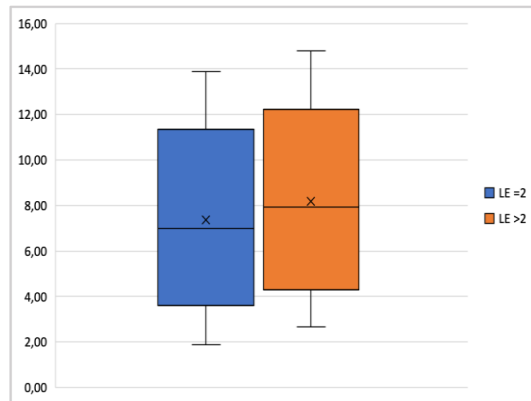


Gráfico 18. Diagrama de cajas para la variable “conocimiento de otras LE”.

De nuevo, la superioridad de los encuestados que conocen más de dos LE se hace manifiesta, pero pese a esta divergencia cuantitativa, se desprende una cierta correspondencia entre las desviaciones de los dos muestreos ya que el rango intercuartílico (la diferencia entre Q3 y Q1) es homogéneo: los intervalos entre cuartiles se revelan equilibrados en lo que atañe a la distribución de respuestas. La correlación de los promedios de vocablos cambia esta tendencia.

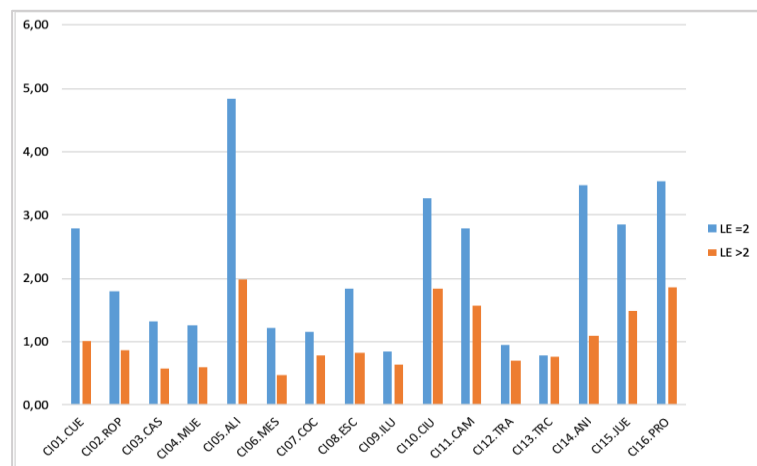


Gráfico 19. Promedio de vocablos según la variable “conocimiento de otras LE”.

De manera evidente, la disparidad se advierte en CI05, “alimentos y bebidas”; CI14, “los animales”; CI01, “partes del cuerpo”.

Al contrario, el desnivel baja en CI13, “trabajos del campo y del jardín” (+3%); CI09, “iluminación y calefacción” (+31%); CI12, “los medios de transporte” (+38%); CI07, “la cocina y sus utensilios” (+49%).

Por último, la cohesión se estudia a partir de la tabla a continuación.

CI	Índice de cohesión	
	LE =2	LE >2
CI01.CUE	0,240	0,165
CI02.ROP	0,209	0,116
CI03.CAS	0,234	0,131
CI04.MUE	0,219	0,144
CI05.ALI	0,151	0,093
CI06.MES	0,233	0,135
CI07.COC	0,144	0,060
CI08.ESC	0,197	0,124
CI09.ILU	0,164	0,069
CI10.CIU	0,164	0,081
CI11.CAM	0,118	0,055
CI12.TRA	0,462	0,143
CI13.TRC	0,126	0,043
CI14.ANI	0,166	0,122
CI15.JUE	0,146	0,065
CI16.PRO	0,112	0,061
PROMEDIO	0,193	0,101

Tabla 34. Índice de cohesión por CI según la variable “conocimiento de otras LE”.

Los índices de del grupo LE =2 son más altos y, por tanto, el conjunto de unidades aportadas es más cerrado y tiene una mayor asociación conceptual de las respuestas con respecto al otro que actualiza más palabras distintas, en particular en el CI12, “medios de transporte”; el CI03, “partes de la casa (sin muebles)”; el CI06, “objetos colocados en la mesa para la comida”. La diferencia mínima se da en el CI14, “los animales”, donde el intervalo entre lo dos se acerca.

El gráfico ilustra la situación.

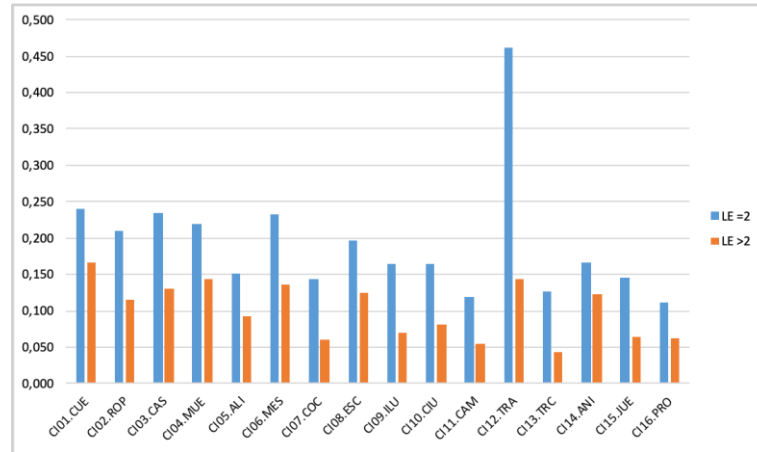


Gráfico 20. Índice de cohesión según la variable “conocimiento de otras LE”.

En conclusión, examinamos los índices de densidad léxica en cada CI.

CI	Densidad léxica			
	LE =2	Rango	LE >2	Rango
CI01.CUE	4,57	2	13,40	1
CI02.ROP	3,97	6	9,37	8
CI03.CAS	4,44	3	10,62	5
CI04.MUE	4,17	5	11,67	2
CI05.ALI	2,87	11	7,49	9
CI06.MES	4,43	4	10,97	4
CI07.COC	2,73	13	4,84	14
CI08.ESC	3,74	7	10,03	6
CI09.ILU	3,13	9	5,62	11
CI10.CIU	3,11	10	6,56	10
CI11.CAM	2,25	15	4,49	15
CI12.TRA	8,78	1	11,61	3
CI13.TRC	2,40	14	3,48	16
CI14.ANI	3,15	8	9,91	7
CI15.JUE	2,78	12	5,24	12
CI16.PRO	2,13	16	4,97	13
PROMEDIO	3,67		8,14	

Tabla 35. Densidad léxica por CI según la variable “conocimiento de otras LE”.

La variante LE =2 presenta una mayor densidad en el CI12, “los medios de transporte”, que se coloca en los rangos 14 por *types* y 5 por *tokens*, debido a su baja riqueza frente a su gran productividad. Lo mismo ocurre en LE >2 donde este centro es uno de los más densos y cohesionados (rango 3), precedido por el CI01, “partes del

cuerpo”, (rango 1) y el CI04, “los muebles de la casa”, (rango 2) que producen una grandísima cantidad de palabras. Al contrario, encontramos una baja media de repeticiones en el CI16, “profesiones y oficios”, que es uno de los más ricos en ambos grupos.

2.3.3.2 Conclusiones parciales

El factor conocimiento de otras LE pone de manifiesto que el grupo LE >2 arroja una cantidad mayor de palabras del 10,54%, pero con respecto a la media de vocablos el otro grupo lo supera del 50,84%.

La capacidad de producción de los informantes que conocen más de dos LE sobresale en la mayoría de los CI, si excluimos que el rendimiento aumenta ligeramente a favor de los que saben dos LE en cuatro campos.

Al contrario, estos aportan más del doble de vocablos, en particular, en CI05, “alimentos y bebidas”; CI14, “los animales”; CI01, “partes del cuerpo”, índice en el cual el grupo LE >2 presenta datos pobres.

En la correlación de los datos del rango por palabras, las dos variantes coinciden en productividad en seis centros (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI11, “el campo”; CI13, “trabajos del campo y del jardín”), los mismos resultan menos variados en ambos grupos.

El análisis de la cohesión interna hace patente que la muestra LE =2 logra valores superiores, las unidades léxicas que aportan tienen una mayor asociación conceptual con respecto a las respuestas de sus compañeros.

2.4 Conclusiones del análisis transversal

Dentro de la población investigada se manifiesta una distribución bastante heterogénea de los informantes en función de los factores sociolingüísticos analizados:

1. Variable sexo:

- Mujeres: 87;
- Hombres: 13.

2. Variable nivel de ELE:

- Nivel B1: 50;

- Nivel B2: 50.

3. Variable conocimiento de otras LE:

- Conoce 2 LE: 19;
- Conoce más de 2 LE: 81.

Hemos trabajado con los promedios para ofrecer un análisis lo más preciso posible, del cual ha surgido que entre los dieciséis centros de interés que hemos propuesto durante la prueba los que resultan más o menos productivos y más o menos ricos son los mismos.

Hay cierta correspondencia sea en el análisis de los resultados generales sea en los resultados por variable: apreciamos cuatro campos que tienen constantemente los valores más altos de productividad léxica: CI01, “partes del cuerpo”; CI05, “alimentos y bebidas”; CI10, “la ciudad”; CI14, “los animales”. Contamos, además, con otros que siempre se han caracterizado como los menos rentables: CI06, “objetos colocados en la mesa para la comida”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”.

De igual manera, los centros más ricos son CI05, “alimentos y bebidas”; CI10, “la ciudad”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios”; y los más pobres son CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI06, “objetos colocados en la mesa para la comida”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”. Creemos que esto se debe al escaso interés que estos temas suscitan en los aprendices o a la poca atención que les dedican las propuestas didácticas.

El análisis del índice de cohesión y de densidad léxica destaca que algunos estímulos temáticos se presentan como más cerrados y homogéneos (CI01, “partes del cuerpo”; CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI12, “medios de transporte”) y otros, al revés, son más abiertos y difusos (CI11, “el campo”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios”). Parece que una mayor concreción semántica se debe a un aprendizaje escolar sistematizado y que los estudiantes no han ampliado sus conocimientos por encima de lo propuesto en clase. En el caso opuesto, cuando las actualizaciones son más variadas, cada informante ha escrito vocablos distintos de los otros ya que se trata de argumentos que ha aprendido, o profundizado, por su cuenta.

En lo que se refiere al factor sexo, pese a que el desnivel entre la productividad de mujer y hombre no sea tan elevado (+9,61% a favor del conjunto femenino), destaca la supremacía de los varones en la producción de vocablos que supera a sus compañeras en un 66,35%, valores que se ven reflejados en el análisis del grado de cohesión de las respuestas.

Seguidamente, como esperábamos, el factor nivel de ELE supone un aumento del léxico disponible activado por los informantes de nivel B2, bien en el número de *tokens* (+7,51%) bien en el número de *types* (+14,05%). Los participantes de nivel B1, por su parte, presentan una cohesión ligeramente mayor, con lo cual sus respuestas son más homogéneas.

Por último, lo que sobresale del análisis del factor conocimiento de otras LE es que los sujetos que conocen más de dos LE arrojan más palabras (+10,24%), confirmando la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979), pero, a la vez, sus aportaciones no rinden tanta variedad léxica: los que conocen solo dos LE ofrecen el 50,84% de vocablos más y alcanzan valores más altos de cohesión.

3. ANÁLISIS LONGITUDINAL

Abordamos ahora el análisis longitudinal: cotejamos los datos procedentes de la primera administración de la prueba y de la segunda, realizadas ambas por los mismos estudiantes de nivel B2 y de las cuales hemos recopilado los corpus B2_a y B2_b. Nuestro propósito es medir la evolución en el tiempo de los conocimientos de los participantes.

Apreciamos un total de 15.008 palabras y 2.405 vocablos como resume la tabla.

	B2_a	B2_b
Informantes	50	50
Palabras	6.517	8.491
Vocablos	1.161	1.244
Promedio de palabras	130,34	169,82
Promedio de vocablos	23,22	24,88
Índice de cohesión	0,118	0,146
Densidad léxica	6,11	7,30

Tabla 36. Índices generales de DL según el análisis longitudinal.

3.1 Número de palabras

Las tablas contienen el total de palabras por cada corpus, sus rangos y promedios.

I suministración			II suministración		
Rango	CI	Palabras	Rango	CI	Palabras
2	CI01.CUE	747	2	CI01.CUE	785
6	CI02.ROP	427	7	CI02.ROP	594
12	CI03.CAS	310	11	CI03.CAS	421
10	CI04.MUE	344	10	CI04.MUE	444
1	CI05.ALI	794	1	CI05.ALI	850
13	CI06.MES	281	13	CI06.MES	399
14	CI07.COC	204	15	CI07.COC	288
8	CI08.ESC	411	5	CI08.ESC	689
15	CI09.ILU	169	14	CI09.ILU	338
3	CI10.CIU	575	4	CI10.CIU	710
11	CI11.CAM	336	12	CI11.CAM	402
7	CI12.TRA	412	8	CI12.TRA	522
16	CI13.TRC	145	16	CI13.TRC	208
4	CI14.ANI	562	3	CI14.ANI	729
9	CI15.JUE	367	9	CI15.JUE	484
5	CI16.PRO	433	6	CI16.PRO	628
TOTAL		6.517	TOTAL		8.491
PROMEDIO		407,31	PROMEDIO		530,69

Tabla 37. Número de palabras por CI según el análisis longitudinal.

En B2_a los informantes alcanzan un total de 6.517 respuestas y constatamos un incremento del 30,29% de los valores en B2_b, donde contabilizamos 8.491 *tokens*.

En la segunda prueba los alumnos han arrojado un número mayor de palabras en todos los campos, desde un +5% en el CI01, “partes del cuerpo”, hasta un +100% en el CI09, “iluminación y calefacción”.

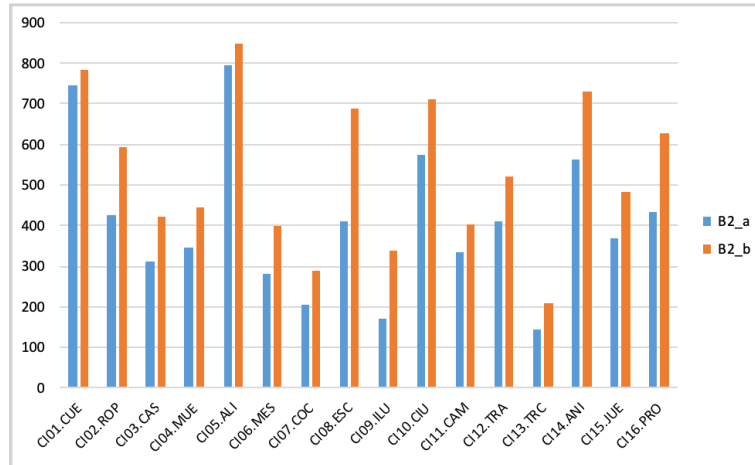


Gráfico 21. Número total de palabras según la fecha de la prueba.

Al considerar el orden de los centros de interés por rango, esto es, de su mayor o menor productividad, podemos clasificar los campos en función de si se encuentran por encima o por debajo de la media de palabras arrojadas.

I suministración			II suministración		
Rango	CI	Palabras	Rango	CI	Palabras
1	CI05.ALI	794	1	CI05.ALI	850
2	CI01.CUE	747	2	CI01.CUE	785
3	CI10.CIU	575	3	CI14.ANI	729
4	CI14.ANI	562	4	CI10.CIU	710
5	CI16.PRO	433	5	CI08.ESC	689
6	CI02.ROP	427	6	CI16.PRO	628
7	CI12.TRA	412	7	CI02.ROP	594
8	CI08.ESC	411	8	CI12.TRA	522
9	CI15.JUE	367	9	CI15.JUE	484
10	CI04.MUE	344	10	CI04.MUE	444
11	CI11.CAM	336	11	CI03.CAS	421
12	CI03.CAS	310	12	CI11.CAM	402
13	CI06.MES	281	13	CI06.MES	399
14	CI07.COC	204	14	CI09.ILU	338
15	CI09.ILU	169	15	CI07.COC	288
16	CI13.TRC	145	16	CI13.TRC	208

Tabla 38. CI ordenados en función de su productividad léxica según el análisis longitudinal.

En la primera prueba los centros que sobrepasan la media son ocho, mientras que en la segunda son siete.

Seis campos se colocan en los mismos rangos (CI05, “alimentos y bebidas”; CI01,

“partes del cuerpo”; CI15, “juegos y distracciones”; CI04, “los muebles de la casa”; CI06, “objetos colocados en la mesa para la comida”; CI13, “trabajos del campo y del jardín”), con lo cual la productividad de los CI es la misma, pero con un crecimiento cuantitativo notable. Dentro de este conjunto, el CI05, “alimentos y bebidas”, ocupa la primera posición en ambas pruebas, donde computamos respectivamente 794 y 850 palabras, calculando un incremento del 7%. El segundo centro más productivo es el CI01, “partes del cuerpo”, que pasa de 747 a 785 *tokens*, crece del 5%. Incluso el campo menos rentable, el CI13, “trabajos del campo y del jardín”, goza de un aumento de 145 a 208 ítems.

Otros centros se han posicionado en rangos superiores, resultando más productivos (CI14, “los animales”; CI08, “la escuela: muebles y materiales”; CI03, “partes de la casa (sin muebles)”; CI09, “iluminación y calefacción”) en detrimento de otros que han retrocedido, pero siempre ofreciendo un enriquecimiento cuantitativo.

CI	I suministración	II suministración
CI01.CUE	14,94	15,70
CI02.ROP	8,54	11,88
CI03.CAS	6,20	8,42
CI04.MUE	6,88	8,88
CI05.ALI	15,88	17,00
CI06.MES	5,62	7,98
CI07.COC	4,08	5,76
CI08.ESC	8,22	13,78
CI09.ILU	3,38	6,76
CI10.CIU	11,50	14,20
CI11.CAM	6,72	8,04
CI12.TRA	8,24	10,44
CI13.TRC	2,90	4,16
CI14.ANI	11,24	14,58
CI15.JUE	7,34	9,68
CI16.PRO	8,66	12,56
TOTAL	8,15	10,61

Tabla 39. Promedio de palabras por informante por CI según el análisis longitudinal.

En la segunda prueba se computa un aumento del 30,29% en la media de respuestas, lo cual refleja el aumento de actualizaciones y destaca una mayor producción léxica que se verifica después de la asistencia al curso. Al traducir estos datos en porcentajes, las diferencias van del +5% al +100%.

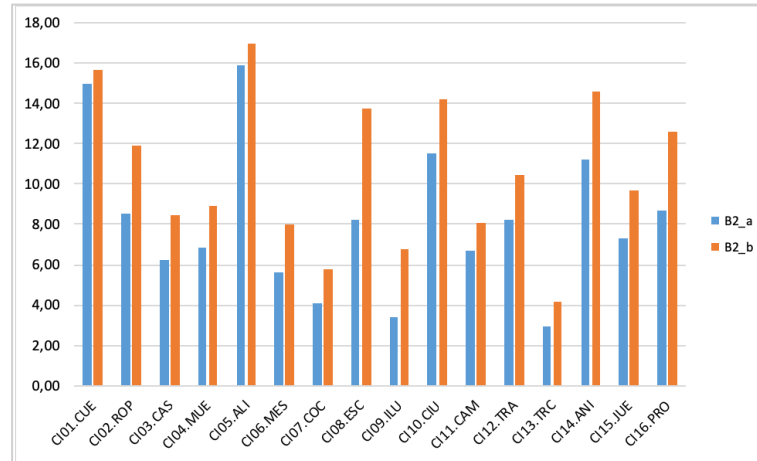


Gráfico 22. Promedio de palabras por informante según la fecha de la prueba.

La tabla proporciona los datos estadísticos que corroboran lo expuesto.

Descriptivos	Suministración	
	I	II
Media	8,15	10,61
Mediana	7,78	10,06
Mínimo	2,90	4,16
Máximo	15,88	17,00

Tabla 40. Estadísticos descriptivos de las dos suministraciones de la prueba.

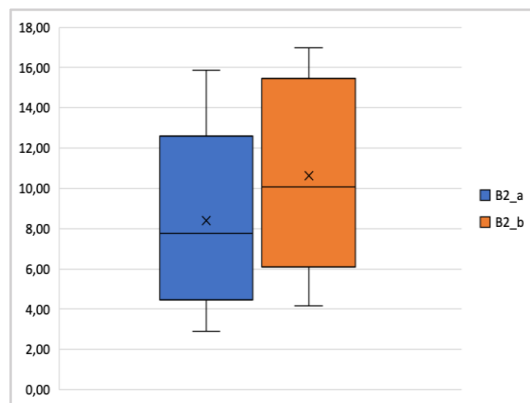


Gráfico 23. Diagrama de cajas según la fecha de la prueba.

Se confirma la mejoría de los resultados alcanzados en la segunda aplicación de la prueba; la posición de las cajas en el gráfico permite observar dónde la variación intragrupal es mayor: en los valores más altos que se posicionan entre la mediana y el tercer cuartil del corpus B2_b. Igualmente, el B2_a presenta una asimetría más grande en el mismo intervalo, pero la diferencia entre cuartiles es menor, la caja es más alargada.

3.2 Número de vocablos

Pasamos al cotejo de la variedad léxica alcanzada en cada uno de los momentos.

I suministración			II suministración		
Rango	CI	Vocablos	Rango	CI	Vocablos
7	CI01.CUE	72	9	CI01.CUE	65
8	CI02.ROP	60	8	CI02.ROP	67
14	CI03.CAS	40	14	CI03.CAS	45
13	CI04.MUE	41	15	CI04.MUE	41
1	CI05.ALI	148	1	CI05.ALI	140
16	CI06.MES	33	16	CI06.MES	35
11	CI07.COC	51	11	CI07.COC	61
10	CI08.ESC	54	4	CI08.ESC	99
12	CI09.ILU	46	12	CI09.ILU	58
3	CI10.CIU	116	2	CI10.CIU	124
4	CI11.CAM	102	6	CI11.CAM	96
15	CI12.TRA	40	10	CI12.TRA	65
9	CI13.TRC	55	13	CI13.TRC	49
6	CI14.ANI	90	5	CI14.ANI	97
5	CI15.JUE	96	7	CI15.JUE	89
2	CI16.PRO	117	3	CI16.PRO	113
TOTAL		1.161	TOTAL		1.244
PROMEDIO		23,22	PROMEDIO		24,88

Tabla 41. Número de vocablos por CI según el análisis longitudinal.

En una primera lectura, al analizar la cantidad de palabras diferentes no se observan valores más altos en el corpus B2_b, ya que resulta menos rico en seis campos semánticos (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI15, “juegos y distracciones”; CI11, “el campo”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios”), hasta disminuir del 5,7% en el CI05, “alimentos y bebidas.

En todo caso, apuntamos en general un aumento de 83 *types* (+7,50%) en las respuestas de la segunda prueba, esto es, de 1,66 vocablos de media por informante.

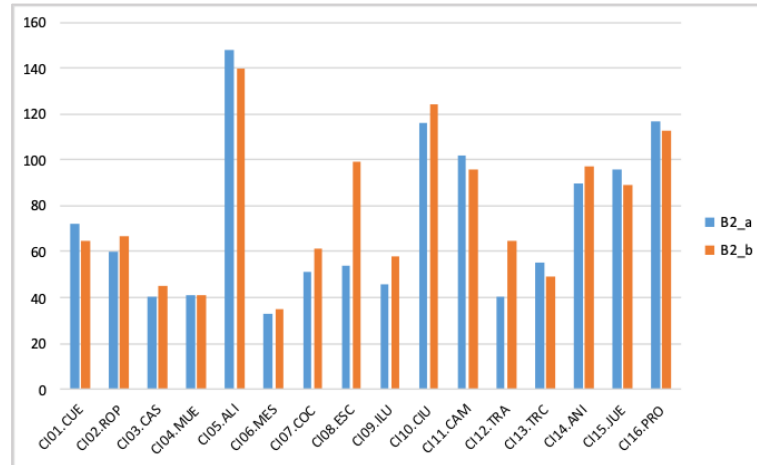


Gráfico 24. Número total de vocablos según la fecha de la prueba.

Los centros quedan ordenados por rango del siguiente modo.

I suministración			II suministración		
Rango	CI	Vocablos	Rango	CI	Vocablos
1	CI05.ALI	148	1	CI05.ALI	140
2	CI16.PRO	117	2	CI10.CIU	124
3	CI10.CIU	116	3	CI16.PRO	113
4	CI11.CAM	102	4	CI08.ESC	99
5	CI15.JUE	96	5	CI14.ANI	97
6	CI14.ANI	90	6	CI11.CAM	96
7	CI01.CUE	72	7	CI15.JUE	89
8	CI02.ROP	60	8	CI02.ROP	67
9	CI13.TRC	55	9	CI01.CUE	65
10	CI08.ESC	54	10	CI12.TRA	65
11	CI07.COC	51	11	CI07.COC	61
12	CI09.ILU	46	12	CI09.ILU	58
13	CI04.MUE	41	13	CI13.TRC	49
14	CI03.CAS	40	14	CI03.CAS	45
15	CI12.TRA	40	15	CI04.MUE	41
16	CI06.MES	33	16	CI06.MES	35

Tabla 42. CI ordenados en función de su variedad léxica según el análisis longitudinal.

En ambos corpus siete centros superan la media, que es respectivamente de 72,5 y 77,7 vocablos, una cantidad cuatro veces mayor con respecto a los menos ricos. Dentro del conjunto examinado seis temas se posicionan exactamente en los mismos rangos en las dos pruebas (CI05, “alimentos y bebidas”; CI02, “la ropa”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI03, “partes de la casa (sin muebles)”;

CI06, “objetos colocados en la mesa para la comida”). Destacamos que los cambios más evidentes que afectan a los centros CI08, “la escuela: muebles y materiales”, que pasa del rango 10 al 4 aumentando su riqueza del 83%, y CI12, “los medios de transporte”, que incrementa del 63%, desplazándose del 15 al 10. Por otro lado, el CI13, “trabajos del campo y del jardín”, baja de la posición 9 a la 13, disminuyendo del 12%. Se confirma lo que acabamos de decir sobre las diferencias entre las pruebas: se ha producido un +7%.

CI	I suministración	II suministración
CI01.CUE	1,44	1,30
CI02.ROP	1,20	1,34
CI03.CAS	0,80	0,90
CI04.MUE	0,82	0,82
CI05.ALI	2,96	2,80
CI06.MES	0,66	0,70
CI07.COC	1,02	1,22
CI08.ESC	1,08	1,98
CI09.ILU	0,92	1,16
CI10.CIU	2,32	2,48
CI11.CAM	2,04	1,92
CI12.TRA	0,80	1,30
CI13.TRC	1,10	0,98
CI14.ANI	1,80	1,94
CI15.JUE	1,92	1,78
CI16.PRO	2,34	2,26
TOTAL	1,45	1,56

Tabla 43. Promedio de vocablos por informante por CI según el análisis longitudinal.

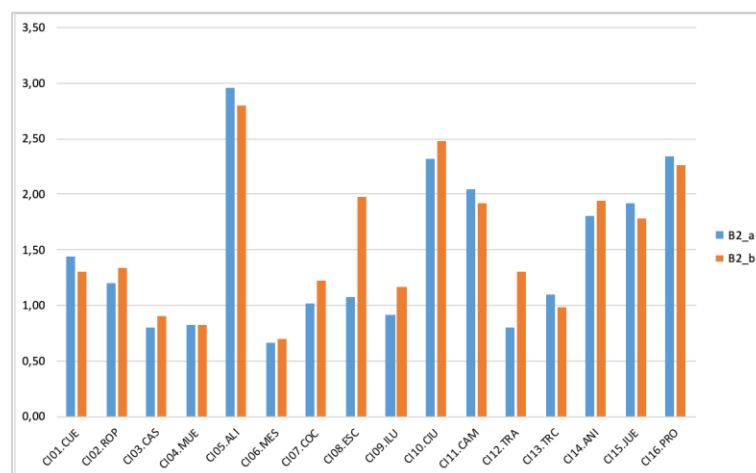


Gráfico 25. Promedio de vocablos por informante según la fecha de la prueba.

3.3 Índice de cohesión y densidad léxica

Cierran la sección los índices de cohesión y de densidad interna de los CI.

CI	I suministración			II suministración		
	Rango	Índice de cohesión	Densidad léxica	Rango	Índice de cohesión	Densidad léxica
CI01.CUE	1	0,208	10,38	1	0,242	12,08
CI02.ROP	7	0,137	7,12	5	0,177	8,87
CI03.CAS	5	0,149	7,75	4	0,187	9,36
CI04.MUE	4	0,161	8,39	3	0,217	10,83
CI05.ALI	9	0,103	5,36	9	0,121	6,07
CI06.MES	3	0,164	8,52	2	0,228	11,40
CI07.COC	11	0,077	4,00	14	0,094	4,72
CI08.ESC	6	0,146	7,61	8	0,139	6,96
CI09.ILU	14	0,071	3,67	10	0,117	5,83
CI10.CIU	10	0,095	4,96	11	0,115	5,73
CI11.CAM	15	0,063	3,29	16	0,084	4,19
CI12.TRA	2	0,198	10,30	6	0,161	8,03
CI13.TRC	16	0,051	2,64	15	0,085	4,24
CI14.ANI	8	0,120	6,24	7	0,150	7,52
CI15.JUE	12	0,074	3,82	13	0,109	5,44
CI16.PRO	13	0,071	3,70	12	0,111	5,56
PROMEDIO		0,118	6,11	PROMEDIO	0,146	7,30

Tabla 44. Índice de cohesión y densidad léxica por CI según el análisis longitudinal.

Las respuestas registradas en el corpus B2_b forman un conjunto de palabras más cerrado con respecto a las del B2_a: el índice de cohesión crece del 23,75%.

En efecto, a lo largo del análisis observamos que en la segunda aplicación los valores aumentan en todos los campos, hasta alcanzar un 67% en el CI13, “trabajos del campo y del jardín”.

Las únicas excepciones son el CI12, “los medios de transporte”, y el CI08, “la escuela: muebles y materiales”, donde bajan del 19% y 5%. Los informantes tienden a actualizar más unidades léxicas distintas al final del año académico que a comienzo.

Se confirma la teoría de Schmitt (2000: 42) sobre la organización del lexicón, que afirma que al alcanzar un nivel lingüístico más elevado se desarrollan más asociaciones mentales y, en consecuencia, el vocabulario resulta más articulado.

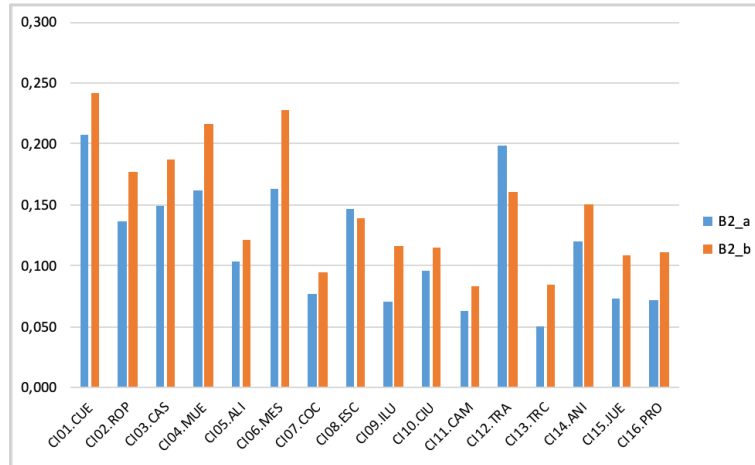


Gráfico 26. Índice de cohesión según la fecha de la prueba.

En lo que se refiere a la densidad léxica, el histograma describe la misma tendencia, es decir, que en la segunda prueba los encuestados presentan una mayor repetición de palabras hasta un 61% más (en promedio +19,49%). De nuevo, encontramos dos casos aislados, el CI12, “los medios de transporte”, y el CI08, “la escuela: muebles y materiales”, en los que hay más palabras diferentes: los valores bajan del 22% y del 9%.

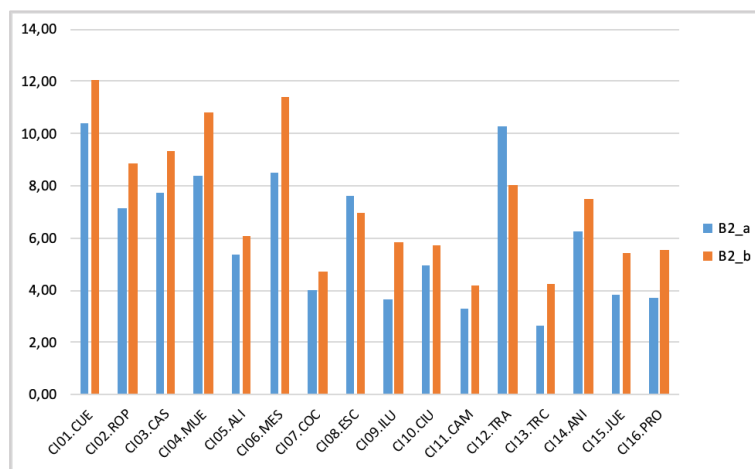


Gráfico 27. Densidad léxica según la fecha de la prueba.

Tras descartar algunos campos donde no se registra ninguna dependencia, en otros (CI01, “partes del cuerpo”; CI12, “medios de transporte”; CI06, “objetos colocados en la mesa para la comida”; CI04, “los muebles de la casa”; CI03, “partes de la casa (sin muebles)”) los valores de coherencia semántica interna son muy elevados. Detectamos la situación opuesta en CI16, “profesiones y oficios” y en CI11, “el campo”, donde los índices resultan menores y suponen una mayor variedad léxica.

3.4 Conclusiones del análisis longitudinal

En la primera aplicación de la prueba, el grupo arroja un total de 6.517 palabras mientras que en la segunda aporta 8.491, conllevando un aumento del 30,29%.

Los informantes han mejorado su capacidad productiva en todos los campos, desde un +5% en el CI01, “partes del cuerpo”, hasta un +100% en el CI09, “iluminación y calefacción”. En lo que atañe al número total de vocablos, los datos del corpus B2_b no siguen dominando en todas las áreas, sino que hay seis campos semánticos menos ricos, aunque siguen teniendo más palabras diferentes: notamos un incremento de 83 *types* (+7,50%), es decir de 1,66 vocablos de media por informante.

Los índices de cohesión manifiestan que en la segunda prueba los participantes producen un conjunto de palabras más cerrado, la cohesión crece del 23,75% debido a que los valores aumentan en todos los CI hasta un +67% en el CI13, “trabajos del campo y del jardín”. La densidad léxica revela lo mismo, o sea que los encuestados al final del año académico consiguen valores más altos que implican una cantidad más alta de repetición de palabras en todos los centros, llegando a un +61%.

Se observa una mayor producción y riqueza léxica en los estudiantes después de la asistencia al curso: el bagaje de vocabulario conocido ha incrementado y la acción docente ha sido fructífera.

4. ANÁLISIS COMPARATIVO

Cotejamos ahora el léxico disponible de distintos grupos de aprendices de ELE, con el propósito de analizar las similitudes y las diferencias entre los conocimientos.

Pese a que “los vocablos de muchos campos muestran una estrecha relación con las peculiaridades culturales” (Šifrar Kalan, 2009: 180) y que “la disponibilidad léxica es una medida de las diferencias culturales” (Mackey, 1971a: 26-27 *apud* Carcedo, 2000c: 54), realizamos este análisis para averiguar si los estudiantes de ELE que comparten el nivel lingüístico tienen la misma competencia, aunque tengan lenguas maternas diferentes, tan lejanas con respecto al italiano (alemán, esloveno, finés, japonés, polaco, sueco) y vivan en entornos culturales y escolares igualmente distintos⁹⁹.

⁹⁹ Habría que considerar otros elementos para llevar a cabo una comparación exhaustiva, por ejemplo, como afirma Šifrar Kalan (2009: 180): “[...] lo que verdaderamente hace falta es tomar en cuenta las

Durante la fase de revisión que nos ha llevado a decidir cuáles investigaciones incluir en esta comparación, hemos tenido en cuenta aprendices del mismo tipo que los de nuestro estudio, es decir, hemos elegido exclusivamente trabajos referidos a estudiantes universitarios de nivel intermedio. Algunos de los estudios que analizamos cuentan también con alumnos de la educación secundaria, pero a la hora de confrontar los resultados hemos recopilado solo los datos relativos a los alumnos procedentes del entorno educativo que nos interesaba, prescindiendo de los demás: en las tablas que siguen exponemos los datos filtrados según nuestros intereses.

Hemos procedido a la selección aun a sabiendas de que no todas las encuestas se centraban en nuestras mismas áreas temáticas. En todo caso, hemos podido realizar la comparación por lo menos entre nueve CI¹⁰⁰.

Los estudios cotejados se dividen en dos bloques conforme al contexto de aprendizaje: estudio del español LE el país de origen y en inmersión. En orden cronológico, se trata de investigaciones llevadas a cabo en Finlandia, España (Salamanca y Andalucía), Eslovenia, China e Italia.

El primer trabajo es de Carcedo (2000c) que se centra en 50 estudiantes finlandeses universitarios, hablantes de finés y sueco. Samper Hernández (2002) trabaja con 14 aprendices en inmersión en la Universidad de Salamanca, cuyas lenguas maternas son inglés, italiano y japonés. Sánchez-Saus (2011) analiza un grupo de 125 informantes que estudia español en las Universidades de Andalucía, nativos de alemán, finés, francés, inglés, italiano y polaco. A continuación, Šifrar Kalan (2014b) proporciona datos procedentes de un estudio dedicado a 100 estudiantes eslovenos. Hidalgo (2017b) se dedica a 294 encuestados sinohablantes y, por último, Del Barrio y Vann (2018) elaboran su proyecto basándose en 68 estudiantes universitarios.

características del input de ELE, de las cuales depende la competencia léxica: el tipo de metodología, el input por parte del profesor y de los manuales sobre todo en los estudiantes que no estudian ELE en inmersión.” Pero, todo esto no se incluye en los objetivos del presente trabajo, aunque puede abrir una nueva vía de investigación.

¹⁰⁰ Nuestros centros de interés coinciden con los analizados por Carcedo (2000c), Samper Hernández (2002), Del Barrio y Vann (2018). Por otro lado, Sánchez-Saus (2011) trabaja con dieciocho campos semánticos, entre los cuales detectamos doce que corresponden con los nuestros, pese a la denominación distinta (“el cuerpo humano”; “la casa”, que dividimos en “partes de la casa (sin muebles)” y “los muebles de la casa”; “escuela y universidad”; “ocio y tiempo libre”; “profesiones y trabajos”). Šifrar Kalan (2014b) analiza once áreas, de las que descartamos “acciones que se realizan todos los días” al no tenerla en nuestra investigación y “la casa” que nosotros separamos en dos CI. Por su parte, Hidalgo (2017b) estudia dieciocho temáticas, diez que coinciden con las nuestras, no obstante, otra vez, la designación sea diferente (“el cuerpo humano”; “la ropa y complementos”; “la casa”, que diferenciamos de “los muebles de la casa”; “comidas y bebidas”; “escuela y universidad”; “ocio y tiempo libre”; “profesiones y trabajos”).

De ahí que hayamos tomado como valores de referencia los promedios de respuesta ya que no dependen del número de participantes¹⁰¹.

4.1 Análisis descriptivo

La tabla 45 resume las características principales de estos trabajos, sobresale enseguida que los tamaños muestrales cambian mucho entre una y otra investigación. Por eso, no hemos empleado los valores en bruto, pues, se habría desvirtuado el resultado de la comparación: analizamos exclusivamente los datos relativos a los nueve campos comunes, además de los promedios en lugar del total de palabras y vocablos.

Investigación	Informantes	País	CI	Promedio palabras	Promedio vocablos
Nalesso (2019) ¹⁰²	100	Italia	16	125,79	14,90
Carcedo (2000c)	50	Finlandia	16	162,12	44,36
Samper Hernández (2002)	14	España	16	174,92	65,07
Sánchez-Saus (2011)	125	España	18	285,66	47,17
Šifrar Kalan (2014b)	100	Eslovenia	11	174,24	28,22
Hidalgo (2017b)	294	China	18	284,99	42,60
Del Barrio y Vann (2018)	68	Italia	16	145,93	26,41

Tabla 45. Datos generales de DL de las investigaciones cotejadas.

4.1.1 Promedio de palabras

Empezamos la comparación del léxico disponible en los nueve CI compartidos por las siete investigaciones analizadas a partir de la siguiente tabla.

Vemos, en general, que el centro de mayor rendimiento es el CI05, “alimentos y bebidas”, mientras que los menos rentables son el CI12, “los medios de transporte”, y el CI02, “la ropa”.

CI	Nalesso (2019)	Carcedo (2000c)	Samper Hernández (2002)	Sánchez-Saus (2011)	Šifrar Kalan (2014b)	Hidalgo (2017b)	Del Barrio y Vann (2018)
CI01. CUE	13,41	16,22	16	18,34	19,07	17,29	14,51

¹⁰¹ Hemos computado el promedio que requeríamos para la comparación en el caso de que no hubiese ya sido calculado en el estudio.

¹⁰² Con Nalesso (2019) nos referimos obviamente a los datos de este trabajo.

CI02. ROP	7,91	12,18	11,85	12,57	11,96	11,9	10,44
CI03. CAS	6,10	9,66	8,85	18,12	18,55	16,74	8,12
CI05. ALI	14,63	18,14	21,21	22,65	20,11	18,06	14,94
CI10. CIU	11,71	12,46	19,92	19,18	16,44	16,73	12,54
CI12. TRA	8,08	10,26	11,35	13,08	12,23	13,38	9,28
CI14. ANI	10,80	15,02	12,35	13,38	15,19	15,58	11,51
CI15. JUE	7,84	8,76	12,14	18,27	14,06	17,88	8,66
CI16. PRO	8,94	12,54	12,21	13,58	13,96	16,54	10,44
PROMEDIO	9,94	12,80	13,99	16,57	15,73	16,01	11,16

Tabla 46. Promedios de palabras de las investigaciones cotejadas.

En orden, los encuestados que han actualizado más palabras son los de: Sánchez-Saus (2011) con una media de 16,57; Hidalgo (2017b) con 16,01; Šifrar Kalan (2014b) con 15,73; Samper Hernández (2002) con 13,99; Carcedo (2000c) con 12,80; Del Barrio y Vann (2018) con 11,16; Nalesso (2019) con 9,94.

Los informantes de Sánchez-Saus (2011) predominan en tres centros, quizá se deba al contexto de aprendizaje en inmersión, lo que podría conllevar una ventaja con respecto a los otros aprendices, aunque no hay esta misma supremacía en Samper Hernández (2002), cuyos alumnos se encontraban también en un contexto de inmersión. Los encuestados por Hidalgo (2017b) también aportan el número más alto de palabras en tres campos. No sorprenden los resultados alcanzados en Šifrar Kalan (2014b) ya que sus estudiantes tienen un nivel B2+, mayor con respecto a los otros, y presentan índices superiores en dos ámbitos. A continuación, se colocan los finlandeses de Carcedo (2000c), los italianos de Del Barrio y Vann (2018) y los nuestros que no sobresalen en ninguno.

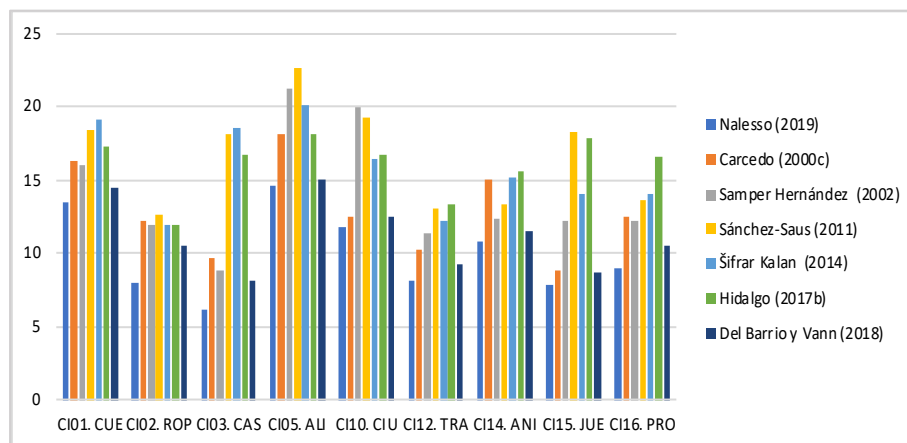


Gráfico 28. Promedios de palabras por informante en las investigaciones cotejadas.

Es interesante observar el promedio de palabras por informante y compararlo con trabajos dedicados a hispanohablantes nativos¹⁰³, que suelen superar las 20 respuestas por sujeto (Samper Hernández, 2002). No extrañan los valores más bajos, derivados de la menor competencia de los participantes en estas encuestas al ser estudiantes extranjeros de nivel intermedio. De todos modos, merece la pena subrayar que tres grupos dominan en el CI05, “alimentos y bebidas”, aportando un promedio de más de 20 palabras (Samper Hernández, 2002; Sánchez-Saus, 2011; Šifrar Kalan, 2014b). A continuación, los informantes de Carcedo (2000c) e Hidalgo (2017b) también arrojan cantidades apreciables, si bien se quedan por debajo de las 20 unidades.

Por último, vemos los descriptores estadísticos.

Investigación	Media	Mediana	Mínimo	Máximo
Nalesso (2019)	8,94	8,94	6,10	14,63
Carcedo (2000c)	11,52	12,46	8,76	18,14
Samper Hernández (2002)	12,59	12,21	8,85	21,21
Sánchez-Saus (2011)	14,92	18,12	12,57	22,65
Šifrar Kalan (2014b)	14,16	15,19	11,96	20,11
Hidalgo (2017b)	14,41	16,73	11,90	18,06
Del Barrio y Vann (2018)	10,04	10,44	8,12	14,94

Tabla 47. Estadísticos descriptivos de las investigaciones cotejadas.

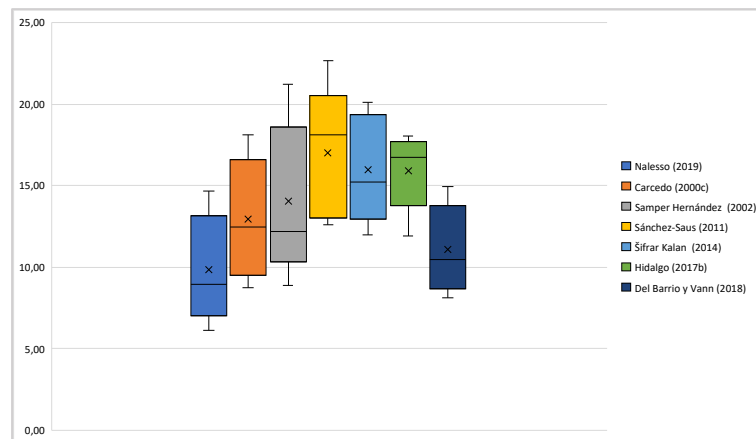


Gráfico 29. Diagrama de cajas de las investigaciones cotejadas.

¹⁰³ Con la misma finalidad, centrándose en la fonética, Arbulu Barturen (1998; 2000) coteja el sistema vocálico de un grupo de discentes de ELE, de diferentes nacionalidades, con el comportamiento del sistema vocálico nativo para (i) detectar las diferencias entre las interlenguas de los no nativos con el sistema español LM y (ii) averiguar si dichas interlenguas difieren en virtud de las lenguas maternas de los aprendices.

Samper Hernández (2002) revela el intervalo intercuartil más elevado, se pasa de un valor mínimo de 8,85 a un máximo de 21,21 y la desviación mayor se computa entre la mediana (12,21) y el Q3 (16), donde se coloca también el promedio de palabras por informante (12,59).

Se detecta una gran diferencia entre los cuartiles segundo y tercero en Carcedo (2000c), Šifrar Kalan (2014b), Del Barrio y Vann (2018), Nalesso (2019) ya que las cajas son más amplias con respecto a las que se posicionan por debajo. Además, en las investigaciones de Samper Hernández (2002), Carcedo (2000c), Del Barrio y Vann (2018) y Nalesso (2019) contamos con una simetría difusa en las cuatro partes del diagrama que no se encuentra en los otros estudios. Al contrario, en Sánchez-Saus (2011) e Hidalgo (2017b) la desviación máxima entre las respuestas está entre los intervalos Q2-Q1 y Q1-mínimo, indicado por el brazo inferior (respectivamente 11,96 y 11,90).

4.1.2 Promedio de vocablos

Estudiamos, ahora, los promedios de vocablos a partir de los siguientes valores.

CI	Nalesso (2019)	Carcedo (2000c)	Samper Hernández (2002)	Sánchez-Saus (2011)	Šifrar Kalan (2014b)	Hidalgo (2017b)	Del Barrio y Vann (2018)
CI01.CUE	0,91	2,40	3,35	1,71	1,45	1,21	1,07
CI02.ROP	0,72	2,26	3,07	1,55	1,93	1,09	1,50
CI03.CAS	0,47	2,06	3,14	2,22	2,85	1,90	0,93
CI05.ALI	1,75	5,44	7,92	2,90	3,28	2,00	2,94
CI10.CIU	1,57	4,50	7,64	3,44	3,43	3,13	2,47
CI12.TRA	0,58	2,00	2,64	2,08	1,97	1,99	1,25
CI14.ANI	1,02	3,38	3,28	1,79	2,13	1,40	1,66
CI15.JUE	1,31	3,68	6,00	3,28	2,87	2,48	2,97
CI16.PRO	1,62	4,76	6,42	3,38	2,63	2,38	2,40
PROMEDIO	1,11	3,39	4,83	2,48	2,50	1,95	1,91

Tabla 48. Promedios de vocablos de las investigaciones cotejadas.

Los campos más ricos son el CI05, “alimentos y bebidas”; el CI10, “la ciudad”; el CI16, “profesiones y oficios”; el CI15, “juegos y distracciones”. Los CI que tienen menor variación son el CI01, “partes del cuerpo”; el CI02, “la ropa”, y el CI12, “medios de transporte”.

Los aprendices que han conseguido aportar el mayor número de vocablos se encuentran en: Samper Hernández (2002) con 4,83; Carcedo (2000c) con 3,39; Šifrar Kalan (2014b) con 2,50; Sánchez-Saus (2011) con 2,48; Hidalgo (2017b) con 1,95; Del Barrio y Vann (2018) con 1,91; Nalesso (2019) con 1,11.

Se registra un cambio de tendencia, como ilustra el gráfico 30: el grupo en inmersión en Salamanca predomina en todas las áreas, excepto en el CI14, “los animales”, donde los encuestados de Carcedo (2000c) arrojan un valor levemente más alto. Los finlandeses se colocan en segunda posición, superando los resultados de las otras investigaciones que son, por orden, los estudiantes en inmersión en Andalucía, los eslovenos, los chinos, los italianos.

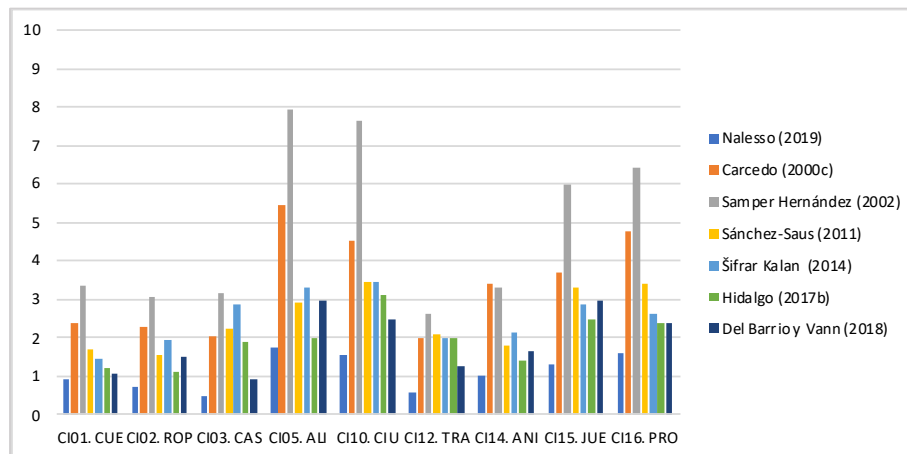


Gráfico 30. Promedios de vocablos por informante en las investigaciones cotejadas.

4.1.3 Índice de cohesión

La tabla permite analizar el grado de cohesión interna de cada centro de interés por cada grupo encuestado¹⁰⁴.

CI	Nalesso (2019)	Carcedo (2000c)	Samper Hernández (2002)	Sánchez-Saus (2011)	Šifrar Kalan (2014)	Hidalgo (2017b)	Del Barrio y Vann (2018)
CI01.CUE	0,15	0,27	0,34	0,09	0,13	0,05	0,20
CI02.ROP	0,11	0,22	0,28	0,06	0,06	0,04	0,10
CI03.CAS	0,13	0,18	0,20	0,07	0,07	0,03	0,13
CI05.ALI	0,08	0,14	0,19	0,06	0,06	0,03	0,07

¹⁰⁴ Hemos dejado fuera el estudio de la densidad léxica ya que es sensible al número de informantes.

CI10.CIU	0,08	0,11	0,19	0,04	0,05	0,02	0,07
CI12.TRA	0,14	0,21	0,31	0,05	0,06	0,02	0,11
CI14.ANI	0,11	0,18	0,27	0,06	0,07	0,04	0,10
CI15.JUE	0,06	0,09	0,14	0,04	0,05	0,02	0,04
CI16.PRO	0,06	0,11	0,14	0,03	0,05	0,02	0,06

Tabla 49. Índice de cohesión de las investigaciones cotejadas.

Merece la pena subrayar que el CI01, “partes del cuerpo”, tiene la mayor cohesión en todos los grupos, resulta que los conocimientos de los discentes son compactos, pese a sus procedencias, ya que el índice es cerrado. Esto podría deberse a que es fruto de un aprendizaje sistematizado: los alumnos han estudiado las palabras en clase. Parece que los campos que se presentan en el aula dan lugar a un conocimiento más compacto, tienen un grado de cohesión semántica más alto con respecto a los que no se trabajan, o que se trabajan menos. Asimismo, es posible que en ciertos centros los índices cambien debido a la realidad socio ambiental en la que viven los informantes (Gómez Molina y Gómez Devís, 2004: 83-84).

En el caso opuesto, los centros CI15, “juegos y distracciones”, y CI16, “profesiones y oficios”, son más difusos, quizás sea porque cada discente ha activado más asociaciones mentales en virtud de su experiencia personal.

Para terminar, los encuestados de Samper Hernández (2002) y Carcedo (2000c) han contestado con mayor cohesión frente a los otros aprendices, siguen los italianos y los eslovenos. Sánchez-Saus (2011) e Hidalgo (2017b) presentan índices que se alejan mucho del 1, con lo cual la producción se manifiesta más abierta y parece que cada sujeto ha actualizado unidades diferentes de sus propios compañeros.

4.2 Conclusiones del análisis comparativo

Los informantes en inmersión de Sánchez-Saus (2011) son los que han producido el número mayor de palabras, seguidos por los de Hidalgo (2017b) y Šifrar Kalan (2014b), hecho que no extraña ya que sus alumnos tienen un nivel de ELE mayor con respecto a los otros encuestados (B2+). Después, se colocan los aprendices de Samper Hernández (2002), que estudiaban también en una situación de inmersión, pero solo predominan en un CI, por tanto, pese a que atribuimos las causas de la gran producción detectada en Sánchez-Saus (2011) a la situación en la que estaban sus participantes, aquí no

constatamos la misma productividad. Se sitúan en los últimos puestos los finlandeses de Carcedo (2000c) seguidos por los dos grupos de italianos, en orden, Del Barrio y Vann (2018) y Nalesso (2019).

Si bien se han manifestado estas diferencias cuantitativas, en todas las investigaciones el CI más productivo es el CI05, “alimentos y bebidas”, mientras que los menos rentables son el CI12, “medios de transporte” y el CI02, “la ropa”.

Observamos una variación de tendencia en el análisis de la variedad léxica, ya que el mayor número de vocablos se ha aportado en Samper Hernández (2002), Carcedo (2000c), Sánchez-Saus (2011), Šifrar Kalan (2014b), Hidalgo (2017b), Del Barrio y Vann (2018), Nalesso (2019).

Igualmente, hay un cambio en la distribución de los campos: el más rico es el CI05, “alimentos y bebidas”, y, a continuación, se ubican el CI10, “la ciudad”; el CI16, “profesiones y oficios”; el CI15, “juegos y distracciones”. Los más pobres son el CI01, “partes del cuerpo”; el CI02, “la ropa” y el CI12, “medios de transporte”.

En lo que se refiere a la cohesión, Samper Hernández (2002) y Carcedo (2000c) presentan los valores más altos, seguidos por Del Barrio y Vann (2018), Nalesso (2019), Šifrar Kalan (2014b), Sánchez-Saus (2011) e Hidalgo (2017b).

Cabe recordar, antes de concluir, que los estudios cotejados no coinciden perfectamente en sus directrices metodológicas, por tanto, es posible que se haya dado un mayor desvío en los resultados, desde el momento en el que algunos investigadores han aplicado criterios más estrictos que otros.

De todas formas, creemos que el desnivel entre la cantidad de léxico disponible de los estudiantes encuestados depende de la atención que les otorgan los programas curriculares de cada universidad, del método didáctico aplicado y del contexto de aprendizaje (país de origen *versus* país hispanohablante). Puede que los trabajos realizados con informantes en inmersión ofrecen resultados superiores gracias al contexto hispanohablante en el que viven. Además, no sorprende que los resultados de los eslovenos puesto que su nivel de español es ligeramente más avanzado.

Destacan, después, los chinos que actualizan un número apreciable de palabras posiblemente porque las dos lenguas son muy lejanas y centrarse en el léxico de la LE estudiada permite la comprensión y la expresión de los discentes en la lengua meta que, al contrario, sería difícil sin una base léxica sólida.

En el caso opuesto, la afinidad lingüística entre LM y LE de los italianos, que presentan los resultados más escasos, puede deberse a que se descuida la enseñanza del vocabulario, quizá por la sensación de facilidad que causa la poca distancia tipológica. Efectivamente, en nuestra Universidad el profesorado confirma la mayor concentración de la docencia en la gramática y en la contrastividad entre italiano y español, en detrimento de otros elementos. Es más, no obstante la gran cantidad de horas dedicadas a clases de literatura y al análisis del texto (por ende, a la lectura en español) no resulta suficiente para el desarrollo de la competencia léxica de los alumnos que muestran un bagaje menor con respecto a los compañeros del mismo nivel.

5. CONCLUSIONES DEL ANÁLISIS CUANTITATIVO DE LA DISPONIBILIDAD LÉXICA

Hemos realizado el análisis de los datos cuantitativos procedentes de las pruebas de disponibilidad léxica suministradas a nuestros informantes y hemos cotejado los resultados con los de otros estudios de la misma índole, dedicados a estudiantes de ELE de otras nacionalidades.

En primer lugar, se han expuesto los datos generales obtenidos: se ha recogido un total de 12.579 palabras, lo que supone una media de 125,79 palabras por informante y de 7,86 palabras por CI. Al contabilizar las palabras diferentes, o vocablos, hemos conseguido un total de 1.490, en promedio 14,90.

Los centros de interés más productivos son: CI01, “partes del cuerpo”; CI05, “alimentos y bebidas”; CI10, “la ciudad”; CI14, “los animales”. Mientras que los menos productivos son CI06, “objetos colocados en la mesa para la comida”; CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”.

Por otro lado, los centros CI05, “alimentos y bebidas”; CI10, “la ciudad”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios” ofrecen la mayor cantidad de vocablos, por lo que se configuran como los más ricos. Al contrario, se han revelado como los más pobres el CI03, “partes de la casa (sin muebles)”; el CI04, “los muebles de la casa”; el CI06, “objetos colocados en la mesa para la comida”; el CI07, “la cocina y sus utensilios”; el CI09, “iluminación y calefacción”; el CI13, “trabajos del campo y del jardín”.

Los datos del índice de cohesión y de la densidad léxica destacan que los campos

semánticos más cerrados, donde la coincidencia en las respuestas de los informantes ha sido mayor, son CI01, “partes del cuerpo”; CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI12, “medios de transporte”; al opuesto, los más abiertos son CI11, “el campo”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios”.

En el siguiente apartado hemos realizado el análisis por variable de los datos, finalizada a comprobar la influencia de los factores sociolingüísticos considerados sobre el léxico disponible.

- (i) La variable sexo destaca que las mujeres obtienen resultados levemente mayores en los promedios de palabras (+9,61%) mientras que los hombres sobrepasan en el conocimiento de los vocablos, rebasan las compañeras del 66,35%.
- (ii) La variable nivel de ELE provoca una variación en los resultados, se calcula un aumento del 7,51% más de palabras y del 14,05% de vocablos a favor de los informantes de nivel más avanzado.
- (iii) La variable conocimiento de otras LE evidencia que los participantes en nuestro estudio que conocen más de dos LE escriben más palabras (+10,24%), pero los que saben solo dos LE conocen significativamente más vocablos (+50,84%).

En segundo lugar, a la luz de los resultados procedentes de la primera y de la segunda aplicación de la encuesta a los estudiantes de nivel B2 hemos realizado el análisis longitudinal. Resulta que la frecuencia a un curso académico, como esperábamos, determina una importancia crucial en el desarrollo de la disponibilidad léxica de los aprendices, ya que al final del año los informantes han mejorado del 30,29% su productividad, pasando de 6.517 palabras a un total de 8.491. Al desglosar los datos en función de cada centro de interés, observamos una superioridad que va del 5% en el CI01, “partes del cuerpo”, al 100% en el CI09, “iluminación y calefacción”.

Igualmente, los datos de la segunda prueba subrayan una superior riqueza léxica desde el momento en el que registramos un aumento de 83 *types* actualizados (en media 1,66) que corresponde a +7,50%.

Asimismo, la concreción interna de los CI es mayor en los resultados de la segunda aplicación, donde las respuestas coinciden en gran medida: el conjunto de palabras es más cohesionado (+23,75%). El grupo ha ampliado sus conocimientos y parece que una vez que se llega a un nivel lingüístico más avanzado se desarrollan más asociaciones mentales y el conocimiento del vocabulario es más articulado.

En tercer lugar, a la hora de correlacionar los resultados de diferentes investigaciones sobre la disponibilidad léxica en ELE, hemos estudiado los datos en función de la nacionalidad de los informantes y el país en el que estudian español. Los aprendientes en inmersión en Andalucía son los que obtienen índices más altos, seguidos por los chinos, los eslovenos, los en inmersión en Salamanca, los finlandeses y, por último, los dos grupos de italianos. No sorprende que los sujetos analizados en Šifrar Kalan (2014b) alcancen valores superiores dado que tienen desarrollado un nivel más avanzado de ELE, como no extrañan los resultados de Samper Hernández (2002) y Sánchez-Saus (2011) gracias al contexto en inmersión.

En lo referente a los datos estadísticos, la desviación mayor se ha manifestado en Samper Hernández (2002). En Carcedo (2000c), Šifrar Kalan (2014b), Del Barrio y Vann (2018), Nalesso (2019) se aprecia una simetría más difusa; mientras que en Sánchez-Saus (2011) e Hidalgo (2017b) se detecta una dispersión mayor.

Por último, Samper Hernández (2002) y Carcedo (2000c) muestran la mayor cohesión interna, seguidos por Del Barrio y Vann (2018), Nalesso (2019), Šifrar Kalan (2014b), Sánchez-Saus (2011) e Hidalgo (2017b).

Sin embargo, una vez que dejamos de un lado los valores en si mismos, los informantes se comportan de manera muy semejante ya que arrojan más palabras en el CI05, “alimentos y bebidas”, y menos en el CI12, “medios de transporte” y el CI02, “la ropa”. El mayor número de vocablos se registra en el CI05, “alimentos y bebidas”; el CI10, “la ciudad”; el CI16, “profesiones y oficios”; el CI15, “juegos y distracciones”, y el menor en el CI01, “partes del cuerpo”; el CI02, “la ropa”, y el CI12, “medios de transporte”.

Suponemos que el desnivel en el léxico disponible de las distintas agrupaciones sea debido a los contenidos de los programas curriculares, el método didáctico aplicado y el contexto de aprendizaje que dedican al vocabulario una atención diferente. Todo esto confirmaría una vez más la importancia de plantear actividades explícitas para la adquisición del léxico, porque el aprendizaje incidental del vocabulario no es suficiente:

El aprendizaje intencionado supone un aprendizaje consciente por parte del aprendiz y suele sucederse de actividades concretas encaminadas a informar a los alumnos sobre los elementos léxicos y sus características formales, semánticas, sintácticas, etc., y a la práctica de los mismos. [...] El aprendizaje intencionado de vocabulario es muy efectivo en términos relativos, es decir da lugar a altos grados de adquisición de palabras [...]. (Agustín Llach, 2013)

CAPÍTULO V.

ANÁLISIS CUALITATIVO DE LA DISPONIBILIDAD LÉXICA

1. INTRODUCCIÓN

Abordamos ahora el análisis cualitativo de nuestro corpus desde las perspectivas transversal (análisis general y por variable) y longitudinal. Examinamos el tipo de léxico disponible activado en los distintos centros de interés a partir del estudio de los índices de disponibilidad léxica presentando los vocablos más disponibles, que alcanzan un ID igual o superior a 0,1.

Todo esto permite averiguar cuáles son las temáticas predominantes en cada campo y qué categorías gramaticales son las más difundidas en las respuestas. No seremos exhaustivos, pues esto supondría una clasificación sin aplicar una medida de corte.

Asimismo, computamos la *cardinalidad del conjunto* (López Chávez, 1992) o *índice de compatibilidad* propuesto por Ávila Muñoz y Sánchez Sáez (2010; 2011).

Por último, el análisis comparativo coteja nuestra muestra con dos corpus de referencia. Contrastamos los datos con la lista de las primeras 5.000 palabras¹⁰⁵ más frecuentes del español procedentes del *Corpus de referencia del español actual* (CREA) con el propósito de verificar si las aportaciones de los informantes están incluidas en este banco de datos. En segunda instancia, comparamos nuestra lista con el *Corpus de aprendices del español* (CAES) para descubrir si coincide con las palabras más conocidas por otros aprendices de ELE.

2. ANÁLISIS TRANSVERSAL

2.1 Resultados generales

El análisis transversal se abre con el estudio de los listados de léxico más disponible por centro de interés.

Las tablas muestran (i) cuántos vocablos superan o igualan el ID del 0,1; (ii) cuáles

¹⁰⁵ Con esta cantidad de palabras una persona puede desenvolverse en situaciones comunicativas cotidianas (Alvar Ezquerro, 2003: 100).

son las temáticas y las clases gramaticales predominantes; (iii) la eventual presencia de locuciones o colocaciones léxicas.

2.1.1 CI01, “partes del cuerpo”

Rango	Vocablo
1	ojos
2	mano
3	cabeza
4	pie
5	pierna
6	brazo
7	boca
8	dedo
9	nariz
10	pelo
11	cara
12	espalda
13	cuello
14	orejas
15	rodilla
16	dientes
17	lengua
18	barriga
19	uñas

Tabla 50. Vocablos más disponibles en el CI01.

Las palabras se adscriben al estímulo prototípico nocional y, en mayor medida, se refieren a la parte superior del cuerpo humano (*ojos, cabeza, boca, nariz, pelo, cara, cuello, orejas, dientes, lengua*).

Son todos sustantivos que describen las partes del cuerpo (71%), pero si indagamos el total de vocablos detectamos otras categorías gramaticales y temáticas ligadas a un estímulo secundario: 11% de adjetivos relacionados con el aspecto físico (*rubio, azul, pelirrojo, guapo, delgado*); 5% de verbos y sentidos (*oír, oído, ver, vista, respirar, tacto*); 12% de otros sustantivos referentes a la biología (*sangre, neurona, aparato*).

La presencia de unidades multipalabra se limita a una, *cuerpo humano*.

2.1.2 CI02, “la ropa”

Rango	Vocablo
1	camiseta
2	zapatos
3	pantalones
4	camisa
5	jersey
6	chaqueta
7	falda
8	vaqueros
9	bufanda
10	calcetines
11	vestido
12	abrigo

Tabla 51. Vocablos más disponibles en el CI02.

Contamos con sustantivos referidos a las prendas de vestir y a los complementos (*bufanda, calcetines*), que se encuentran en gran cantidad a lo largo de todo el listado (*anillo, pendientes, pulsera, gafas, bolso, mochila*).

En los rangos por debajo del ID del 0,1 detectamos un porcentaje escaso de verbos (*ponerse, vestirse*) y otros sustantivos adscritos a diferentes campos semánticos, como los materiales (*seda, algodón*).

Los encuestados han aportado algunas locuciones sustantivas: *gafas de sol, lente de contacto, ropa interior, traje de baño, zapatilla de deporte*.

2.1.3 CI03, “partes de la casa (sin muebles)”

Rango	Vocablo
1	cocina
2	cuarto-de-baño
3	habitación
4	salón
5	jardín
6	cuarto
7	garaje

Tabla 52. Vocablos más disponibles en el CI03.

En esta área temática los vocablos más disponibles son siete (que cubren el 5% del entero listado). De nuevo se trata de sustantivos y una locución, *cuarto de baño*, a la que se añaden otras, aportadas con menor frecuencia (*cuarto de estar*, *cuarto trasero*, *cuarto matrimonial*, *salón de baile*).

Al examinar la lista por entero, vemos que el sustantivo es la única categoría gramatical para las nociones referidas a las habitaciones y a las partes estructurales de la casa (*pared*, *muro*, *techo*, *puerta*, *portal*, *ventana*, *pasillo*, *escalera*, *corredor*).

2.1.4 CI04, “los muebles de la casa”

Rango	Vocablo
1	cama
2	silla
3	mesa
4	sofá
5	armario
6	sillón
7	ducha
8	lámpara
9	tele(visión)
10	mesilla-de-noche

Tabla 53. Vocablos más disponibles en el CI04.

Son diez los ítems que en este centro de interés superan o igualan el ID preestablecido, entre ellos constatamos una unidad pluriverbal: *mesilla de noche*, la única que hemos encontrado en este listado.

De nuevo, todas las actualizaciones son sustantivos vinculados a temáticas distintas ya que hemos concedido una mayor amplitud de las asociaciones. Además del estímulo primario, los informantes han escrito palabras que indican electrodomésticos, complementos y decoraciones (*televisión*, *lavadora*, *horno*, *frigorífico*, *lavavajillas*, *cocina*, *refrigerador*, *cuadro*, *alfombra*, *cortina*, *almohada*).

2.1.5 CI05, “alimentos y bebidas”

Rango	Vocablo
1	agua
2	cerveza

3	tomate
4	leche
5	manzana
6	naranja
7	vino
8	paella
9	jamón
10	pescado
11	pan
12	arroz
13	carne
14	zumos
15	pasta
16	café
17	bocadillo
18	azúcar
19	huevo
20	queso
21	patata
22	chocolate
23	sangría
24	plátano
25	sal

Tabla 54. Vocablos más disponibles en el CI05.

La tabla muestra que este CI es el que cuenta con más *types* que tienen un ID \geq 0,1 entre los cuales encontramos vocablos relacionados con la comida y la bebida que hacen lo propio bajo la forma de sustantivos (93%). Notamos productos típicos de la cocina española (*paella, jamón, sangría*), que han sido aportados también en posiciones más bajas (*tortilla, tapas, calimochos, gazpacho*) juntos a características de Hispanoamérica (*aguacate, palta, papa, quesadilla, frijoles*).

Aparecen también técnicas de cocción (*asado, flambeado, parrilla*), condimentos y aliños (*sal, azúcar, aceite, vinagre, orégano, azafrán*).

Extraña el hecho de que en un campo como este no aparezca ningún verbo, mientras que se han actualizado adjetivos relativos a los sabores (*dulce, salado, amargo*) y al tipo de alimento (*animal, vegetal*).

Hemos aceptado *Coca-Cola*, ya que se trata de un ejemplo prototípico de lexicalización de una marca comercial que designa un tipo de refresco, ya no es un mero

distintivo comercial (Hidalgo, 2017b: 335).

Se pone de relieve la aparición de varias locuciones (7%): *arroz con leche, crema catalana, piña colada, dulce de leche, agua sin gas, agua con gas, vino tinto, vino blanco, tinto de verano, zumo de naranja, zumo de fruta, pechuga de pollo*.

2.1.6 CI06, “objetos colocados en la mesa para la comida”

Rango	Vocablo
1	plato
2	vaso
3	cuchillo
4	cuchara
5	botella
6	tenedor

Tabla 55. Vocablos más disponibles en el CI06.

Los vocablos más disponibles son sustantivos que designan los objetos más comunes que se utilizan para la comida, como requería el estímulo.

El 91% está formado por nombres de comida y bebida (*agua, alimento, comida*), utensilios (*mantel, olla, botella, bandeja, copa, tenedor, servilleta, jarra*), condimentos (*aceite, sal, pimienta, vinagre*) y otras lexías (*desayuno, mesa, comedor*). El restante 9% se compone de verbos que denotan acciones relacionadas al estímulo (*cambiar, poner, tomar, pulir*).

En total, contamos con cuatro colocaciones: *plato hondo, plato llano, vino tinto, vino blanco*.

2.1.7 CI07, “la cocina y sus utensilios”

Rango	Vocablo
1	horno
2	plato
3	cuchillo
4	cuchara
5	sartén
6	vaso
7	frigorífico
8	nevera

Tabla 56. Vocablos más disponibles en el CI07.

Los lexemas más disponibles para este campo nocional son ocho, todos sustantivos. Aparecen electrodomésticos y utensilios, como en el CI06, “objetos colocados en la mesa para la comida”, a lo largo de toda la lista. En la opinión de Hidalgo (2017b: 521) se trataría de

[...] *aprovechamiento léxico*, es decir, los encuestados se valen de sus respuestas en centros de interés anteriores para completar el centro sucesivo, de manera que se reproducen secuencias léxicas con unidades idénticas e, incluso, en el mismo orden en distintos campos nocionales. [...] en la mayoría de las ocasiones esta repetición se produce cuando los informantes no saben qué otras unidades aportar, ya sea porque apenas conocen vocablos o porque ya han registrado todos los que conocían.

Señalamos la presencia de unidades que hemos aceptado pese a que no encajen exactamente en el estímulo nocional, debido a la mayor amplitud de relaciones asociativas concedida: *comedor, cubierto, comida*.

Los participantes han aportado la locución *libro de cocina* y las unidades compuestas *lavaplatos* y *lavavajillas*.

Los sustantivos cubren el 92% de la lista frente al 8% de verbos (*abrir, cocinar, enharinar, apagar, mezclar*), colocados en distintos rangos.

2.1.8 CI08, “la escuela: muebles y materiales”

Rango	Vocablo
1	libro
2	mesa
3	silla
4	bolí(grafo)
5	lápiz
6	cuaderno
7	pizarra
8	mochila
9	ordenador
10	papel
11	borrador

Tabla 57. Vocablos más disponibles en el CI08.

Estas unidades se adscriben principalmente al segundo estímulo semántico, los materiales escolares (*libro, bolígrafo, lápiz, cuaderno, mochila, ordenador, papel, borrador*), mientras que las demás responden al primero, los muebles (*mesa, silla, pizarra*).

Siguiendo la misma tendencia, el 7% se compone de verbos (*escribir, ir, pinchar, subrayar, volver*) y el 93% de sustantivos. Si desglosamos este último porcentaje contamos con la siguiente distribución de temáticas: 68% de materiales (entre otros, *goma de borrar, marcador, proyector, registro, sacapuntas, papel, mapa*), 15% de muebles (*armario, cátedra, mesa, perchero, pupitre, reloj, silla*), 7% de personas (*alumno, clase, compañero, estudiante, profesor*), 2% misceláneo (*aula, luz, edificio*).

2.1.9 CI09, “iluminación y calefacción”

Rango	Vocablo
1	lámpara
2	luz
3	fuego

Tabla 58. Vocablos más disponibles en el CI09.

En este campo semántico, las lexías con $ID \geq 0,1$ son tres y presentan índices de frecuencia más elevados con respecto a la que se posiciona después, *sol*, que alcanza 12 apariciones mientras que estas logran 80, 59, 39 ocurrencias respectivamente.

En general, se trata de sustantivos (91%) que responden a los dos estímulos del CI. En la mayoría de los casos, son nombres que pertenecen al ámbito de la calefacción (*calentador, termosifón, temperatura, calorífico, calefacción, chimenea, estufa, ventilador, leña*). Los únicos adjetivos detectados, *caliente* y *frío*, cubren el 4%. Tampoco los verbos son muchos ya que forman el 5% de la lista: *calentar, apagar, encender*.

Detectamos también palabras relacionadas con otras temáticas (*electricidad, pantalla, noche, estrella, circuito, enchufe*) que hemos admitido debido a la dificultad que, a nuestro juicio, conllevaba este centro para los aprendices.

El análisis ha revelado la actualización de algunas unidades multipalabra: *aire acondicionado, energía eléctrica, energía renovable, hacer frío, hacer calor, central nuclear*.

2.1.10 CI10, “la ciudad”

Rango	Vocablo
1	calle
2	coche
3	tienda

4	parque
5	escuela
6	casa
7	edificio
8	(auto)bús
9	iglesia
10	plaza
11	estación
12	bar
13	carretera
14	restaurante
15	tren
16	ayuntamiento
17	árbol

Tabla 59. Vocablos más disponibles en el CI10.

Este centro de interés es uno de los más ricos en la producción de vocablos (157), los más disponibles son diecisiete y comprenden nombres de edificios y construcciones de la ciudad (*escuela, casa, edificio, iglesia, plaza, estación, ayuntamiento, árbol*), medios de transporte (*coche, autobús, tren*), servicios (*tienda, parque, bar, restaurante*), vías públicas (*calle, carretera*).

En la lista aparecen sinónimos, hiperónimos, hipónimos, merónimos y holónimos de estas lexías, entre otros: *museo, ayuntamiento, rascacielos, tranvía, metro, farmacia, gimnasio, universidad, banco, correos, avenida, cruce, esquina, acera*.

Aún más, se han escrito sustantivos referidos a los habitantes (*gente, persona, peatón, alcalde, turista, pueblo, ciudadano*), la diversión (*movida, cine, teatro, discoteca, tapas*), las tiendas y los restaurantes (*peluquería, panadería, carnicería, heladería, cafetería, estanco, joyería, agencia, tabaquería, frutería, florería, café, floristería, pizzería*), el trabajo (*oficina, negocio, oficio, trabajo, despacho, fábrica*), los problemas de la ciudad (*esmog, tráfico, ruido*) y una miscelánea de otros vocablos (*centro, policía, fuente, río, aeropuerto, derecha, cartel, muralla, lampión, playa, rincón, sitio, luz, arte, billete, flor*).

En lo que atañe a la variación diatópica del español sobresale la palabra *carro*, típica de la variedad hispanoamericana.

En la totalidad de los vocablos, sin tener en cuenta las locuciones, el 9% está formado por verbos (*pasear, cruzar, aparcar, girar, comer, vivir, viajar, beber, salir*,

comprar, bajar, vender, caminar) y el 3% por adjetivos (*verde, cosmopolita, antiguo, moderno, histórico*).

Las combinaciones léxicas abarcan el 6% de las aportaciones: *casco antiguo, transporte público, centro comercial, bloque de piso, señal de tráfico, centro histórico, ir de compras, zona peatonal, oficina de correos*.

2.1.11 CIII, “el campo”

Rango	Vocablo
1	árbol
2	animal
3	flor
4	tierra
5	hierba
6	vaca
7	campesino
8	planta
9	caballo

Tabla 60. Vocablos más disponibles en el CIII.

Todas las unidades más disponibles son sustantivos, perfectamente encajados en el ámbito semántico. Tenemos pocos verbos (*pasear, cultivar, andar, correr, jugar, trabajar*) y aún menos adjetivos (*verde, fértil*).

Los informantes han asociado este estímulo con las personas que viven en el campo y sus oficios (*campesino, gaucho, cultivador, agricultor, granjero*), los lugares característicos (*molino, río, campo, granja, huerto, aldea, arboreto*), los animales (*vaca, caballo, gallina, pato, conejo, toro, pavo*), los productos (*árbol, flor, planta, verdura, harina, vino, aceite*), los medios de transporte (*tractor, bicicleta*), el clima y el tiempo atmosférico (*viento, sol, niebla*).

Se han encontrado relaciones secundarias vinculadas con sensaciones (*tranquilidad, silencio, paz, libertad*), entre las cuales aparece la locución *al aire libre*.

2.1.12 CII2, “medios de transporte”

Rango	Vocablo
1	coche
2	(auto)bús

3	tren
4	bici(cleta)
5	avión
6	taxi
7	metro
8	moto(cicleta)
9	barco
10	tranvía

Tabla 61. Vocablos más disponibles en el CI12.

Constatamos diez elementos estrictamente ligados a los medios de transporte, tal cual se observa en todo el listado (*camión, nave, automóvil, autobús, tándem, velero, funicular, submarino*) a los que se añaden otras formas de desplazamiento (*a pie, caballo, patinetes, a nado, correr, caminar, andar*).

Entre las respuestas aparecen títulos de viaje (*billete*), documentos (*carne de conducir*) y otras palabras relacionadas con el tráfico (*esmog, horario, parada, gasolina*).

El porcentaje de sustantivos arrojados es del 81%, dentro del cual hay también americanismos: *carro, colectivo, guagua, máquina*. Los verbos cubren el 9%: *conducir, tomar, coger, correr, subir, esperar, viajar, caminar, ir, andar*.

Hemos encontrado la sigla *AVE* y algunas locuciones como *a pie, a nado, transporte público, carne de conducir, hacer dedo, globo aerostático*.

2.1.13 CI13, “trabajos del campo y del jardín”

Rango	Vocablo
1	jardinero
2	campesino
3	plantar
4	cortar

Tabla 62. Vocablos más disponibles en el CI13.

Los *types* que se revelan más disponibles para este campo nocional son cuatro y, por primera vez, aparecen verbos: *plantar* y *cortar*. Los demás vocablos alternan la presencia de estas dos categorías gramaticales a lo largo de la lista, a pesar del predominio de los sustantivos.

La mayoría de los lexemas pertenece al ámbito prototípico del centro de interés (*cultivar, agricultor, regar, granjero, pastor, arar, recoger, cosechar, gaucho*), las otras

temáticas abarcan los animales asociados con el campo (*gato, perro, vaca, oveja, burro, mariposa*), productos que derivan de los cultivos y de otras actividades (*flor, fruta, verdura, mantequilla, aceite*), herramientas (*tijeras, arado*), lugares (*campo, jardín*).

Dar agua es la única combinación que hemos encontrado.

2.1.14 CII4, “los animales”

Rango	Vocablo
1	perro
2	gato
3	caballo
4	vaca
5	pez
6	pájaro
7	león
8	cerdo
9	gallina
10	tigre
11	conejo
12	toro

Tabla 63. Vocablos más disponibles en el CII4.

Como era esperable en este campo nocional los doce lexemas más disponibles son exclusivamente sustantivos referidos a clases de animales. Constatamos dos colocaciones, *caballito de mar* y *estrella marina*.

Sorprende el hecho de que no haya surgido ninguna otra asociación, como por ejemplo podrían haber sido sus partes del cuerpo, los lugares donde viven, los productos de origen animal.

Por último, notamos los siguientes americanismos: *tigrillo, hicotea, chancho, cochino, alpaca*.

2.1.15 CII5, “juegos y distracciones”

Rango	Vocablo
1	fútbol
2	baloncesto
3	leer
4	libro

5	bailar
6	cartas
7	pelota
8	tele(visión)
9	ordenador
10	videojuegos
11	cine
12	escuchar- música
13	natación

Tabla 64. Vocablos más disponibles en el CI15.

En el conjunto total de vocablos arrojados en esta área, son trece los que revelan un $ID \geq 0,1$, que corresponde con el 10% de la lista.

Por lo general, se presenta una gran variedad temática y tipológica de juegos y diversiones, que incluyen juguetes, actividades deportivas, lugares en los que se pueden realizar dichas acciones, personas y animales de compañía.

La mayoría son sustantivos, pero hay un 27% de verbos (*correr, nadar, pasear, cantar, dibujar, salir, viajar, descansar, esquiar, chatear, escribir, patinar, fotografiar*), una cantidad que destaca con respecto a la contabilizada en los demás CI, entre los que constatamos una combinación léxica (*escuchar música*).

Además, las unidades multipalabra alcanzan un porcentaje del 11%: *juego de mesa, parque de diversiones, instrumento musical, dibujos animados, soldadito de plomo, redes sociales, escuchar música, salir con amigos, ir de compras, quedar con amigos, pasárselo bien, ir de vacaciones, ir de tapas, ir de bares*.

Es oportuno señalar la presencia de los extranjerismos *Monopoly* y *Play Station*, que hemos aceptado al tratarse de la lexicalización de una marca comercial.

2.1.16 CI16, “profesiones y oficios”

Rango	Vocablo
1	profesor
2	médico
3	cocinero
4	abogado
5	doctor
6	camarero

7	fontanero
8	actor
9	estudiante
10	empleado

Tabla 65. Vocablos más disponibles en el CI16.

Todas las lexías más disponibles son sustantivos, como era de esperar.

En cuanto al resto de unidades de la lista y su categoría gramatical, se observa un 98% de nombres comunes y un 2% de verbos, hay un solo adjetivo muy en línea con el estímulo del campo, *trabajador*.

Los nombres se reparten en dos macroáreas temáticas: la primera (94%) agrupa profesiones y profesionales (*enfermero, peluquero, traductor, canguero, escritor, jefe, periodista, albañil, sastre, cirujano, empresario, estilista, logopeda*), que hemos recopilado en la forma masculina singular para dar uniformidad a la muestra, pero manteniendo los siguientes heterónimos: *actor, actriz, azafata, preste, cura, monja*; la segunda (4%) reúne lugares y tipos de trabajo (*oficina, banco, trabajo, tienda, estudio, escuela, fábrica, negocios, restaurante, hospital*).

Los verbos están igualmente en relación con el centro de interés ya que se refieren a acciones ligadas a los empleos mencionados (*enseñar, bailar, ordenar, cocinar*).

Constatamos la actualización de americanismos (*plomero, mucamo, botón*) y de algunas locuciones (*guía de turismo, ama de casa, señora de la limpieza, cuidador de perro*).

2.2 Conclusiones parciales

En primera instancia, la categoría gramatical sobresaliente es la de los sustantivos. Los verbos aparecen, pero en porcentajes menores. El más alto (27%) se detecta en el CI15, “juegos y distracciones”. Los adjetivos aparecen aún menos y no hemos encontrado ninguna unidad perteneciente a otras clases morfológicas.

En segundo lugar, los vocablos más disponibles que han sido activados se adscriben a los ámbitos prototípicos de cada campo, solo en las posiciones más bajas (Anexo III) y en particular en los centros CI04, “los muebles de la casa”; CI11, “el campo”; CI12, “medios de transporte”; CI13, “trabajos del campo y del jardín”; CI15, “juegos y distracciones”; CI16, “profesiones y oficios”, los informantes han escrito lexías

referentes a estímulos secundarios.

La tabla resume las cantidades de vocablos con $ID \geq 0,1$ según centro de interés.

CI	Vocablos con $ID \geq 0,1$
CI01. CUE	19
CI02. ROP	12
CI03. CAS	7
CI04. MUE	10
CI05. ALI	25
CI06. MES	6
CI07. COC	8
CI08. ESC	11
CI09. ILU	3
CI10. CIU	17
CI11. CAM	9
CI12. TRA	10
CI13. TRC	4
CI14. ANI	12
CI15. JUE	13
CI16. PRO	10

Tabla 66. Número de vocablos más disponibles por CI.

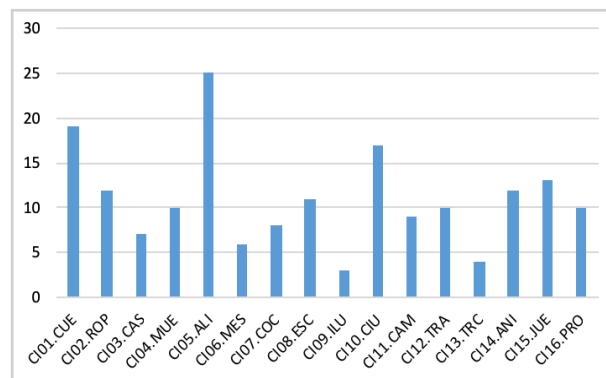


Gráfico 31. Número de vocablos más disponibles por CI.

La oscilación va de un mínimo de 3 a un máximo de 25 vocablos. En los extremos se encuentran los centros que tienen más palabras, el CI05, “alimentos y bebidas”; el CI01, “partes del cuerpo”, y el CI10, “la ciudad”, mientras que los que presentan un número menor son el CI09, “iluminación y calefacción”, y el CI13, “trabajos del campo

y del jardín”. No extraña que estos, al mismo tiempo, se hayan manifestado como los más y menos productivos en el análisis cuantitativo.

2.3 Resultados por variable

Realizamos el análisis de los resultados por variable según el mismo protocolo de la sección previa, añadiendo el estudio de la *cardinalidad del conjunto* que permite establecer los porcentajes de compatibilidad de las respuestas.

Las lexías marcadas en gris son compartidas por las dos variantes mientras que las que no están señaladas no aparecen entre las más disponibles de una de las dos (aunque aparezcan en posiciones más bajas de los registros, no alcanzan el $ID \geq 0,1$).

2.3.1 Variable “sexo”

2.3.1.1 CI01, “partes del cuerpo”

Rango	Mujer	Hombre
1	ojos	mano
2	mano	cabeza
3	cabeza	ojos
4	pierna	pie
5	pie	boca
6	brazo	pelo
7	boca	brazo
8	dedo	dedo
9	nariz	rodilla
10	pelo	cara
11	cara	pierna
12	espalda	nariz
13	orejas	espalda
14	cuello	cejas
15	rodilla	
16	dientes	
17	lengua	
18	barriga	
19	uñas	

Tabla 67. Vocablos más disponibles del CI01 según la variable “sexo”.

Los lexemas más disponibles compartidos por los dos sociolectos son trece, las

mujeres han arrojado seis unidades que los hombres no han escrito o que en su lista no alcanzan el ID requerido (orejas, cuello, dientes, lengua, barriga, uñas). Los chicos presentan una sola palabra que sus compañeras no presentan (cejas). De este modo la cardinalidad computada es del 65%, vista la gran coincidencia en los rangos superiores.

2.3.1.2 CI02, “la ropa”

Rango	Mujer	Hombre
1	camiseta	camiseta
2	zapatos	zapatos
3	pantalones	camisa
4	falda	pantalones
5	jersey	calcetines
6	camisa	vaqueros
7	chaqueta	bufanda
8	vaqueros	chaqueta
9	bufanda	jersey
10	abrigo	
11	calcetines	
12	vestido	
13	traje	

Tabla 68. Vocablos más disponibles del CI02 según la variable “sexo”.

La tabla indica que se ha producido una coincidencia de nueve ítems en las listas de palabras con ID mayor o igual a 0,1. En particular, las primeras dos posiciones (*camiseta, zapatos*) coinciden. Las mujeres tienen una lista más larga, de trece vocablos (añaden *falda, abrigo, vestido, traje*). A partir de estos valores, hemos obtenido un porcentaje de compatibilidad del 69,23%.

2.3.1.3 CI03, “partes de la casa (sin muebles)”

Rango	Mujer	Hombre
1	cocina	cuarto-de-baño
2	cuarto-de-baño	cocina
3	habitación	jardín
4	salón	habitación
5	jardín	cuarto

6	cuarto	salón
7	garaje	escalera
8		garaje
9		puerta

Tabla 69. Vocablos más disponibles del CI03 según la variable “sexo”.

Todas las palabras compartidas se corresponden con las lexías aportadas por las mujeres (*cocina, cuarto de baño, habitación, salón, jardín, cuarto, garaje*) ya que estas unidades se encuentran en la lista de hombres, donde hallamos dos unidades exclusivas (*escalera, puerta*). Por tanto, la cardinalidad es del 77,78%.

2.3.1.4 CI04, “los muebles de la casa”

Rango	Mujer	Hombre
1	cama	cama
2	silla	armario
3	mesa	sofá
4	sofá	silla
5	armario	mesa
6	sillón	
7	ducha	
8	lámpara	
9	tele(visión)	
10	mesilla-de-noche	

Tabla 70. Vocablos más disponibles del CI04 según la variable “sexo”.

El 50% de cardinalidad que hemos computado deriva del hecho de que los vocablos comunes son cinco y el grupo de las chicas ha aportado cinco unidades más que sus compañeros (*sillón, ducha, lámpara, televisión, mesilla de noche*).

2.3.1.5 CI05, “alimentos y bebidas”

Rango	Mujer	Hombre
1	agua	agua
2	tomate	cerveza
3	manzana	vino
4	cerveza	naranja
5	leche	paella
6	jamón	pan

7	naranja	carne
8	paella	pescado
9	vino	tomate
10	pescado	leche
11	pan	jamón
12	arroz	limón
13	zumos	arroz
14	pasta	manzana
15	carne	<i>Coca-Cola</i>
16	café	sangría
17	azúcar	pasta
18	huevo	plátano
19	bocadillo	
20	queso	
21	chocolate	
22	patata	
23	sal	
24	tapas	
25	plátano	
26	sangría	
27	aceite	

Tabla 71. Vocablos más disponibles del CI05 según la variable “sexo”.

Se revela un desnivel notable entre el número de respuestas de los dos conjuntos de participantes: un total de treinta siete para las mujeres y dieciocho para los hombres. Las chicas han actualizado once vocablos que los chicos no tienen en este registro (*zumos, café, azúcar, huevo, bocadillo, queso, chocolate, patata, sal, tapas, aceite*). Ellos, por su parte, han arrojado dos divergentes (*limón, Coca-Cola*).

Estos datos suponen un 55,17% de cardinalidad entre las respuestas que manifiestan la presencia exclusiva de sustantivos.

2.3.1.6 CI06, “objetos colocados en la mesa para la comida”

Rango	Mujer	Hombre
1	plato	plato
2	vaso	cuchillo
3	cuchillo	vaso
4	cuchara	cuchara

5	botella	tenedor
6	tenedor	botella
7	cucharilla	
8	servilleta	
9	sal	
10	aceite	
11	mantel	
12	copa	
13	taza	
14	jarra	
15	ensaladera	
16	agua	
17	sartén	
18	plato- llano	
19	bebida	
20	plato- hondo	
21	cucharada	

Tabla 72. Vocablos más disponibles del CI06 según la variable “sexo”.

La compatibilidad es inferior con respecto a lo visto hasta ahora, se trata del 28,57%, pese a la gran correspondencia en las primeras posiciones de los vocablos. En efecto, las mujeres sobrepasan a los varones en quince elementos: las lexías compartidas son seis, esto es, la entera lista masculina.

Si excluimos *cucharada*, que indica una porción, las demás unidades son exclusivamente nombres que describen objetos que se utilizan para comer.

2.3.1.7 CI07, “la cocina y sus utensilios”

Rango	Mujer	Hombre
1	horno	horno
2	plato	refrigerador
3	cuchillo	nevera
4	cuchara	olla
5	vaso	plato
6	frigorífico	
7	sartén	
8	nevera	

Tabla 73. Vocablos más disponibles del CI07 según la variable “sexo”.

Entre mujeres y hombres hay tres coincidencias. El grupo femenino presenta cinco palabras no presentes en el masculino (*cuchillo, cuchara, vaso, frigorífico, sartén*), mientras que los varones presentan dos unidades que las mujeres no han activado entre las más disponibles (*refrigerador, olla*). Por esta razón, la cardinalidad del conjunto es del 30%.

2.3.1.8 CI08, “la escuela: muebles y materiales”

Rango	Mujer	Hombre
1	mesa	libro
2	silla	lápiz
3	libro	cuaderno
4	bolí(grafo)	bolí(grafo)
5	cuaderno	silla
6	lápiz	mochila
7	pizarra	pizarra
8	ordenador	
9	mochila	
10	papel	
11	borrador	

Tabla 74. Vocablos más disponibles del CI08 según la variable “sexo”.

La tabla pone de relieve una coincidencia de catorce vocablos, todas las lexías presentes en la lista de los hombres también están en la de las mujeres, que por su parte tienen cuatro vocablos divergentes (*mesa, ordenador, papel, borrador*). Esto lleva a una compatibilidad del 63,64%.

2.3.1.9 CI09, “iluminación y calefacción”

Rango	Mujer	Hombre
1		lámpara
2		luz

Tabla 75. Vocablos más disponibles del CI09 según la variable “sexo”.

Aquí no hay correspondencia entre los vocablos ya que los hombres han aportado en su listado dos ítems que superan o igualan el ID de 0,1 mientras que las compañeras no tienen ninguno que presenta la característica necesaria para aparecer en esta tabla.

2.3.1.10 CI10, “la ciudad”

Rango	Mujer	Hombre
1	calle	casa
2	coche	plaza
3	tienda	calle
4	parque	iglesia
5	escuela	centro
6	edificio	parque
7	(auto)bús	edificio
8	casa	(auto)bús
9	iglesia	estación
10	plaza	
11	bar	
12	carretera	
13	estación	
14	restaurante	
15	árbol	
16	tren	
17	museo	
18	ayuntamiento	
19	universidad	

Tabla 76. Vocablos más disponibles del CI10 según la variable “sexo”.

En este campo semántico se ha producido un total de ocho coincidencias. Las mujeres registran más vocablos exclusivos con $ID \geq 0,1$: un total de once (*coche, tienda, escuela, bar, carretera, restaurante, árbol, tren, museo, ayuntamiento, universidad*) mientras que los varones ofrecen solo uno divergente (*centro*). Basándonos en estos datos, la compatibilidad es del 40%.

2.3.1.11 CI11, “el campo”

Rango	Mujer	Hombre
1	árbol	árbol
2	animal	vaca
3	flor	animal
4	tierra	campesino
5	hierba	gallina
6	vaca	
7	campesino	

8	planta	
9	caballo	
10	naturaleza	
11	casa	

Tabla 77. Vocablos más disponibles del CI11 según la variable “sexo”.

La cardinalidad del conjunto aquí llega al 33,34% ya que los vocablos compartidos por los sociolectos son cuatro. Las mujeres han activado siete *types* exclusivos (*flor, tierra, hierba, planta, caballo, naturaleza, casa*), los hombres solo han escrito uno (*gallina*).

2.3.1.12 CI12, “medios de transporte”

Rango	Mujer	Hombre
1	coche	coche
2	(auto)bús	(auto)bús
3	tren	tren
4	bici(cleta)	bici(cleta)
5	avión	avión
6	taxi	metro
7	metro	moto(cicleta)
8	barco	taxi
9	moto(cicleta)	tranvía
10	tranvía	barco

Tabla 78. Vocablos más disponibles del CI12 según la variable “sexo”.

Este CI es particularmente interesante ya que por primera vez el número de vocablos más disponibles coincide en cada registro, donde aparecen las mismas unidades: la compatibilidad es del 100%. Es más, los lexemas de los rangos 1, 2, 3, 4 y 5 se corresponden perfectamente (*coche, autobús, tren, bicicleta, avión*).

2.3.1.13 CI13, “trabajos del campo y del jardín”

Rango	Mujer	Hombre
1	jardinero	campesino
2	campesino	jardinero
3	plantar	cortar
4	cortar	

Tabla 79. Vocablos más disponibles del CI13 según la variable “sexo”.

Los vocablos de los hombres (*campesino, jardinero, cortar*) están contenidos en la lista de las mujeres, que proporcionan una sola unidad más (*plantar*), que nos ha llevado a un porcentaje de cardinalidad del 75%.

Cabe señalar una novedad con respecto a los centros analizados hasta ahora ya que contamos con la aparición de dos verbos: *plantar* y *cortar*.

2.3.1.14 CI14, “los animales”

Rango	Mujer	Hombre
1	perro	perro
2	gato	gato
3	caballo	caballo
4	vaca	cerdo
5	pez	vaca
6	león	pez
7	pájaro	
8	cerdo	
9	gallina	
10	conejo	
11	tigre	
12	toro	
13	mariposa	
14	elefante	

Tabla 80. Vocablos más disponibles del CI14 según la variable “sexo”.

Otra vez, las unidades compartidas coinciden con la lista de los participantes masculinos que está formada por seis palabras. Constatamos ocho unidades en el listado de las mujeres que están ausentes en los hombres (*león, pájaro, gallina, conejo, tigre, toro, mariposa, elefante*). La cardinalidad es del 42,86%.

2.3.1.15 CI15, “juegos y distracciones”

Rango	Mujer	Hombre
1	fútbol	fútbol
2	baloncesto	baloncesto
3	leer	videojuegos
4	cartas	pelota
5	libro	ordenador
6	bailar	voleibol

7	cine	tele(visión)
8	escuchar- música	
9	tele(visión)	
10	pelota	
11	ordenador	
12	correr	
13	videojuegos	
14	natación	
15	pasear	

Tabla 81. Vocablos más disponibles del CI15 según la variable “sexo”.

Las palabras compartidas aquí son seis, lo que supone una compatibilidad del 37,50% desde el momento en el que los varones tienen como más disponible un solo vocablo (*voleibol*) que no está en la lista de las mujeres. Estas, a su vez, ofrecen nueve lexemas más (*leer, cartas, libro, bailar, cine, escuchar-música, correr, natación, pasear*).

Como se ha notado en el CI13, “trabajos del campo y del jardín”, no sorprende la aportación de algunos verbos debido a las asociaciones surgidas a partir del estímulo temático del centro.

2.3.1.16 CI16, “profesiones y oficios”

Rango	Mujer	Hombre
1	profesor	profesor
2	médico	abogado
3	cocinero	doctor
4	abogado	bombero
5	doctor	cocinero
6	camarero	
7	fontanero	
8	actor	
9	secretario	
10	estudiante	
11	empleado	

Tabla 82. Vocablos más disponibles del CI16 según la variable “sexo”.

Mujeres y hombres comparten cuatro palabras, pero la lista de las primeras sobrepasa la de los hombres en siete unidades (*médico, camarero, fontanero, actor, secretario, estudiante, empleado*) ya que ellos solo ofrecen una palabra (*bombero*) que

ellas no tienen (por lo menos entre los vocablos con $ID \geq 0,1$). Esto implica una compatibilidad del 33,34%.

2.3.2 Conclusiones parciales

Los sustantivos dominan las listas de los vocablos con $ID \geq 0,1$ ya que solo se han detectado tres verbos en el CI13, “trabajos del campo y del jardín”, y cuatro en el CI15, “juegos y distracciones”.

En lo que atañe a las temáticas más difundidas, hemos observado que los encuestados siguen fielmente los estímulos de cada área, otros lexemas originados por asociaciones secundarias se quedan en posiciones inferiores de los listados y no forman parte del léxico más disponible.

Pese a que en el análisis cuantitativo de los resultados hemos detectado un número superior de vocablos arrojados por los hombres (un promedio más elevado del 66,35%), aquí las alumnas ofrecen una cantidad mayor de palabras con $ID \geq 0,1$.

En resumen, los cómputos de la cardinalidad del conjunto han revelado los valores visibles en la tabla y en el gráfico siguientes.

CI	Cardinalidad del conjunto
CI01.CUE	65%
CI02.ROP	69,23%
CI03.CAS	77,78%
CI04.MUE	50%
CI05.ALI	55,17%
CI06.MES	28,57%
CI07.COC	30%
CI08.ESC	63,64%
CI09.ILU	0%
CI10.CIU	40%
CI11.CAM	33,34%
CI12.TRA	100%
CI13.TRC	75%
CI14.ANI	42,86%
CI15.JUE	37,50%
CI16.PRO	33,34%

Tabla 83. Cardinalidad del conjunto por CI según la variable “sexo”.

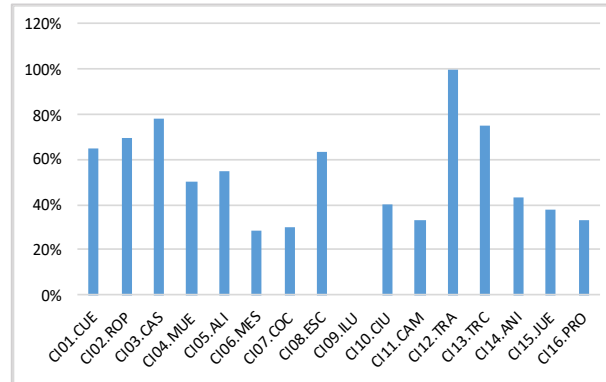


Gráfico 32. Cardinalidad del conjunto por CI según la variable "sexo".

No obstante la gran estabilidad en los primeros rangos de las listas analizadas, se observa una oscilación de la compatibilidad que va del 0% en el CI09, "iluminación y calefacción", al 100% en el CI12, "los medios de transporte", posiblemente por el hecho de que este es uno de los campos semántico menos rico y, por tanto, las aportaciones de los dos grupos coinciden en gran medida, además de que, quizás, se haya estudiado en clase y forme un conjunto compacto de palabras en el lexicón de los encuestados, a diferencia del CI09 que es un tema que no se desarrolla mucho en la ASL.

2.3.3 Variable "nivel de ELE"

2.3.3.1 CI01, "partes del cuerpo"

Rango	B1	B2
1	ojos	mano
2	mano	ojos
3	cabeza	pie
4	pie	pierna
5	pierna	brazo
6	nariz	cabeza
7	pelo	dedo
8	boca	boca
9	brazo	cara
10	cara	pelo
11	dedo	nariz
12	espalda	uñas
13	cuello	espalda
14	rodilla	orejas
15	orejas	dientes

16	lengua	cuello
17	barriga	rodilla
18		lengua
19		hombro
20		labios
21		barriga
22		cejas

Tabla 84. Vocablos más disponibles del CI01 según la variable “nivel de ELE”.

Los vocablos compartidos son diecisiete, que coinciden con todas las aportaciones de los informantes de nivel B1. La tabla muestra que los de nivel B2 presentan otras cinco unidades con $ID \geq 0,1$ (*uñas, dientes, hombro, labios, cejas*). Por tanto, la cardinalidad del conjunto es bastante elevada, 77,27%.

2.3.3.2 CI02, “la ropa”

Rango	B1	B2
1	camiseta	camiseta
2	zapatos	pantalones
3	pantalones	zapatos
4	vaqueros	camisa
5	jersey	falda
6	chaqueta	bufanda
7	camisa	chaqueta
8	falda	jersey
9	vestido	calcetines
10	abrigo	vaqueros
11		abrigo
12		vestido
13		blusa

Tabla 85. Vocablos más disponibles del CI02 según la variable “nivel de ELE”.

Otra vez, los lexemas comunes se corresponden perfectamente con la lista de vocablos de los B1 ya que los B2 arrojan una cantidad mayor de tres elementos (*bufanda, calcetines, blusa*). La cardinalidad sigue siendo elevada: hemos computado un 76,92%.

2.3.3.3 CI03, “partes de la casa (sin muebles)”

Rango	B1	B2
1	cocina	cocina

2	cuarto-de-baño	cuarto-de-baño
3	habitación	cuarto
4	salón	salón
5	jardín	jardín
6	garaje	habitación
7		techo
8		garaje
9		puerta
10		escalera
11		ventana

Tabla 86. Vocablos más disponibles del CI03 según la variable “nivel de ELE”.

La compatibilidad del conjunto baja al 55,56% ya que la suma disyuntiva entre los dos registros es mayor con respecto a lo visto hasta ahora. Los informantes de nivel B1 tienen una sola lexía exclusiva (*garaje*) y los más avanzados ofrecen seis (*cuarto*, *techo*, *garaje*, *puerta*, *escalera*, *ventana*).

Constatamos la actualización de palabras que encajan bien en el ámbito temático del campo y una gran cantidad de sustantivos, además de la locución *cuarto de baño*.

2.3.3.4 CI04, “los muebles de la casa”

Rango	B1	B2
1	cama	silla
2	silla	mesa
3	mesa	cama
4	sofá	armario
5	armario	sofá
6	sillón	sillón
7	ducha	mesilla-de-noche
8	lámpara	ducha
9		tele(visión)
10		lámpara
11		cajón

Tabla 87. Vocablos más disponibles del CI04 según la variable “nivel de ELE”.

La lista de unidades más disponibles se ajusta a la de los aprendices B1, compuesta por ocho vocablos. La tabla muestra un desnivel de tres unidades entre los dos grupos,

donde prevalece el B2 con once palabras en total, que lleva a una cardinalidad del 72,73%.

2.3.3.5 CI05, “alimentos y bebidas”

Rango	B1	B2
1	agua	agua
2	cerveza	leche
3	manzana	tomate
4	tomate	cerveza
5	vino	naranja
6	jamón	paella
7	naranja	pan
8	paella	manzana
9	pescado	pasta
10	pan	vino
11	leche	pescado
12	carne	arroz
13	zumos	jamón
14	arroz	huevo
15	bocadillo	zumos
16	azúcar	café
17	queso	carne
18	sangría	zanahoria
19	tapas	chocolate
20	café	patata
21	pasta	azúcar
22	aceite	cebolla
23	plátano	sal
24		galleta
25		plátano
26		pollo
27		limón
28		bocadillo
29		queso
30		ensalada

Tabla 88. Vocablos más disponibles del CI05 según la variable “nivel de ELE”.

Existe una coincidencia de veinte vocablos, ya que los B1 proporcionan tres voces que no están presentes en el otro listado (*sangría, tapas, aceite*) y los B2 hacen lo propio con diez (*huevo, zanahoria, chocolate, patata, cebolla, sal, galleta, pollo, limón,*

ensalada).

Estas cifras producen una compatibilidad del 60,61% debido a la alta coincidencia en las posiciones iniciales, pese a las divergencias de las finales.

2.3.3.6 CI06, “objetos colocados en la mesa para la comida”

Rango	B1	B2
1	plato	plato
2	vaso	vaso
3	cuchillo	cuchillo
4	cuchara	cuchara
5	botella	botella
6	tenedor	tenedor
7		cucharilla

Tabla 89. Vocablos más disponibles del CI06 según la variable “nivel de ELE”.

La compatibilidad del conjunto es elevada: alcanza el 85,71% desde el momento en el que la lista de vocablos con índice de disponibilidad mayor o igual a 0,1 comprende seis unidades por cada grupo y solo hay una divergente en el B2 (*cucharilla*). Es más, las lexías coincidentes se colocan exactamente en los mismos rangos en las dos opciones (*plato, vaso, cuchillo, cuchara, botella, tenedor*).

2.3.3.7 CI07, “la cocina y sus utensilios”

Rango	B1	B2
1	horno	horno
2	plato	cuchillo
3	frigorífico	plato
4	lavavajillas	cuchara
5	nevera	sartén
6	vaso	olla
7		vaso

Tabla 90. Vocablos más disponibles del CI07 según la variable “nivel de ELE”.

Los vocablos compartidos son tres (*horno, plato, vaso*), cada variante presenta en su lista algunos ítems exclusivos con $ID \geq 0,1$: los B1 proporcionan tres (*frigorífico, lavavajillas, nevera*) y los B2 cuatro (*cuchillo, cuchara, sartén, olla*), lo que supone una cardinalidad del 30%.

2.3.3.8 CI08, “la escuela: muebles y materiales”

Rango	B1	B2
1	mesa	bolí(grafo)
2	silla	libro
3	libro	silla
4	lápiz	lápiz
5	cuaderno	cuaderno
6	pizarra	mesa
7	bolí(grafo)	pizarra
8	ordenador	mochila
9	mochila	ordenador
10	papel	papel
11	borrador	borrador

Tabla 91. Vocablos más disponibles del CI08 según la variable “nivel de ELE”.

La correspondencia llega a su máximo: la cardinalidad del conjunto alcanza el 100% ya que no hay ninguna lexía divergente. Asimismo, se detecta cierta compatibilidad entre el orden de las aportaciones según rango, en particular las posiciones 4, 5, 10, 11 coinciden.

Los participantes han arrojado sustantivos relacionados tanto con el primer estímulo prototípico del centro, como con el segundo.

2.3.3.9 CI09, “iluminación y calefacción”

Rango	B1	B2
1	lámpara	lámpara
2	luz	luz
3	fuego	fuego

Tabla 92. Vocablos más disponibles del CI09 según la variable “nivel de ELE”.

De nuevo, se ha alcanzado el 100% de la cardinalidad entre los dos conjuntos que presentan tres voces totales en sus listados, con la misma ordenación.

2.3.3.10 CI10, “la ciudad”

Rango	B1	B2
1	calle	calle
2	parque	coche
3	tienda	tienda

4	escuela	casa
5	coche	parque
6	casa	(auto)bús
7	iglesia	edificio
8	edificio	escuela
9	plaza	plaza
10	(auto)bús	iglesia
11	ayuntamiento	carretera
12	estación	bar
13	árbol	estación
14	carretera	tren
15	museo	bici(cleta)
16	bar	restaurante
17	cine	palacio
18	universidad	

Tabla 93. Vocablos más disponibles del CI10 según la variable “nivel de ELE”.

La compatibilidad disminuye en este campo semántico hasta el 59,09% puesto trece vocablos coinciden en las listas y cada muestreo, por su parte, ofrece lexías divergentes. Los B1 tienen cinco unidades que no están entre las de los más avanzados (*ayuntamiento, árbol, museo, cine, universidad*), que a su vez presentan cuatro elementos exclusivos (*tren, bicicleta, restaurante, palacio*).

La mayoría de los sustantivos aportados encajan en el tema principal del centro, pero, al mismo tiempo, conviven con otros ligados a asociaciones secundarias (*coche, bicicleta, árbol, tren*).

2.3.3.11 CIII, “el campo”

Rango	B1	B2
1	árbol	árbol
2	animal	animal
3	flor	tierra
4	vaca	hierba
5	naturaleza	flor
6	caballo	planta
7	hierba	campesino
8	campesino	fruta
9	perro	vaca
10	casa	gallina

11		verdura
12		tractor
13		caballo

Tabla 94. Vocablos más disponibles del CI11 según la variable “nivel de ELE”.

En este CI, se pone de relieve la supremacía de los alumnos B2, los cuales producen una lista de vocablos con $ID \geq 0,1$ de trece elementos. Los B1 alcanzan un total de diez. Las lexías compartidas son siete (*árbol, animal, flor, vaca, caballo, hierba, campesino*), en consecuencia, hemos contabilizado una cardinalidad del 43,75%.

2.3.3.12 CI12, “medios de transporte”

Rango	B1	B2
1	coche	coche
2	(auto)bús	(auto)bús
3	tren	tren
4	bici(cleta)	bici(cleta)
5	avión	avión
6	barco	taxi
7	moto(cicleta)	metro
8	metro	moto(cicleta)
9	taxi	barco
10	tranvía	tranvía

Tabla 95. Vocablos más disponibles del CI12 según la variable “nivel de ELE”.

La única diferencia entre los registros son las colocaciones distintas de las palabras en los rangos de 6 a 9 de los dos grupos ya que las respuestas coinciden en las demás posiciones. Así pues, la compatibilidad alcanzada es del 100% ya que ambos han arrojado el mismo número de vocablos con $ID \geq 0,1$.

2.3.3.13 CI13, “trabajos del campo y del jardín”

Rango	B1	B2
1	jardinero	jardinero
2	campesino	plantar
3		campesino
4		cortar

Tabla 96. Vocablos más disponibles del CI13 según la variable “nivel de ELE”.

Cuatro lexemas coinciden aquí, los B2 tienen en su lista 2 vocablos que los B1 no

presentan (*plantar, cortar*), por lo tanto, la compatibilidad es del 50%.

2.3.3.14 CII4, “los animales”

Rango	B1	B2
1	perro	perro
2	gato	gato
3	caballo	caballo
4	vaca	vaca
5	pez	pájaro
6	pájaro	cerdo
7	león	león
8	toro	pez
9	cerdo	gallina
10	conejo	elefante
11	tigre	tigre
12	gallina	pollo
13	burro	cebra
14		conejo
15		ave

Tabla 97. Vocablos más disponibles del CII4 según la variable “nivel de ELE”.

Se notan once vocablos coincidentes y seis que no están compartidos, que se reparten entre las variantes: los estudiantes B1 ofrecen dos piezas exclusivas (*toro, burro*) y los B2 hacen lo propio con cuatro (*elefante, pollo, cebra, ave*). La cardinalidad es 64,71%.

2.3.3.15 CII5, “juegos y distracciones”

Rango	B1	B2
1	fútbol	fútbol
2	baloncesto	baloncesto
3	bailar	leer
4	leer	cartas
5	libro	libro
6	pelota	videojuegos
7	cine	ordenador
8	cartas	tele(visión)
9	tele(visión)	bailar
10	jugar	pelota

11	pasear	nadar
12	natación	escuchar- música
13	escuchar- música	
14	ordenador	
15	tocar	
16	correr	

Tabla 98. Vocablos más disponibles del CI15 según la variable “nivel de ELE”.

En esta área temática, el grupo B1 prevalece sobre el B2, aportando en total dieciséis vocablos más disponibles, contra los doce de los compañeros. En efecto, tiene seis voces divergentes (*cine, jugar, pasear, natación, tocar, correr*) mientras que los otros solo logran dos (*videojuegos, nadar*). De todas formas, las palabras compartidas son diez, lo que lleva a calcular una cardinalidad del conjunto del 55,56%.

Aparece aquí, por encima de los sustantivos, una cantidad notable de verbos debido a la tipología de asociaciones vinculadas con este campo.

2.3.3.16 CI16, “profesiones y oficios”

Rango	B1	B2
1	profesor	profesor
2	camarero	médico
3	cocinero	abogado
4	fontanero	cocinero
5	médico	doctor
6	estudiante	actor
7	doctor	secretario
8	abogado	empleado
9	entrenador	bombero
10	actor	policía
11		camarero
12		enfermero

Tabla 99. Vocablos más disponibles del CI16 según la variable “nivel de ELE”.

La compatibilidad computada es del 46,67% debido a que los vocablos compartidos son siete y los aprendices de nivel B1 revelan tres lexías nuevas en su lista (*fontanero, estudiante, entrenador*), los de nivel B2 hacen lo mismo con cinco (*secretario, empleado, bombero, policía, enfermero*).

2.3.4 Conclusiones parciales

Sobresale, de nuevo, que los sustantivos son la clase gramatical más difundida en los registros. Los verbos aparecen en las primeras posiciones en el CI13, “trabajos del campo y del jardín”, y en el CI15, “juegos y distracciones”, con un porcentaje del 30%. Al igual que lo observado en el análisis cuantitativo, el grupo de aprendices de nivel B2 ha arrojado una cantidad mayor de *tokens* (+7,52%) y de *types* (+14,17%), aquí también las listas contienen más unidades léxicas. En lo que se refiere a la compatibilidad del conjunto, la tabla muestra los valores según centro de interés, así como el gráfico.

CI	Cardinalidad del conjunto
CI01.CUE	77,27%
CI02.ROP	76,92%
CI03.CAS	55,56%
CI04.MUE	72,73%
CI05.ALI	60,61%
CI06.MES	85,71%
CI07.COC	30%
CI08.ESC	100%
CI09.ILU	100%
CI10.CIU	59,09%
CI11.CAM	43,75%
CI12.TRA	100%
CI13.TRC	50%
CI14.ANI	64,71%
CI15.JUE	55,56%
CI16.PRO	46,67%

Tabla 100. Cardinalidad del conjunto por CI según la variable “nivel de ELE”.

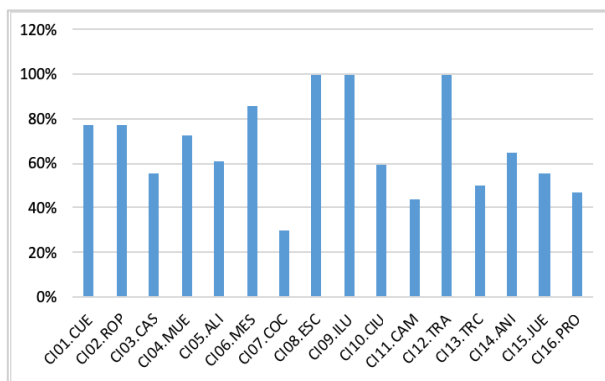


Gráfico 33. Cardinalidad del conjunto por CI según la variable “nivel de ELE”.

Los porcentajes que hemos computado varían del 30% en el CI07, “la cocina y sus utensilios”, al 100% en los CI08, “la escuela: muebles y materiales”; CI09, “iluminación y calefacción”; CI12, “medios de transporte”, donde ambos grupos de participantes han activado los mismos vocablos, sea en lo que concierne a la categoría gramatical sea a la temática de referencia.

2.3.5 Variable “conocimiento de otras LE”

2.3.5.1 CI01, “partes del cuerpo”

Rango	=2 LE	>2 LE
1	ojos	mano
2	mano	ojos
3	cabeza	cabeza
4	pie	pie
5	cara	pierna
6	pelo	brazo
7	nariz	boca
8	boca	dedo
9	pierna	nariz
10	dedo	pelo
11	dientes	cara
12	rodilla	espalda
13	espalda	cuello
14	brazo	orejas
15	barriga	rodilla
16	uñas	lengua
17	lengua	dientes
18	hombro	barriga
19		uñas

Tabla 101. Vocablos más disponibles del CI01 según la variable “conocimiento de otras LE”.

La cardinalidad del conjunto es 68,18% ya que los vocablos comunes son quince y cada sociolecto aporta algunos elementos divergentes

El grupo =2 LE tiene tres (*cara, dedo, hombro*) y el otro cuatro (*boca, cara, cuello, orejas*), se trata siempre de sustantivos relacionados con el estímulo temático principal del centro de interés.

2.3.5.2 CI02, “la ropa”

Rango	=2 LE	>2 LE
1	camiseta	camiseta
2	zapatos	zapatos
3	chaqueta	pantalones
4	pantalones	camisa
5	falda	jersey
6	jersey	vaqueros
7	camisa	falda
8	vaqueros	chaqueta
9	calcetines	bufanda
10		abrigo
11		vestido
12		calcetines
13		traje

Tabla 102. Vocablos más disponibles del CI02 según la variable “conocimiento de otras LE”.

La correspondencia entre las dos variantes es de nueve lexemas, que coinciden con la lista de los estudiantes que conocen dos LE. En efecto, los que conocen más han escrito más *types* con $ID \geq 0,1$: tienen cuatro voces exclusivas (*bufanda*, *abrigo*, *vestido*, *traje*), con lo cual la compatibilidad es del 69,23%.

2.3.5.3 CI03, “partes de la casa (sin muebles)”

Rango	=2 LE	>2 LE
1	cocina	cocina
2	cuarto-de-baño	cuarto-de-baño
3	habitación	habitación
4	salón	salón
5	jardín	jardín
6	garaje	cuarto
7	cuarto	garaje
8	escalera	
9	techo	

Tabla 103. Vocablos más disponibles del CI03 según la variable “conocimiento de otras LE”.

En este CI hemos computado el 77,78% de cardinalidad, un porcentaje mayor con respecto a los contabilizados antes puesto que la discrepancia entre los dos conjuntos es

solo de dos palabras (*escalera, techo*) que están en el listado del grupo =2 LE.

Si excluimos los rangos 6 y 7, las unidades se colocan exactamente en las mismas posiciones, son sustantivos y están en línea con el área temática. Quizás esta correspondencia sea debida al hecho de que se trata de un centro que, normalmente, se estudia en las primeras etapas de la ASL en clase y después se descuida, dejando fosilizado el bagaje de vocabulario aprendido.

2.3.5.4 CI04, “los muebles de la casa”

Rango	=2 LE	>2 LE
1	cama	cama
2	mesa	silla
3	silla	mesa
4	sofá	sofá
5	armario	armario
6	tele(visión)	sillón
7	sillón	ducha
8	frigorífico	lámpara
9	espejo	mesilla-de-noche

Tabla 104. Vocablos más disponibles del CI04 según la variable “conocimiento de otras LE”.

Las coincidencias son seis en total, la cardinalidad baja hasta el 50% porque se ha detectado el mismo número de vocablos divergentes en cada una de las opciones. Los encuestados del muestreo >2 LE proporcionan entre sus respuestas la unidad pluriverbal *mesilla de noche*, que no está entre las de los compañeros =2 LE, que, a su vez, han aportado *televisión, frigorífico, espejo*, ausentes en la otra lista.

2.3.5.5 CI05, “alimentos y bebidas”

Rango	=2 LE	>2 LE
1	agua	agua
2	cerveza	leche
3	vino	tomate
4	naranja	manzana
5	zumos	cerveza
6	café	paella
7	manzana	pescado

8	tomate	jamón
9	jamón	pan
10	paella	naranja
11	pan	arroz
12	leche	vino
13	pescado	pasta
14	carne	carne
15	bocadillo	zumos
16	sangría	queso
17	zanahoria	azúcar
18	galleta	bocadillo
19	huevo	huevo
20	chocolate	patata
21		café
22		plátano
23		sal
24		chocolate
25		tapas

Tabla 105. Vocablos más disponibles del CI05 según la variable “conocimiento de otras LE”.

Cada conjunto ofrece algunos vocablos peculiares: en la lista de los alumnos que saben menos LE aparecen tres (*sangría, zanahoria, galleta*) mientras que los que conocen más ofrecen ocho (*arroz, pasta, queso, azúcar, patata, plátano, sal, tapas*). En total, los vocablos compartidos son diecisiete y permiten contabilizar una cardinalidad del 60,71%.

2.3.5.6 CI06, “objetos colocados en la mesa para la comida”

Rango	=2 LE	>2 LE
1	plato	plato
2	vaso	vaso
3	cuchillo	cuchillo
4	botella	cuchara
5	cuchara	botella
6	tenedor	tenedor

Tabla 106. Vocablos más disponibles del CI06 según la variable “conocimiento de otras LE”.

Al producirse una correspondencia total en todas las posiciones, la cardinalidad logra el 100%. No hay ningún vocablo divergente, la única diferencia está en los rangos 4 y 5 que intercambian las lexías *botella* y *cuchara*.

Posiblemente esta correspondencia se deba a que no se preste mucha atención a este asunto en las clases y en los materiales de ELE: puede que una vez que se haya presentado no se profundice y, por ende, el caudal aprendido sea igual para todos.

2.3.5.7 CI07, “la cocina y sus utensilios”

Rango	=2 LE	>2 LE
1	horno	horno
2	nevera	plato
3	cuchillo	cuchillo
4	sartén	cuchara
5		vaso
6		sartén
7		frigorífico

Tabla 107. Vocablos más disponibles del CI07 según la variable “conocimiento de otras LE”.

La compatibilidad de este campo semántico es la menor que hemos computado hasta ahora en lo que atañe a esta variable: no supera el 37,50%. Las palabras comunes son tres, cada grupo tiene en su lista por los menos un vocablo exclusivo: los informantes =2 LE presentan *nevera*; los >2 LE hacen lo propio con *plato*, *cuchara*, *vaso*, *frigorífico*.

2.3.5.8 CI08, “la escuela: muebles y materiales”

Rango	=2 LE	>2 LE
1	libro	silla
2	bolí(grafo)	mesa
3	lápiz	libro
4	cuaderno	cuaderno
5	pizarra	lápiz
6	mesa	bolí(grafo)
7	silla	pizarra
8	papel	mochila
9	ordenador	ordenador
10		papel
11		borrador

Tabla 108. Vocablos más disponibles del CI08 según la variable “conocimiento de otras LE”.

Los sustantivos compartidos son nueve y forman la lista entera de los estudiantes que conocen dos LE. Del otro lado, la lista de los compañeros es más larga ya que contiene

diez elementos, de los cuales dos son divergentes (*mochila, borrador*). La cardinalidad entre las dos opciones de la variable es bastante elevada: 81,82%.

2.3.5.9 CI09, “iluminación y calefacción”

Rango	=2 LE	>2 LE
1	lámpara	lámpara
2	luz	luz
3	fuego	fuego

Tabla 109. Vocablos más disponibles del CI09 según la variable “conocimiento de otras LE”.

Como no se ha detectado ninguna diferencia entre los dos grupos, se ha producido el 100% de cardinalidad. Asimismo, las tres voces comunes se colocan en los mismos rangos en ambos registros.

2.3.5.10 CI10, “la ciudad”

Rango	=2 LE	>2 LE
1	calle	calle
2	tienda	parque
3	edificio	coche
4	casa	escuela
5	coche	tienda
6	iglesia	(auto)bús
7	plaza	casa
8	parque	edificio
9	árbol	iglesia
10	bar	plaza
11	(auto)bús	estación
12	museo	carretera
13	carretera	bar
14	escuela	tren
15		ayuntamiento
16		restaurante
17		universidad

Tabla 110. Vocablos más disponibles del CI10 según la variable “conocimiento de otras LE”.

En este CI los lexemas compartidos son doce y cada grupo tiene algunos exclusivos: dos (*árbol, museo*) en el grupo =2 LE y cinco (*estación, tren, ayuntamiento, restaurante, universidad*) en el >2 LE.

Según eso, la cardinalidad del conjunto es 63,16%.

Observamos la presencia de aportaciones ligadas al estímulo prototípico del tema y también de algunas procedentes de asociaciones secundarias.

2.3.5.11 CIII, “el campo”

Rango	=2 LE	>2 LE
1	árbol	árbol
2	animal	animal
3	campesino	flor
4	caballo	hierba
5	vaca	tierra
6	perro	vaca
7	flor	planta
8		campesino
9		caballo

Tabla 111. Vocablos más disponibles del CI11 según la variable “conocimiento de otras LE”.

La compatibilidad es del 60% debido a los seis vocablos compartidos y a los vocablos peculiares de cada grupo: uno en los estudiantes que conocen dos LE (*perro*) y tres en los que saben más (*hierba, tierra, planta*).

2.3.5.12 CI12, “medios de transporte”

Rango	=2 LE	>2 LE
1	coche	coche
2	avión	(auto)bús
3	(auto)bús	tren
4	tren	bici(cleta)
5	bici(cleta)	avión
6	metro	barco
7	tranvía	taxi
8	taxi	moto(cicleta)
9	moto(cicleta)	metro
10		tranvía

Tabla 112. Vocablos más disponibles del CI12 según la variable “conocimiento de otras LE”.

El único lexema divergente (*barco*) se encuentra en el listado >2 LE, los elementos comunes son diez, como indica la tabla, y esto hace que la compatibilidad

llegue a un 90%.

2.3.5.13 CI13, “trabajos del campo y del jardín”

Rango	=2 LE	>2 LE
1	jardinero	jardinero
2	campesino	campesino
3		plantar
4		cortar

Tabla 113. Vocablos más disponibles del CI13 según la variable “conocimiento de otras LE”.

La lista del conjunto =2 LE engloba ambas lexías compartidas (*jardinero*, *campesino*). Los >2 LE ofrecen, por su parte, otros dos vocablos, se trata de verbos: *plantar*, *cortar*. Basándonos en estos datos la cardinalidad es del 50%.

2.3.5.14 CI14, “los animales”

Rango	=2 LE	>2 LE
1	perro	perro
2	gato	gato
3	caballo	caballo
4	cerdo	vaca
5	pez	león
6	pájaro	pez
7	vaca	pájaro
8	león	gallina
9	pollo	conejo
10	tigre	cerdo
11	gallina	tigre
12		toro
13		elefante
14		burro

Tabla 114. Vocablos más disponibles del CI14 según la variable “conocimiento de otras LE”.

En este centro de interés las unidades que alcanzan el $ID \geq 0,1$ son diez y la compatibilidad es del 66,67% puesto que cada muestreo tiene vocablos exclusivos. Efectivamente, el primero actualiza *pollo* mientras que el segundo proporciona *conejo*, *toro*, *elefante*, *burro*.

2.3.5.15 CII5, “juegos y distracciones”

Rango	=2 LE	>2 LE
1	fútbol	fútbol
2	baloncesto	baloncesto
3	leer	cartas
4	libro	leer
5	tele(visión)	bailar
6	bailar	libro
7	ordenador	pelota
8	pelota	videojuegos
9	<i>Play-Station</i>	escuchar-música
10	nadar	cine
11	videojuegos	ordenador
12	cine	natación
13	disco(teca)	correr
14		tele(visión)

Tabla 115. Vocablos más disponibles del CII5 según la variable “conocimiento de otras LE”.

Son diez las palabras comunes en esta área temática, donde al analizar por separado las listas, observamos que un grupo ha activado tres vocablos exclusivos (*Play Station, nadar, discoteca*) y el otro, cuatro (*cartas, escuchar música, natación, correr*), con lo cual el porcentaje de cardinalidad es del 58,82%.

Constatamos, en lo que se refiere a las categorías gramaticales, la aparición del 26% de verbos frente al 74% de sustantivos.

2.3.5.16 CII6, “profesiones y oficios”

Rango	=2 LE	>2 LE
1	profesor	profesor
2	médico	cocinero
3	abogado	médico
4	fontanero	abogado
5	cocinero	doctor
6	doctor	camarero
7	camarero	actor
8	jardinero	empleado
9		fontanero
10		estudiante

11		secretario
12		cantante

Tabla 116. Vocablos más disponibles del CI16 según la variable “conocimiento de otras LE”.

La compatibilidad es del 53,85%, pues las palabras que los dos muestreos tienen en común son siete y ambos presentan por lo menos una divergente. La tabla muestra que los aprendices =2 LE tienen otra unidad con $ID \geq 0,1$ (*jardinero*) y los >2 LE presentan cinco (*actor, empleado, estudiante, secretario, cantante*).

2.3.6 Conclusiones parciales

De la lectura de las tablas se deduce que los participantes que conocen más lenguas han arrojado una cantidad mayor de vocablos con respecto a los compañeros que solo saben dos, no obstante en ciertos campos semánticos la compatibilidad del conjunto haya alcanzado el 100%.

Se ha confirmado la tendencia detectada en los análisis previos que ve la prevalencia de los sustantivos sobre las otras categorías gramaticales ya que solo han aparecido verbos en los CI13, “trabajos del campo y del jardín” (33%), y CI15, “juegos y distracciones” (26%).

Hemos detectado la activación de unidades con ID mayor o igual a 0,1 perfectamente en línea con las asociaciones primarias de cada centro, el único que ha presentado otras conexiones ha sido el CI10, “la ciudad”, a causa de la amplitud de relaciones léxico-semánticas que podían crearse en la mente de los informantes. En los demás casos las palabras pertenecientes a nexos secundarios se encuentran en posiciones más bajas de las listas.

Para concluir, la cardinalidad se resume en la tabla y en el gráfico que siguen.

CI	Cardinalidad del conjunto
CI01.CUE	68,18%
CI02.ROP	69,23%
CI03.CAS	77,78%
CI04.MUE	50%
CI05.ALI	60,71%
CI06.MES	100%
CI07.COC	37,50%

CI08.ESC	81,82%
CI09.ILU	100%
CI10.CIU	63,16%
CI11.CAM	60%
CI12.TRA	90%
CI13.TRC	50%
CI14.ANI	66,67%
CI15.JUE	58,82%
CI16.PRO	53,85%

Tabla 117. Cardinalidad del conjunto por CI según la variable “conocimiento de otras LE”.

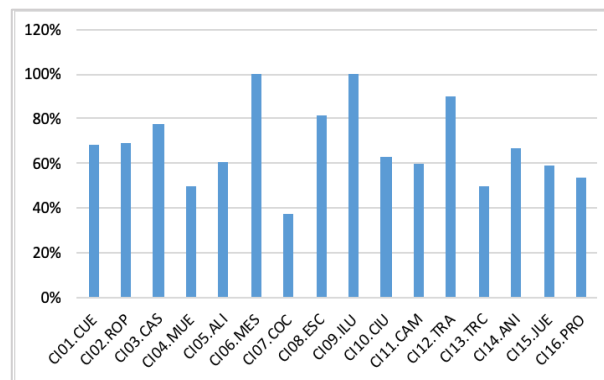


Gráfico 34. Cardinalidad del conjunto por CI según la variable “conocimiento de otras LE”.

La compatibilidad oscila de valores bajos (37,50% en el CI07, “la cocina y sus utensilios”) a muy elevados (90% en CI12, “medios de transporte”, 100% en los CI06, “objetos colocados en la mesa para la comida”, y CI09, “iluminación y calefacción”). Quizás esto derive de que ciertos argumentos no se traten detenidamente durante la ASL escolar y, por ende, los estudiantes actualizan lexemas distintos unos de los otros. Al contrario, en los temas más comunes la coincidencia es mayor posiblemente porque es fruto de un aprendizaje sistematizado.

2.4 Conclusiones del análisis transversal

Los sustantivos son la categoría morfológica que domina las listas de los vocablos más disponibles, o sea que tienen un $ID \geq 0,1$, conforme a nuestra medida de corte. Análogamente, constatamos algunos verbos en los campos semánticos CI13, “trabajos del campo y del jardín”, y CI15, “juegos y distracciones”, cuya presencia cubre aproximadamente el 30% de los listados.

Los informantes han activado, por lo general, lexemas que se adscriben perfectamente a los estímulos principales de cada centro propuesto, mientras que en los rangos inferiores se han detectado lexías referentes a asociaciones secundarias.

En lo que se refiere al análisis general, los registros varían de tres a veinticinco unidades léxicas. Los centros CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”, y CI10, “la ciudad”, son los más ricos, mientras que los más pobres son el CI09, “iluminación y calefacción”, y el CI13, “trabajos del campo y del jardín”.

Pasando a los resultados por variable, se nota la supremacía de las siguientes variantes en la cantidad de *types* más disponibles: mujer, nivel B2, conoce más de dos LE. Aún más, los cómputos que han determinado la cardinalidad del conjunto por centro de interés han revelado que en ciertos ámbitos se ha alcanzado el 100% de compatibilidad y en otros, al contrario, ha bajado hasta el 0%.

3. ANÁLISIS LONGITUDINAL

Analizamos según la metodología empírica ya utilizada las diferencias entre las respuestas para comprobar si tras la asistencia a un curso de ELE de un año las palabras con $ID \geq 0,1$ activadas son las mismas o cambian.

3.1 Análisis descriptivo

3.1.1 CI01, “partes del cuerpo”

Rango	B2_a	B2_b
1	mano	mano
2	ojos	ojos
3	pie	pierna
4	pierna	cabeza
5	brazo	pie
6	cabeza	boca
7	dedo	dedo
8	boca	brazo
9	cara	nariz
10	pelo	pelo
11	nariz	cara
12	uñas	orejas

13	espalda	espalda
14	orejas	cuello
15	dientes	rodilla
16	cuello	uñas
17	rodilla	lengua
18	lengua	dientes
19	hombro	hombro
20	labios	labios
21	barriga	barriga
22	cejas	

Tabla 118. Vocablos más disponibles del CI01 según la fecha de la prueba.

En la primera prueba los encuestados han activado una palabra más, *cejas*, con respecto a la segunda y que la correspondencia entre las dos es de veintiuno lexemas, por tanto, la cardinalidad llega al 95,45%.

Las diferencias más evidentes se detectan entre los rangos en los que se posicionan las unidades léxicas dentro de las listas.

3.1.2 CI02, “la ropa”

Rango	B2_a	B2_b
1	camiseta	camiseta
2	pantalones	pantalones
3	zapatos	falda
4	camisa	zapatos
5	falda	camisa
6	bufanda	sudadera
7	chaqueta	vaqueros
8	jersey	chaqueta
9	calcetines	sombrero
10	vaqueros	bufanda
11	abrigo	calcetines
12	vestido	vestido
13	blusa	jersey
14		bolso
15		cinturón
16		traje
17		minifalda
18		zapatillas
19		abrigo

20		tacones
21		blusa

Tabla 119. Vocablos más disponibles del CI02 según la fecha de la prueba.

Aquí se pone de manifiesto que el corpus B2_b tiene ocho lexemas exclusivos (*sudadera, sombrero, bolso, cinturón, traje, minifalda, zapatillas, tacones*) ya que las trece palabras del otro están compartidas en las dos listas.

La compatibilidad del conjunto es del 61,90%.

3.1.3 CI03, “partes de la casa (sin muebles)”

Rango	B2_a	B2_b
1	cocina	cuarto-de-baño
2	cuarto-de-baño	cocina
3	cuarto	habitación
4	salón	jardín
5	jardín	salón
6	habitación	garaje
7	techo	techo
8	garaje	cuarto
9	puerta	pasillo
10	escalera	comedor
11	ventana	escalera
12		piso

Tabla 120. Vocablos más disponibles del CI03 según la fecha de la prueba.

Cada conjunto presenta algunas palabras divergentes: en la primera prueba constatamos dos (*puerta, ventana*), en la segunda tres (*pasillo, comedor, piso*), mientras que las comunes son nueve. Todo esto permite contabilizar una compatibilidad del 64,29%.

3.1.4 CI04, “los muebles de la casa”

Rango	B2_a	B2_b
1	silla	silla
2	mesa	mesa
3	cama	cama

4	armario	armario
5	sofá	sofá
6	sillón	sillón
7	mesilla-de-noche	mesilla-de-noche
8	ducha	horno
9	tele(visión)	lámpara
10	lámpara	estantería
11	cajón	lavadora
12		nevera

Tabla 121. Vocablos más disponibles del CI04 según la fecha de la prueba.

La cardinalidad es del 53,33% y ha sido computada a partir de las ocho voces comunes y de la suma disyuntiva que ve tres vocablos peculiares en B2_a (*ducha, televisión, cajón*) y cuatro en B2_b (*horno, estantería, lavadora, nevera*).

Ambos registros contienen, además de la unidad multipalabra *mesilla de noche*, otros sustantivos que se adscriben bien al principal estímulo del campo, bien a estímulos secundarios.

3.1.5 CI05, “alimentos y bebidas”

Rango	B2_a	B2_b
1	agua	agua
2	leche	manzana
3	tomate	cerveza
4	cerveza	naranja
5	naranja	jamón
6	paella	paella
7	pan	pescado
8	manzana	tomate
9	pasta	leche
10	vino	sangría
11	pescado	vino
12	arroz	bocadillo
13	jamón	plátano
14	huevo	pan
15	zumos	pasta
16	café	zumos
17	carne	queso

18	zanahoria	fresa
19	chocolate	tequila
20	patata	limón
21	azúcar	mojito
22	cebolla	zanahoria
23	sal	arroz
24	galleta	pollo
25	plátano	gazpacho
26	pollo	patata
27	limón	hamburguesa
28	bocadillo	tortilla
29	queso	
30	ensalada	

Tabla 122. Vocablos más disponibles del CI05 según la fecha de la prueba.

Hay veinte elementos compartidos y destaca que el corpus B2_a tiene diez divergentes (*huevo, café, carne, chocolate, patata, azúcar, cebolla, sal, galleta, ensalada*) mientras que hay ocho exclusivos en el B2_b (*sangría, fresa, tequila, gazpacho, patata, hamburguesa, tortilla*). Estos datos llevan a computar una compatibilidad del 52,63%.

3.1.6 CI06, “objetos colocados en la mesa para la comida”

Rango	B2_a	B2_b
1	plato	cuchillo
2	vaso	plato
3	cuchillo	vaso
4	cuchara	cuchara
5	botella	tenedor
6	tenedor	mantel
7	cucharilla	botella
8		servilleta
9		sal
10		cucharilla

Tabla 123. Vocablos más disponibles del CI06 según la fecha de la prueba.

Debido a la gran correlación entre los vocablos ubicados en las primeras posiciones, la cardinalidad del conjunto es del 70%. Los ítems compartidos son siete y coinciden con el listado de la primera prueba. Por el contrario, la segunda contiene tres unidades originales (*mantel, servilleta, sal*).

3.1.7 CI07, “la cocina y sus utensilios”

Rango	B2_a	B2_b
1	horno	horno
2	cuchillo	sartén
3	plato	olla
4	cuchara	plato
5	sartén	cuchara
6	olla	cuchillo
7	vaso	nevera
8		tenedor
9		(horno) microondas

Tabla 124. Vocablos más disponibles del CI07 según la fecha de la prueba.

Las lexías compartidas son seis, los informantes en la primera prueba ofrecen una palabra exclusiva (*vaso*) y tres en la segunda (*nevera, tenedor, horno microondas*), lo que supone que la cardinalidad sea del 60%.

Creemos que los encuestados han utilizado las respuestas del centro anterior para completar este, se trataría de *aprovechamiento léxico*.

3.1.8 CI08, “la escuela: muebles y materiales”

Rango	B2_a	B2_b
1	bolí(grafo)	bolí(grafo)
2	libro	silla
3	silla	pizarra
4	lápiz	lápiz
5	cuaderno	libro
6	mesa	cuaderno
7	pizarra	mesa
8	mochila	estuche
9	ordenador	mochila
10	papel	profe(sor)
11	borrador	estudiar
12		ordenador
13		borrador
14		estudiante
15		escribir
16		leer

17		papel
18		sacapuntas

Tabla 125. Vocablos más disponibles del CI08 según la fecha de la prueba.

La compatibilidad del conjunto calculada es del 61,11%. En efecto, son once los vocablos comunes en las dos suministraciones, que están comprendidos en la lista de la primera prueba, y contamos con siete exclusivos en la segunda (*estuche, profesor, estudiar, estudiante, escribir, leer, sacapuntas*).

3.1.9 CI09, “iluminación y calefacción”

Rango	B2_a	B2_b
1	lámpara	lámpara
2	luz	luz
3	fuego	fuego
4		bombilla
5		calor
6		sol
7		radiador
8		iluminar
9		calefacción
10		encender

Tabla 126. Vocablos más disponibles del CI09 según la fecha de la prueba.

Este campo semántico logra el menor porcentaje de cardinalidad con respecto a los demás, el 30%. Las palabras presentes en ambos listados son tres (*lámpara, luz, fuego*) y las exclusivas del corpus B2_b son siete (*bombilla, calor, sol, radiador, iluminar, calefacción, encender*). Esto corrobora el incremento del bagaje léxico al final del año académico ya que los vocablos más disponibles pasan de tres a diez.

Constatamos tanto la activación de sustantivos asociados al primer estímulo temático, como al segundo.

3.1.10 CII0, “la ciudad”

Rango	B2_a	B2_b
1	calle	calle
2	coche	plaza
3	tienda	iglesia

4	casa	coche
5	parque	tienda
6	(auto)bús	parque
7	edificio	escuela
8	escuela	(auto)bús
9	plaza	estación
10	iglesia	casa
11	carretera	carretera
12	bar	mercado
13	estación	semáforo
14	tren	hospital
15	bici(cleta)	museo
16	restaurante	restaurante
17	palacio	monumento
18		ayuntamiento
19		avenida
20		cine
21		bici(cleta)
22		barrio

Tabla 127. Vocablos más disponibles del CI10 según la fecha de la prueba.

Contamos con trece ítems compartidos que se suman a los divergentes en cada muestreo: *edificio, bar, tren, palacio* en el B2_a y *mercado, semáforo, hospital, museo, monumento, ayuntamiento, avenida, cine, barrio* en el B2_b. La compatibilidad, por ende, es del 50%.

3.1.11 CIII, “el campo”

Rango	B2_a	B2_b
1	árbol	árbol
2	animal	tierra
3	tierra	animal
4	hierba	flor
5	flor	hierba
6	planta	campo
7	campesino	cortijo
8	fruta	vaca
9	vaca	campesino
10	gallina	tractor
11	verdura	cultivar

12	tractor	río
13	caballo	caballo

Tabla 128. Vocablos más disponibles del CI11 según la fecha de la prueba.

Con una intersección de ocho vocablos (*árbol, animal, tierra, hierba, flor, campesino, vaca, caballo*) y una suma disyuntiva de diez, que se divide respectivamente en cinco elementos exclusivos por cada corpus (B2_a: *planta, fruta, gallina, verdura, tractor*; B2_b: *campo, cortijo, tractor, cultivar, río*), la compatibilidad es del 44,44%.

3.1.12 CI12, “medios de transporte”

Rango	B2_a	B2_b
1	coche	coche
2	(auto)bús	(auto)bús
3	tren	tren
4	bici(cleta)	avión
5	avión	bici(cleta)
6	taxi	moto(cicleta)
7	metro	metro
8	moto(cicleta)	taxi
9	barco	barco
10	tranvía	tranvía
11		a-pie
12		carro

Tabla 129. Vocablos más disponibles del CI12 según la fecha de la prueba.

De un total de veintitrés vocablos, solo diez están presentes en ambos listados y corresponden con la lista precedente de la prueba suministrada a comienzo del año académico; las unidades exclusivas de la otra son dos (*a pie, carro*). La cardinalidad total, pues, es del 83,33%.

Señalamos una alta coincidencia entre las posiciones de los vocablos dentro de los dos registros, donde se han recopilado, además de los sustantivos más comunes la unidad *a pie* y una palabra típica de la variedad hispanoamericana del español, *carro*.

3.1.13 CI13, “trabajos del campo y del jardín”

Rango	B2_a	B2_b
1	jardinero	cortar

2	plantar	plantar
3	campesino	jardinero
4	cortar	cultivar
5		regar
6		campesino
7		flor

Tabla 130. Vocablos más disponibles del CI13 según la fecha de la prueba.

La compatibilidad es del 57,14% puesto que las dos agrupaciones comparten cuatro vocablos (*jardinero, plantar, campesino, cortar*) y al realizar la prueba al final del año académico los informantes han actualizado tres vocablos más con $ID \geq 0,1$ (*cultivar, regar, flor*).

Las temáticas predominantes están estrechamente vinculadas al CI: por encima de los sustantivos, destaca una cantidad notable de verbos que describen las actividades que se desarrollan en el campo y el jardín (*plantar, cortar, cultivar, regar*).

3.1.14 CII4, “los animales”

Rango	B2_a	B2_b
1	perro	perro
2	gato	gato
3	caballo	pájaro
4	vaca	caballo
5	pájaro	vaca
6	cerdo	león
7	león	pez
8	pez	cerdo
9	gallina	conejo
10	elefante	ratón
11	tigre	jirafa
12	pollo	tigre
13	cebra	gallina
14	conejo	burro
15	ave	pollo
16		tortuga
17		hámster
18		mono
19		elefante
20		mariposa

21		serpiente
----	--	-----------

Tabla 131. Vocablos más disponibles del CI14 según la fecha de la prueba.

Los corpus comparten trece vocablos y difieren en diez: dos son exclusivos del B2_a (*cebra, ave*) y ocho del B2_b (*ratón, jirafa, burro, tortuga, hámster, mono, mariposa, serpiente*). Conforme a esto, la compatibilidad del conjunto es del 56,52%.

3.1.15 CI15, “juegos y distracciones”

Rango	B2_a	B2_b
1	fútbol	jugar
2	baloncesto	leer
3	leer	baloncesto
4	cartas	fútbol
5	libro	libro
6	videojuegos	tele(visión)
7	ordenador	cartas
8	tele(visión)	bailar
9	bailar	ordenador
10	pelota	móvil
11	nadar	correr
12	escuchar- música	cine
13		tenis
14		nadar
15		videojuegos
16		escuchar- música
17		pelota
18		deporte
19		pasear
20		balón
21		redes- sociales
22		voleibol
23		natación

Tabla 132. Vocablos más disponibles del CI15 según la fecha de la prueba.

La tabla contiene las doce unidades léxicas más disponibles comunes. En la segunda aplicación, disponemos de once vocablos más con $ID \geq 0,1$ (*jugar, móvil, correr,*

cine, tenis, deporte, pasear, balón, redes-sociales, voleibol, natación). La correlación de estos datos ha llevado a una cardinalidad del conjunto del 52,17%.

Encontramos sustantivos y verbos a la vez, estos últimos (*jugar, leer, bailar, nadar, correr, escuchar música, pasear*) cubren el 31% de las actualizaciones totales.

3.1.16 CII6, “profesiones y oficios”

Rango	B2_a	B2_b
1	profesor	profesor
2	médico	abogado
3	abogado	camarero
4	cocinero	doctor
5	doctor	cocinero
6	actor	actor
7	secretario	peluquero
8	empleado	secretario
9	bombero	enfermero
10	policía	médico
11	camarero	estudiante
12	enfermero	policía
13		jardinero
14		juez
15		escritor
16		panadero
17		cantante
18		trabajar
19		bombero
20		bailarín

Tabla 133. Vocablos más disponibles del CII6 según la fecha de la prueba.

La compatibilidad es del 52,38% debido a que los lexemas comunes son once y los exclusivos de cada muestreo son uno (*empleado*) y nueve (*peluquero, estudiante, jardinero, juez, escritor, panadero, cantante, trabajar, bailarín*). De nuevo, esto parece confirmar que los encuestados al terminar el año académico pueden activar un caudal mayor de vocablos más disponibles, pasando de doce a veinte.

3.2 Conclusiones parciales

Al igual que en los otros análisis, ha surgido que los sustantivos son la clase

dominante ya que solo hemos detectado verbos en los CI08, “la escuela: muebles y materiales” (10%); CI13, “trabajos del campo y del jardín” (55%); CI16, “profesiones y oficios” (31%). Ninguna otra categoría gramatical ha aparecido en los listados que contienen los vocablos con $ID \geq 0,1$ (Anexo V).

Las temáticas de las respuestas siempre están ligadas al estímulo prototípico de los centros de interés, pese a que hayan venido a la mente de los encuestados otras asociaciones, cuyas palabras se colocan en posiciones inferiores de los registros.

La cardinalidad por CI que hemos computado se recapitula en la tabla y el gráfico.

CI	Cardinalidad del conjunto
CI01.CUE	95,45%
CI02.ROP	61,90%
CI03.CAS	64,29%
CI04.MUE	53,33%
CI05.ALI	52,63%
CI06.MES	70%
CI07.COC	60%
CI08.ESC	61,11%
CI09.ILU	30%
CI10.CIU	50%
CI11.CAM	44,44%
CI12.TRA	83,33%
CI13.TRC	57,14%
CI14.ANI	56,52%
CI15.JUE	45,83%
CI16.PRO	100%

Tabla 134. Cardinalidad del conjunto por CI según la fecha de la prueba.

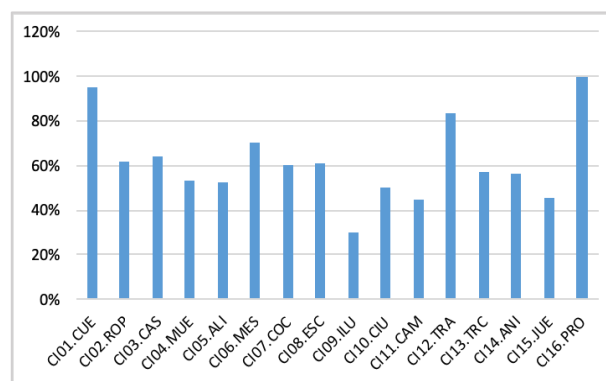


Gráfico 35. Cardinalidad del conjunto por CI según la fecha de la prueba.

Los porcentajes crecen de un 30% en el CI09, “iluminación y calefacción”, pasando por cifras elevadas como 95,45% en CI01, “partes del cuerpo”, hasta llegar al 100% en el CI16, “profesiones y oficios”.

Los lexemas más disponibles actualizados se distribuyen heterogéneamente en los distintos campos configurando grupos más cerrados o abiertos, quizás debido al tratamiento de la temática en fase de aprendizaje o al interés suscitado en los informantes.

El análisis cualitativo nos permite afirmar con mayor asentimiento que los estudiantes han potenciado su vocabulario español, el curso ha surtido algún efecto en el conocimiento compartido por el grupo-clase.

El estudio de la cardinalidad del conjunto ha corroborado el desarrollo de la competencia léxica: los encuestados son capaces de activar un número mayor de lexías con $ID \geq 0,1$. En otras palabras, tienen a disposición una cantidad mayor de léxico disponible, según la siguiente distribución:

- CI01, “partes del cuerpo”, -4,55%;
- CI02, “la ropa”, +38,18%;
- CI03, “partes de la casa (sin muebles)”, +8,33%;
- CI04, “los muebles de la casa”, +8,33%;
- CI05, “alimentos y bebidas”, -6,67%;
- CI06, “objetos colocados en la mesa para la comida”, +30%;
- CI07, “la cocina y sus utensilios”, +22,22%;
- CI08, “la escuela: muebles y materiales”, +38,89%;
- CI09, “iluminación y calefacción”, +70%;
- CI10, “la ciudad”, +22,73%;
- CI11, “el campo”, ±0%;
- CI12, “medios de transporte”, +16,67%;
- CI13, “trabajos del campo y del jardín”, +42,86%;
- CI14, “los animales”, +28,57%;
- CI15, “juegos y distracciones”, +47,83%;
- CI16, “profesiones y oficios”, +40.

Solo un centro de interés ha permanecido invariado (CI11, “el campo”) y dos han arrojado resultados inferiores (CI01, “partes del cuerpo”; CI05, “alimentos y bebidas”). Por otro lado, los demás han registrado un aumento que alcanza porcentajes notablemente

superiores que confirman lo dicho.

4. ANÁLISIS COMPARATIVO

Cotejamos los 1.490 vocablos recopilados en el corpus con las listas de frecuencia extraídas del CREA y del CAES con el objetivo de averiguar cuántas y cuáles son las unidades compartidas entre estos listados y los ítems activados por nuestros encuestados. Además de comparar las listas enteras de los dos repertorios (respectivamente 5.000 y 5.360), hemos considerado como medidas de corte las primeras 500 y 1.000 entradas para averiguar la cantidad de lexemas compartida en estos intervalos menores¹⁰⁶.

4.1 Análisis comparativo con el CREA

Procedemos a la comparación entre el léxico disponible de nuestros informantes y la lista de frecuencia proporcionada por el CREA, el *Corpus de Referencia del Español Actual*, que encierra las 5.000 palabras más frecuentes del español¹⁰⁷.

4.1.1 Análisis descriptivo

Pese a la diversidad de la conformación interna de nuestro corpus con respecto al CREA, que se compone también de palabras atemáticas, nombres propios y formas declinadas, el vocabulario disponible de los informantes está incluido en el léxico más frecuente del español, aunque de manera reducida.

Vocablos CREA	Vocablos compartidos	
	Total	Porcentaje
500	32	6,40%
1.000	76	7,60%
5.000	278	5,56%

Tabla 135. Número de vocablos compartidos con el CREA.

Además, se hace patente que el 20% de estas unidades compartidas coincide con

¹⁰⁶ Claro está que no se trata de un análisis exhaustivo ya que nuestro corpus ha sido editado y lematizado, mientras que los corpus de referencia contienen palabras funcionales, nombres propios y formas conjugadas, flexivas, derivadas, precisamente porque proceden de material auténtico.

¹⁰⁷ Nos basamos en esta cantidad puesto que debería cubrir el 90% de un texto (Alvar Ezquerro, 2003; Nation, 2001; Izquierdo Gil, 2005).

los vocablos más disponibles de nuestro corpus general, esto es, los que tienen un ID \geq 0,1. En orden alfabético son: *actor, animal, árbol, arroz, avión, bailar, bar, barco, botella, brazo, caballo, café, calle, cara, carne, carretera, cartas, cerveza, cine, cocina, cortar, cuarto, cuello, dientes, doctor, edificio, estación, estudiante, flor, fuego, fútbol, gato, habitación, horno, iglesia, jardín, leche, lengua, león, libro, luz, naranja, nariz, ojos, orejas, papel, pasta, pelo, pelota, perro, pie, pierna, planta, plato, plaza, profesor, restaurante, sal, salón, silla, sillón, taxi, televisión, tienda, tierra, toro, tren, vaso, vestido, vino.*

La repartición según categoría gramatical, como muestra el gráfico 35, sigue la tendencia surgida a lo largo de los análisis transversal y longitudinal: 88% de sustantivos; 8% de verbos; 4% de adjetivos.

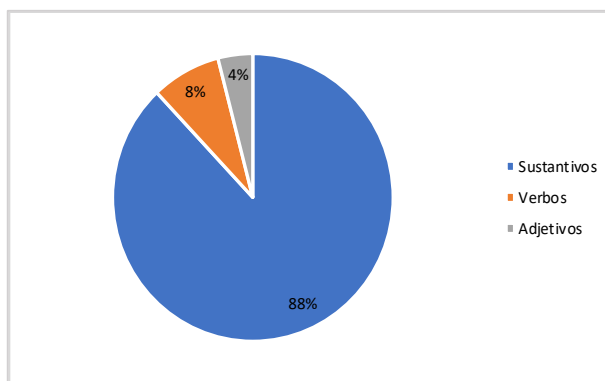


Gráfico 36. Categorías gramaticales de los vocablos compartidos con el CREA.

4.2 Análisis comparativo con el CAES

Continuamos el contraste con la lista de frecuencia procedente del CAES, el *Corpus de aprendices de español*, de la cual hemos tenido en cuenta exclusivamente los datos del nivel intermedio¹⁰⁸.

4.2.1 Análisis descriptivo

Antes de proceder hemos quitado todas las direcciones de correo electrónico que están registradas en el corpus, ya que no hacen parte del léxico español que nos interesa, y sucesivamente, hemos eliminado los datos relativos a los otros niveles lingüísticos, por

¹⁰⁸ Hemos sumado los datos relativos a los niveles B1 (umbral) y B2 (avanzado) de manera que el cotejo resultase más fiable conforme al dominio de ELE de nuestros informantes, puesto que el CAES recoge todos los niveles del MCER, del A1 (*acceso*) al C1 (*dominio operativo eficaz*).

lo que consideramos una lista compuesta por un total de 5.364 ítems.

Sobresale la mayor coincidencia de nuestro repertorio con estos datos con respecto a los del CREA.

Lista frecuencia CAES	Vocablos compartidos	
	Total	Porcentaje
500	74	14,80%
1.000	153	15,30%
5.364	559	10,42%

Tabla 136. Número de vocablos compartidos con el CAES.

Los porcentajes de vocablos compartidos casi doblan los que hemos calculado anteriormente: logramos un máximo de 15,30% al considerar la lista de las 1.000 palabras, intervalo en el que se detecta más correspondencia. Computamos cifras ligeramente inferiores en las otras dos, pero siempre mayores si las contrastamos con el CREA. De todos modos, aquí tampoco trabajamos con números elevados debido a la conformación interna de los dos listados que contienen distintas tipologías morfosintácticas.

En lo que se refiere a la subdivisión según categoría gramatical, otra vez, se detecta una presencia masiva de sustantivos que son 496 (88,73%), seguidos por 46 verbos (8,23%) y por una cantidad aún menor de adjetivos, en total 16 (2,86%) y por un solo adverbio (0,18%), como ilustra el gráfico.

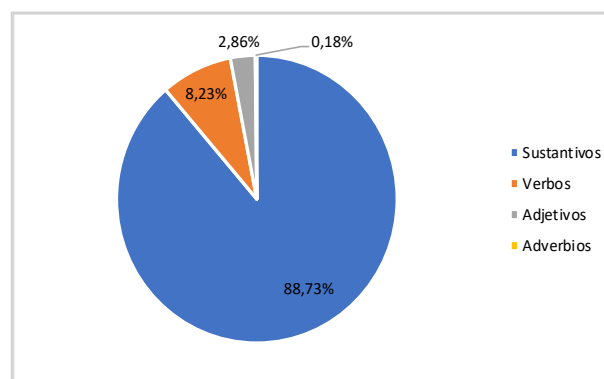


Gráfico 37. Categorías gramaticales de los vocablos compartidos con el CAES.

Sin duda, el caudal de vocablos compartidos depende del tipo de prueba a la que se han sometido los aprendices, el cual implica el tratamiento de determinados temas sobre otros. Los tests de disponibilidad léxica intentan tratar todos los intereses humanos

y ser lo más variados posible gracias a las áreas semánticas propuestas, mientras que las tareas utilizadas para la recopilación del CAES están más circunscritas:

Nivel B1

- Escribir una carta a un amigo solicitándole varios favores (175-200 palabras).
- Escribir una carta con una queja dirigida a una compañía aérea por pérdida de equipaje (175-200 palabras).
- Escribir una historia (real o imaginaria) que sea graciosa (175-200 palabras).

Nivel B2

- Redactar una solicitud de admisión en un programa universitario motivando la petición y preguntando sobre las características del programa (275-300 palabras).
- Escribir un ensayo sobre la importancia de las nuevas tecnologías en la sociedad actual (275-300 palabras). (Instituto Cervantes, CAES)

4.3 Conclusiones del análisis comparativo

De la comparación entre los vocablos arrojados en las encuestas con la lista frecuencia proporcionada por el *Corpus de Referencia del Español Actual* ha surgido que un reducido porcentaje de las respuestas de los informantes coincide con el léxico más frecuente del español: no obstante se hayan revelado porcentajes bajos hemos visto que hay cierta correspondencia.

En segundo lugar, el cotejo con el *Corpus de aprendices de español* ha revelado una mayor coincidencia entre las entradas, que se desprende al mirar las tablas 135 y 136.

De todas formas, se ha manifestado una tendencia común en el análisis ya que, a pesar de las diferencias de los porcentajes, en ambos cotejos hemos observado que al tener en cuenta 1.000 palabras se detectan más vocablos compartidos que si se consideran solo las primeras 500 o si la comparación se extiende a 5.000.

Por último, si consideramos que ambos corpus de referencia están formados también por unidades funcionales del discurso y formas declinadas, mientras que nuestra lista ha sido enmendada cuidadosamente, es obvio que los porcentajes de vocablos coincidentes aumentarían en gran medida si no hubiésemos lematizado el corpus, destacando que el vocabulario de los participantes encaja en los repertorios de palabras más frecuentes sea de los hispanohablantes nativos sea de otros aprendices de ELE.

5. CONCLUSIONES DEL ANÁLISIS CUALITATIVO DE LA DISPONIBILIDAD LÉXICA

Para este capítulo nos hemos basado tanto en la tipología gramatical de las respuestas para averiguar cuál es la clase con mayor número de apariciones, como en las

temáticas predominantes, para indagar si cumplían, y en qué medida, con los estímulos principales de cada centro de interés o si procedían de relaciones mentales secundarias.

La categoría gramatical que prevalece es la de los sustantivos, los verbos aparecen solo en ciertos campos y en cantidad menor (sobre todo en el CI13, “trabajos del campo y del jardín”, y en el CI15, “juegos y distracciones”), los adjetivos se manifiestan aun menos y no hemos detectado ninguna otra clase de palabras. Se confirma otra vez que

[...] en el léxico disponible la gran mayoría de los vocablos son sustantivos, salvo algunos verbos, adverbios y adjetivos –debido, fundamentalmente, a la metodología empleada y al tipo de centros de interés seleccionados–. (Sánchez-Saus, 2009: 143)

Los vocablos más disponibles se adscriben a los ámbitos prototípicos de cada área ya que solo en los rangos más bajos de los listados aparecen unidades referentes a otras asociaciones.

En lo que atañe al análisis general, el número de vocablos oscila entre tres y veinticinco. El CI05, “alimentos y bebidas”; el CI01, “partes del cuerpo”; el CI10, “la ciudad”, presentan la cantidad mayor de lexías con $ID \geq 0,1$. Del otro lado, el CI09, “iluminación y calefacción”, y el CI13, “trabajos del campo y del jardín”, tienen la cantidad menor.

Al analizar los resultados por variable, se nota la supremacía de las mujeres sobre los hombres, de los estudiantes de nivel B2 sobre los B1 y, por último, de los que estudian más de dos LE sobre los que solo están aprendiendo una.

Análogamente, en el análisis longitudinal, hemos observado que los encuestados han activado más *types* con $ID \geq 0,1$ en la segunda prueba, han ampliado su bagaje léxico y el curso del español al que han asistido parece haber sido fructífero.

Como colofón, hemos proporcionado una comparación de los datos entre las variantes de cada variable y entre las dos suministraciones de la prueba con el fin de medir el grado de cardinalidad del conjunto, es decir la compatibilidad entre las respuestas. Esta correlación ha destacado los puntos en común y ha permitido corroborar si el conocimiento se debe a un aprendizaje estructurado gracias a la enseñanza formal o no, como sigue.

En ciertos CI se ha alcanzado el 100% de compatibilidad:

- en el CI12, “los medios de transporte”, según el factor sexo;
- en CI08, “la escuela: muebles y materiales”; CI09, “iluminación y

- calefacción”; CI12, “medios de transporte”, según el nivel de ELE;
- en CI06, “objetos colocados en la mesa para la comida”; CI09, “iluminación y calefacción”, según el conocimiento de otras LE.

Al contrario, se han contabilizado porcentajes bajos en los CI:

- CI09, “iluminación y calefacción”, en el cotejo entre el género (0%);
- CI07, “la cocina y sus utensilios”, en la comparación entre niveles (30%);
- CI07, “la cocina y sus utensilios”, según las LE conocidas (37,50%).

En el apartado del análisis longitudinal, los porcentajes de compatibilidad del conjunto varían del 30% en el CI09, “iluminación y calefacción”, pasando por cifras elevadas como el 95,45% en CI01, “partes del cuerpo”, hasta llegar al 100% en el CI16, “profesiones y oficios”. Se supone que estos valores dependen del espacio dedicado a determinados temas a lo largo de la enseñanza-aprendizaje de ELE ya que es posible que los ámbitos más cerrados y compactos deriven de un proceso escolar sistematizado, mientras que los más abiertos y difusos sean fruto de una adquisición cumplida por iniciativa de los aprendices, por la que cada uno ha asimilado palabras distintas.

El cotejo final con el CREA destaca que la lista de frecuencia analizada contiene los vocablos activados por nuestros participantes, con lo cual cierto vocabulario que tienen a su disposición pertenece al léxico más comúnmente utilizado en español. De igual manera, la comparación realizada con las palabras más frecuentes de otros aprendices de ELE, extraídas del CAES, ha evidenciado una correspondencia mayor con la producción de nuestros encuestados.

CAPÍTULO VI.

ANÁLISIS CUANTITATIVO DE LA RIQUEZA LÉXICA

1. INTRODUCCIÓN

A partir de los índices que hemos elegido, analizamos la riqueza léxica producida en el test realizado por nuestros informantes para añadir información complementaria sobre su competencia, además de la obtenida en el apartado de disponibilidad.

El análisis transversal abre el capítulo con la discusión de los resultados procedentes de los datos de la variación y la densidad del vocabulario. En segunda instancia, discutimos los resultados por variable (sexo, nivel de ELE, conocimiento de otras LE) siguiendo el mismo proceso analítico y extendemos el análisis mediante el estudio de los datos estadísticos descriptivos que informan sobre la simetría o la dispersión del comportamiento intragrupal e intergrupala de los sujetos (Anexo VII).

Igualmente, el análisis longitudinal coteja los resultados de la primera prueba con los de la segunda suministradas a los alumnos de nivel B2 para comprobar si ha habido un incremento del bagaje léxico (Anexo VIII)¹⁰⁹.

El análisis comparativo concluye el capítulo: comparamos nuestros datos con los de cuatro investigaciones de la misma índole, centradas en otros estudiantes no nativos de español con el objetivo de examinar si aprendices de ELE de igual dominio lingüístico y de distinta procedencia (LM, entorno cultural y educativo) presentan la misma riqueza léxica en su producción escrita.

2. ANÁLISIS TRANSVERSAL

2.1 Resultados generales

Los datos generales extraídos de los cómputos expresan la variedad y la densidad léxica a partir de los siguientes valores repartidos en el conjunto textual.

¹⁰⁹ De manera parecida, Laufer (1991) realiza en un estudio longitudinal sobre el desarrollo de la riqueza léxica de aprendices de inglés L2.

Textos	<i>Tokens</i>	<i>Types</i>	Promedio de <i>tokens</i>	Promedio de <i>types</i>
100	10.000	1.072	100	63

Tabla 137. Datos generales de RL.

Los textos redactados por los informantes reúnen un corpus de 10.000 *tokens* entre los que los *types* son 1.072. En detalle, los índices se distribuyen como sigue en cada relato.

Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,64	64%	1,23	48%	2,08
2	0,53	53%	1,51	55%	1,82
3	0,66	66%	1,43	56%	1,79
4	0,62	62%	1,24	50%	2,00
5	0,61	61%	1,27	59%	1,69
6	0,57	57%	1,50	55%	1,82
7	0,60	60%	1,43	59%	1,69
8	0,60	60%	1,43	55%	1,82
9	0,60	60%	1,43	55%	1,82
10	0,64	64%	1,36	57%	1,75
11	0,62	62%	1,32	59%	1,69
12	0,65	65%	1,48	60%	1,67
13	0,63	63%	1,40	52%	1,92
14	0,59	59%	1,51	55%	1,82
15	0,68	68%	1,39	63%	1,59
16	0,53	53%	1,77	52%	1,92
17	0,65	65%	1,35	53%	1,89
18	0,61	61%	1,45	43%	2,33
19	0,53	53%	1,47	51%	1,96
20	0,66	66%	1,14	50%	2,00
21	0,59	59%	1,51	53%	1,89
22	0,65	65%	1,27	52%	1,92
23	0,59	59%	1,51	52%	1,92
24	0,66	66%	1,27	55%	1,82

25	0,71	71%	1,18	56%	1,79
26	0,67	67%	1,31	55%	1,82
27	0,59	59%	1,40	55%	1,82
28	0,56	56%	1,65	60%	1,67
29	0,58	58%	1,41	55%	1,82
30	0,64	64%	1,56	60%	1,67
31	0,62	62%	1,07	53%	1,89
32	0,75	75%	1,29	64%	1,56
33	0,64	64%	1,25	53%	1,89
34	0,61	61%	1,39	48%	2,08
35	0,64	64%	1,25	56%	1,79
36	0,69	69%	1,41	66%	1,52
37	0,59	59%	1,55	51%	1,96
38	0,69	69%	1,33	54%	1,85
39	0,63	63%	1,34	50%	2,00
40	0,54	54%	1,69	49%	2,04
41	0,63	63%	1,54	61%	1,64
42	0,70	70%	1,27	58%	1,72
43	0,61	61%	1,42	60%	1,67
44	0,62	62%	1,38	59%	1,69
45	0,57	57%	1,43	54%	1,85
46	0,66	66%	1,35	60%	1,67
47	0,66	66%	1,29	60%	1,67
48	0,64	64%	1,36	57%	1,75
49	0,65	65%	1,25	55%	1,82
50	0,61	61%	1,45	54%	1,85
51	0,67	67%	1,37	61%	1,64
52	0,54	54%	1,64	50%	2,00
53	0,62	62%	1,38	50%	2,00
54	0,60	60%	1,50	54%	1,85
55	0,59	59%	1,31	53%	1,89
56	0,62	62%	1,41	53%	1,89
57	0,55	55%	1,45	48%	2,08
58	0,65	65%	1,38	63%	1,59

59	0,71	71%	1,25	57%	1,75
60	0,66	66%	1,38	56%	1,79
61	0,61	61%	1,39	48%	2,08
62	0,67	67%	1,31	60%	1,67
63	0,62	62%	1,32	53%	1,89
64	0,64	64%	1,33	55%	1,82
65	0,61	61%	1,45	58%	1,72
66	0,63	63%	1,50	55%	1,82
67	0,54	54%	1,59	50%	2,00
68	0,66	66%	1,29	57%	1,75
69	0,68	68%	1,28	59%	1,69
70	0,53	53%	1,56	49%	2,04
71	0,65	65%	1,27	52%	1,92
72	0,64	64%	1,28	52%	1,92
73	0,67	67%	1,34	52%	1,92
74	0,57	57%	1,46	50%	2,00
75	0,73	73%	1,30	56%	1,79
76	0,66	66%	1,47	55%	1,82
77	0,70	70%	1,37	61%	1,64
78	0,58	58%	1,41	46%	2,17
79	0,56	56%	1,65	55%	1,82
80	0,59	59%	1,40	58%	1,72
81	0,72	72%	1,26	59%	1,69
82	0,67	67%	1,34	59%	1,69
83	0,64	64%	1,33	58%	1,72
84	0,67	67%	1,29	52%	1,92
85	0,61	61%	1,45	54%	1,85
86	0,65	65%	1,41	57%	1,75
87	0,63	63%	1,31	49%	2,04
88	0,66	66%	1,27	53%	1,89
89	0,65	65%	1,30	48%	2,08
90	0,61	61%	1,42	53%	1,89
91	0,63	63%	1,29	52%	1,92
92	0,58	58%	1,53	54%	1,85

93	0,60	60%	1,50	57%	1,75
94	0,69	69%	1,35	57%	1,75
95	0,67	67%	1,34	58%	1,72
96	0,66	66%	1,29	55%	1,82
97	0,67	67%	1,29	54%	1,85
98	0,59	59%	1,48	56%	1,79
99	0,67	67%	1,37	57%	1,75
100	0,61	61%	1,39	50%	2,00
PROMEDIO	0,63	63%	1,39	55%	1,84

Tabla 138. Índices de RL.

2.1.1 Diversidad léxica

La primera etapa del análisis consiste en aplicar las fórmulas que diferencian los *tokens* de los *types* y aportan informaciones sobre la variedad del corpus y de cada texto:

- (i) *Type/Token Ratio*;
- (ii) variación léxica;
- (iii) índice de hápax.

(i) TTR

Analizamos la relación entre las palabras y los vocablos, recordando que el índice de la *Type/Token Ratio* puede variar, idealmente, de 0 a 1.

El valor 1 es improbable de alcanzar en un escrito que no sea una oración ya que implica que no se repita ninguna palabra. Análogamente, es imposible bajar hasta 0 porque siempre tiene que aparecer un *type* como mínimo de manera que un texto exista. Por ende, cuanto más el resultado se acerca a 1, tanto más el índice revela riqueza léxica.

En este caso, el mínimo se ha calculado en cuatro textos (2, 16, 19, 70) que no sobrepasan el 0,53; el máximo se detecta en el texto 32 que logra un 0,75; la media es de 0,63 y se ha alcanzado en seis textos (13, 39, 41, 66, 87, 91).

Es posible dividir el corpus en tres franjas en función de la posición de los textos, si se encuentran por debajo o por encima del promedio.

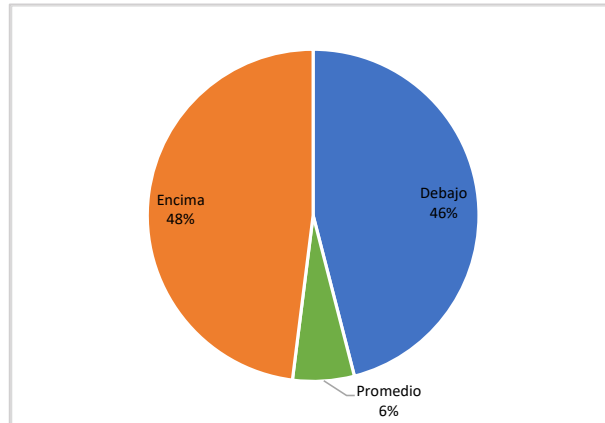


Gráfico 38. Distribución de los textos según el promedio del TTR.

El gráfico indica que el 46% de las aportaciones se coloca bajo la media mientras que el 48% la supera, la oscilación va de 0,64 a 0,75. El TTR es bueno: sobre un total de 100 ítems que componen el texto, más de cincuenta palabras son diferentes.

(ii) Variación léxica

Si traducimos en porcentaje estos datos obtenemos el índice de la variación léxica, pues corresponde a la *Types/Tokens Ratio* expresada en tantos por ciento. Los valores mínimo y máximo y el promedio coinciden con los vistos arriba: respectivamente 53%, 75% y 63%.

El gráfico 38 representa la misma distribución en tres ramos.

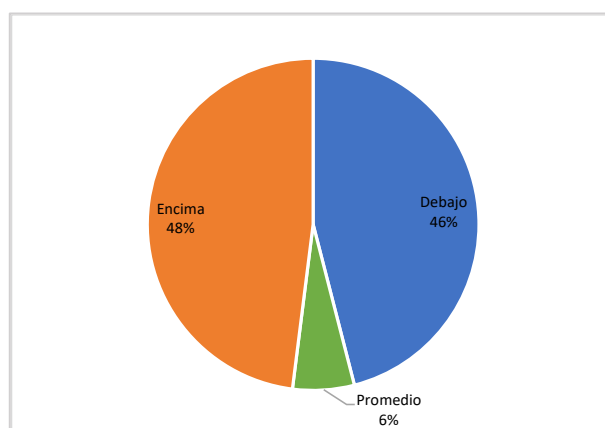


Gráfico 39. Distribución de los textos según el promedio de la variación léxica.

El resultado de los cálculos es satisfactorio, dado el nivel lingüístico intermedio de los participantes ya que más de la mitad de las lexías aportadas en cada producción son unidades que no se repiten.

El corpus se divide en tres secciones: una que presenta una variedad de

vocabulario media que va del 53% al 59% (veinticuatro textos), otra encierra la gran mayoría de los textos que han alcanzado una diversidad comprendida entre el 60% y el 69% (sesenta nueve textos) y la última se compone de los resultados más sobresalientes que han conseguido hasta un 75% de variación (siete textos).

(iii) Índice de hápax

Las palabras hápax, los lexemas cuya ocurrencia es 1 dentro de un relato, constituyen aquí un dato relevante puesto que en promedio se han contabilizado 46 sobre una media de 63 *types* en el total de 100 palabras por texto.

Destaca una gran oscilación de valores, hasta el doble, que supone una capacidad potencial de variación de vocabulario muy diferente dentro del grupo: el número mínimo de hápax es 30 (texto 25) y el máximo es 60 (texto 16).

El texto 31 logra el resultado menor (1,07), donde se cuentan 58 hápax en un total de 62 *types*; mientras que el texto 16 alcanza el mayor (1,77), en el que los hápax son 30 y los vocablos 53. El promedio es de 1,39 (textos 15, 34, 61, 100).

A partir de estas cifras, las producciones se distribuyen como sigue.

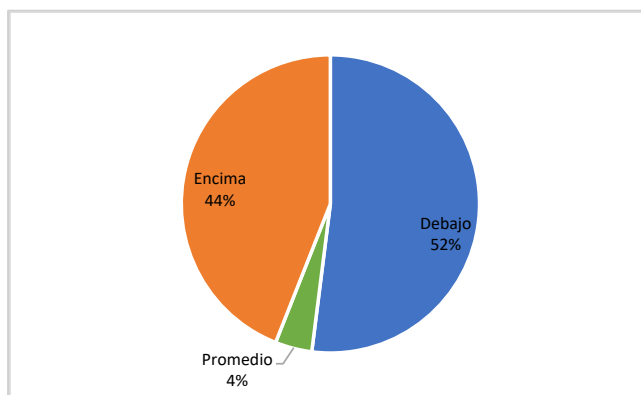


Gráfico 40. Distribución de los textos según el promedio del índice de hápax.

El 44% de las aportaciones supera la media, que se ha computado en cuatro textos (4%), y el 52% queda por debajo, lo que implica que la mayoría de los informantes ha elaborado textos más ricos.

Como esta medida es inversamente proporcional a la riqueza léxica (siendo la suma de los hápax el denominador de la fórmula) cuanto más el índice se acerca a 1, tanto más el vocabulario es rico porque corresponde a la condición en la que los *types* son hápax al mismo tiempo. Por ende, los datos revelan un índice medio-bajo y un grado alto de riqueza ya que la mayoría de los valores se aproximan más a 1 que a 2 (hemos tomado

como medida de corte el valor 2, si bien no se haya establecido un límite máximo universal, porque destaca cuanto el valor se aleja del ideal). Aún más, gran parte de los textos (81%) está bajo el 1,50 que es equidistante de los puntos de referencia.

2.1.2 Densidad léxica

La segunda fase analítica se centra en el uso de palabras nocionales, basándose en la relación entre *tokens* y *tokens* léxicos expresada por los índices de:

- (i) densidad léxica;
- (ii) IAT.

(i) Densidad léxica

La densidad léxica computada revela que más de la mitad de las palabras de un texto es nocional. Este resultado es notable si consideramos que solo el 10% de los textos no alcanza el 50%, el 8% lo iguala, la mayoría (82%) lo sobrepasa.

El corpus se divide en tres grupos conforme a los intervalos porcentuales calculados: diez textos manifiestan datos más pobres de 43% a 49%; 76 textos tienen una densidad de 50% a 59%; catorce textos consiguen un rendimiento mayor, de 60% a 66%.

Dentro de estos subconjuntos se ha contabilizado un mínimo de 43% del texto 18; un promedio de 55% en los textos 2, 6, 8, 9, 14, 24, 26, 27, 29, 49, 64, 66, 76, 79, 96; un máximo de 66% en el texto 36.

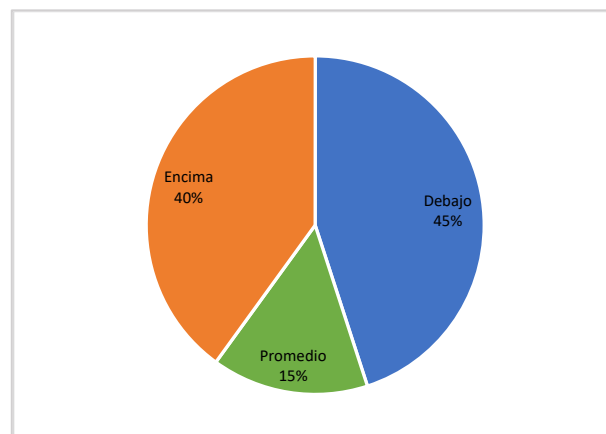


Gráfico 41. Distribución de los textos según el promedio de la densidad léxica.

Según estos datos, hay tres franjas en las que se colocan los textos según la media. El 45% presenta valores pequeños, debajo del promedio que se compone del 15% de las aportaciones totales. Aunque todos han obtenido resultados satisfactorios, el 40% se pone

encima, lo cual significa que son cuarenta los estudiantes que retienen un vocabulario más denso y rico.

(ii) IAT

En cuanto al IAT, se ha revelado un empleo reducido de las palabras nocionales a favor de las funcionales: la media es de 1,84. Esto indica la presencia de un ítem léxico casi cada dos gramaticales y no representa un alto grado de riqueza ya que el resultado mejor se alcanza cuando el intervalo se acerca a 1: “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 Díaz, 2010: 148).

El texto 36 ha obtenido el índice más reducido (1,52) y, por tanto, el uso de lexemas semánticos mayor con respecto a los otros. Por el contrario, el que más se distancia del valor ideal es el relato 18, donde el IAT es 2,33.

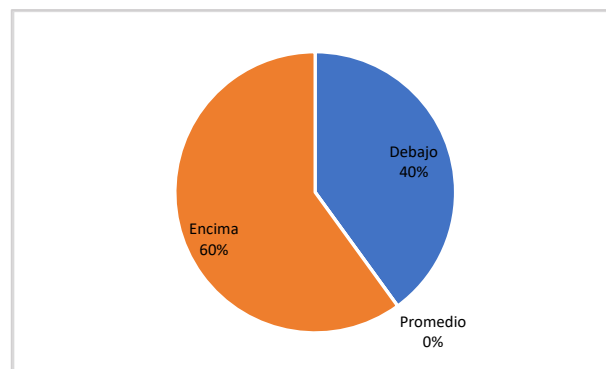


Gráfico 42. Distribución de los textos según el promedio del IAT.

El 40% de los encuestados ha escrito textos que se quedan debajo de la media, proporcionan datos escasos puesto que las cifras procedentes de los cómputos se acercan más a 2 que a 1. Análogamente, aquellos que están por encima (60%) consiguen un resultado aun peor, que supera el 2.

2.2 Conclusiones parciales

Pese al buen grado general de riqueza surgido del análisis de la diversidad léxica, no ha ocurrido lo mismo en lo que se refiere a la densidad.

Los resultados del análisis de la variedad ponen de manifiesto que (i) todos los escritos tienen una buena variación de vocabulario ya que en cada uno se ha detectado más del 50% de unidades léxicas diferentes; (ii) el índice de hápax también manifiesta un

buen nivel de riqueza léxica porque los productos de la fórmula han llevado a cifras medio-bajas.

Es interesante observar la correlación de estos resultados puesto que el texto menos rico según estos indicadores es el 16 porque presenta, a la vez, los valores más bajos de TTR y variación (0,53 y 53%) y el índice mayor de hápax (1,77). Por el contrario, el texto 32 alcanza los mejores TTR y variación (0,75 y 75%) pero los hápax no logran el mismo resultado, aunque sea reducido (1,29).

El estudio del uso de los *tokens* léxico aporta resultados divergentes. El 90% de los informantes ha alcanzado una buena densidad léxica, más de la mitad de las palabras de un texto es temática, lo cual indica un rendimiento satisfactorio (la media es de 55%). Los cálculos del IAT no son igualmente notables ya que se parte de un valor mínimo de 1,52 hasta llegar a un máximo de 2,33 (la media es de 1,84).

De nuevo, sacamos a colación la correspondencia entre los resultados de los relatos que se han revelado los más y menos ricos: el texto 36 ha conquistado el nivel mayor de densidad (66%) y menor de IAT (1,52); por el contrario, el 18 ha conseguido poca riqueza al obtener el valor mínimo de densidad (43%) y máximo de IAT (2,33).

2.3 Resultados por variable

En este apartado estudiamos la riqueza léxica en función de las tres variables sociolingüísticas contempladas aplicando las mismas medidas utilizadas antes y añadimos el análisis de los descriptores estadísticos.

2.3.1 Variable “sexo”

Recordamos que el número de encuestados de los dos conjuntos no es homogéneo. Llevamos a cabo la comparación de las dos variantes según los datos de los promedios.

Informantes	<i>Tokens</i>	<i>Types</i>	Promedio de <i>tokens</i>	Promedio de <i>types</i>
Mujer (87)	8.700	997	100	63
Hombre (13)	1.300	358	100	62

Tabla 139. Datos generales de RL según la variable “sexo”.

Aunque no utilizamos los datos peculiares de los textos, presentamos los resultados de cada uno en las tablas 140 y 141, a partir de los cuales procedemos al estudio

de los datos descriptivos estadísticos para establecer el grado de riqueza de las dos muestras.

Mujer					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,64	64%	1,23	48%	2,08
2	0,53	53%	1,51	55%	1,82
3	0,57	57%	1,50	55%	1,82
4	0,60	60%	1,43	59%	1,69
5	0,60	60%	1,43	55%	1,82
6	0,60	60%	1,43	55%	1,82
7	0,62	62%	1,32	59%	1,69
8	0,65	65%	1,48	60%	1,67
9	0,63	63%	1,40	52%	1,92
10	0,59	59%	1,51	55%	1,82
11	0,68	68%	1,39	63%	1,59
12	0,53	53%	1,77	52%	1,92
13	0,65	65%	1,35	53%	1,89
14	0,61	61%	1,45	43%	2,33
15	0,53	53%	1,47	51%	1,96
16	0,66	66%	1,14	50%	2,00
17	0,59	59%	1,51	53%	1,89
18	0,65	65%	1,27	52%	1,92
19	0,59	59%	1,51	52%	1,92
20	0,66	66%	1,27	55%	1,82
21	0,71	71%	1,18	56%	1,79
22	0,67	67%	1,31	55%	1,82
23	0,59	59%	1,40	55%	1,82
24	0,56	56%	1,65	60%	1,67
25	0,58	58%	1,41	55%	1,82
26	0,64	64%	1,56	60%	1,67
27	0,62	62%	1,07	53%	1,89
28	0,75	75%	1,29	64%	1,56
29	0,64	64%	1,25	53%	1,89
30	0,61	61%	1,39	48%	2,08
31	0,64	64%	1,25	56%	1,79
32	0,69	69%	1,41	66%	1,52
33	0,59	59%	1,55	51%	1,96
34	0,69	69%	1,33	54%	1,85

35	0,63	63%	1,34	50%	2,00
36	0,54	54%	1,69	49%	2,04
37	0,63	63%	1,54	61%	1,64
38	0,70	70%	1,27	58%	1,72
39	0,61	61%	1,42	60%	1,67
40	0,62	62%	1,38	59%	1,69
41	0,57	57%	1,43	54%	1,85
42	0,66	66%	1,35	60%	1,67
43	0,66	66%	1,29	60%	1,67
44	0,64	64%	1,36	57%	1,75
45	0,65	65%	1,25	55%	1,82
46	0,61	61%	1,45	54%	1,85
47	0,67	67%	1,37	61%	1,64
48	0,60	60%	1,50	54%	1,85
49	0,59	59%	1,31	53%	1,89
50	0,55	55%	1,45	48%	2,08
51	0,65	65%	1,38	63%	1,59
52	0,61	61%	1,39	48%	2,08
53	0,67	67%	1,31	60%	1,67
54	0,62	62%	1,32	53%	1,89
55	0,63	63%	1,50	55%	1,82
56	0,66	66%	1,29	57%	1,75
57	0,68	68%	1,28	59%	1,69
58	0,53	53%	1,56	49%	2,04
59	0,65	65%	1,27	52%	1,92
60	0,64	64%	1,28	52%	1,92
61	0,67	67%	1,34	52%	1,92
62	0,57	57%	1,46	50%	2,00
63	0,73	73%	1,30	56%	1,79
64	0,66	66%	1,47	55%	1,82
65	0,70	70%	1,37	61%	1,64
66	0,58	58%	1,41	46%	2,17
67	0,56	56%	1,65	55%	1,82
68	0,59	59%	1,40	58%	1,72
69	0,72	72%	1,26	59%	1,69
70	0,67	67%	1,34	59%	1,69
71	0,64	64%	1,33	58%	1,72
72	0,67	67%	1,29	52%	1,92
73	0,61	61%	1,45	54%	1,85
74	0,65	65%	1,41	57%	1,75

75	0,63	63%	1,31	49%	2,04
76	0,66	66%	1,27	53%	1,89
77	0,65	65%	1,30	48%	2,08
78	0,61	61%	1,42	53%	1,89
79	0,63	63%	1,29	52%	1,92
80	0,58	58%	1,53	54%	1,85
81	0,60	60%	1,50	57%	1,75
82	0,69	69%	1,35	57%	1,75
83	0,67	67%	1,34	58%	1,72
84	0,66	66%	1,29	55%	1,82
85	0,67	67%	1,29	54%	1,85
86	0,59	59%	1,48	56%	1,79
87	0,67	67%	1,37	57%	1,75
PROMEDIO	0,63	63%	1,39	55%	1,83

Tabla 140. Índices de RL según la variante “mujer”.

Hombre					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,66	66%	1,43	56%	1,79
2	0,62	62%	1,24	50%	2,00
3	0,61	61%	1,27	59%	1,69
4	0,64	64%	1,36	57%	1,75
5	0,54	54%	1,64	50%	2,00
6	0,62	62%	1,38	50%	2,00
7	0,62	62%	1,41	53%	1,89
8	0,71	71%	1,25	57%	1,75
9	0,66	66%	1,38	56%	1,79
10	0,64	64%	1,33	55%	1,82
11	0,61	61%	1,45	58%	1,72
12	0,54	54%	1,59	50%	2,00
13	0,61	61%	1,39	50%	2,00
PROMEDIO	0,62	62%	1,39	54%	1,86

Tabla 141. Índices de RL según la variante “hombre”.

Tendremos en cuenta la media, la mediana, los valores mínimo y máximo de forma que el desnivel entre el número de componentes femeninos y masculinos del grupo no influya en el análisis.

2.3.1.1 Diversidad léxica

(i) TTR

Partimos de los siguientes descriptivos estadísticos que revelan una relación similar entre *tokens* y *types* en los textos de mujeres y hombres ya que no se detecta una gran discrepancia entre las secciones.

Descriptivos	Variante	
	Mujer	Hombre
Media	0,63	0,62
Mediana	0,63	0,62
Mínimo	0,53	0,54
Máximo	0,75	0,71

Tabla 142. Estadísticos descriptivos del TTR para la variable “sexo”.

El índice es satisfactorio ya que los datos se acercan más a 1 que a 0. Se observa, en particular, una ligera superioridad de las informantes en el valor máximo (0,75) que superan al grupo de hombres (0,71).

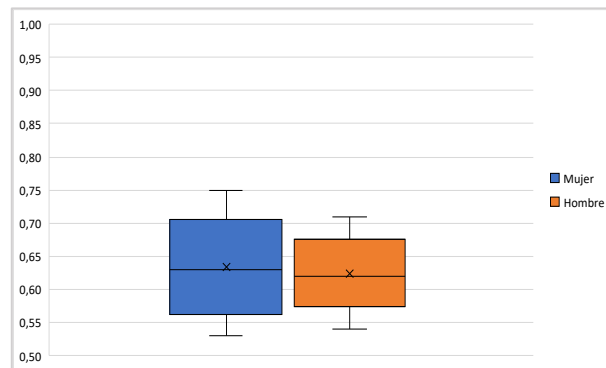


Gráfico 43. Diagrama de cajas del TTR para la variable “sexo”.

Las respuestas de las mujeres ponen de relieve una desviación intragrupal más elevada: se ve una mayor extensión de los bigotes de la caja azul, que marca una oscilación superior de los datos con respecto a la caja naranja.

(ii) Variación léxica

Al traducir en porcentaje estos datos, observamos los siguientes índices de variación léxica. Obviamente, la condición de los datos concuerda con la que acabamos de discutir.

Descriptivos	Variante	
	Mujer	Hombre
Media	63%	62%
Mediana	63%	62%
Mínimo	53%	54%
Máximo	75%	71%

Tabla 143. Estadísticos descriptivos de la variación léxica para la variable “sexo”.

Las mujeres tocan el vértice mayor (75%), el menor (53%) y el promedio sigue a su favor (63%) ya que cambia de un punto con respecto de la media alcanzada por los informantes masculinos (62%).

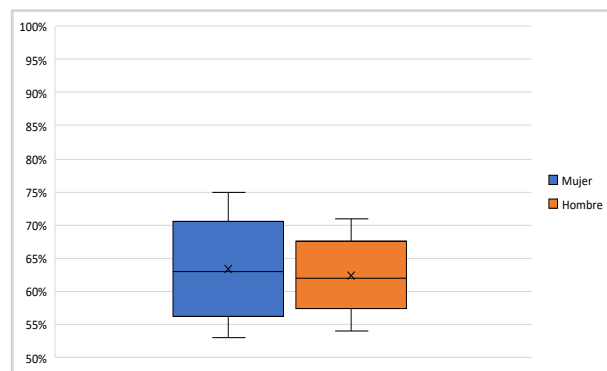


Gráfico 44. Diagrama de cajas de la variación léxica para la variable “sexo”.

Al examinar pormenorizadamente los valores de TTR y de variación léxica de cada relato de las muestras, se hace patente que todos han alcanzado más del 50% de variedad textual (el índice mínimo es 53%), con lo cual el grado de repetición de lexemas es bastante bajo: los porcentajes de *types* son mayores en relación con los *tokens* totales.

(iii) Índice de hápax

El índice de hápax manifiesta una notable oscilación de los resultados en general, pero si tomamos como punto de partida las medias y las medianas no hay diferencia entre las dos variantes.

Descriptivos	Variante	
	Mujer	Hombre
Media	1,39	1,39
Mediana	1,38	1,38
Mínimo	1,07	1,24

Máximo	1,77	1,64
--------	------	------

Tabla 144. Estadísticos descriptivos del índice de hápax para la variable “sexo”.

Las estudiantes arrojan los valores más altos (1,77) y bajos (1,07); mientras que los indicadores medios están compartidos por las dos secciones (1,39 y 1,38). Las mujeres obtienen el índice menor y, sin embargo, mejor, en el texto 131 donde se hallan 58 hápax sobre un total de 62 *types* (93%) mientras que el texto 116 resulta el menos rico ya que se ha computado un índice de 1,77: hay 30 hápax y 53 vocablos (56%).

Fijémonos ahora en la distribución gráfica de los datos.

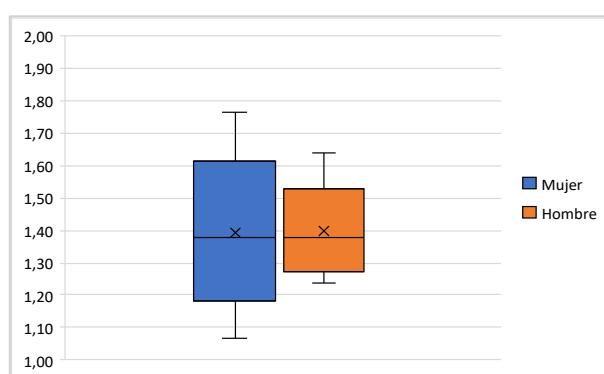


Gráfico 45. Diagrama de cajas del índice de hápax para la variable “sexo”.

Puesto que cuanto más pequeño es este índice, más rico es un texto, las mujeres han aportado un caudal mayor de hápax en sus escritos como indican las partes inferiores de los diagramas, en las que se colocan los resultados mejores (la extensión del bigote, que baja hasta el valor mínimo, y del Q1, que sube hasta la mediana). Los hombres presentan un bigote más corto, que implica una menor cantidad de palabras “originales”. Por otra parte, la presencia de hápax en la producción de los varones resulta más compacta y la desviación es reducida (de 1,69 a 2). Por último, hemos contabilizado que el 85% de los textos de las mujeres y el 85% de los relatos de los hombres tienen un índice que no supera el 1,50. Esto indica que no se detecta ninguna supremacía.

De acuerdo con esto, los resultados son satisfactorios en ambos casos.

2.3.1.2 Densidad léxica

(i) Densidad léxica

La densidad léxica mide el porcentaje de palabras llenas en un texto, que se reparten como sigue según los datos estadísticos.

Descriptivos	Variante	
	Mujer	Hombre
Media	55%	54%
Mediana	55%	55%
Mínimo	43%	50%
Máximo	66%	59%

Tabla 145. Estadísticos descriptivos de la densidad léxica para la variable “sexo”.

La aplicación de la fórmula es significativa, por lo general, ya que la mayoría de las producciones alcanza o supera el 50%. Las mujeres siguen presentando el valor superior (66%) y el inferior (43%). Los promedios corresponden (55%), por tanto, la riqueza léxica de la producción de los dos grupos se reparte equitativamente.

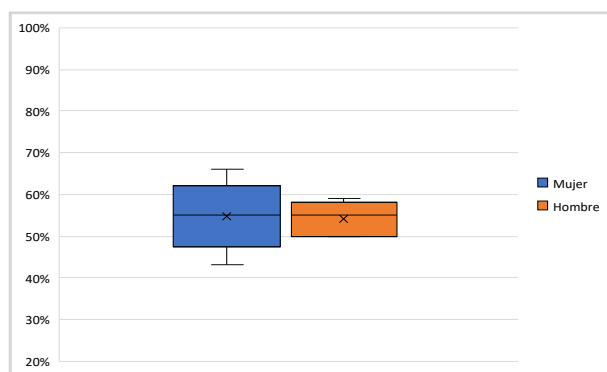


Gráfico 46. Diagrama de cajas de la densidad léxica para la variable “sexo”.

Se nota una ligera diferencia entre los valores medios: media y mediana de las mujeres se solapan (55%); mientras que en los hombres la media matemática (54%) se posiciona por debajo de la línea de la mediana de un punto (55%).

La componente masculina presenta un índice más compacto (de 50% a 59%) frente a la otra, donde es más difuso y la desviación es mayor debido a resultados más variados (de 43% a 66%). El intervalo de datos entre cuartiles es más amplio, de ahí que la caja sea más alargada.

Al desglosar los resultados de cada relato, se ha contabilizado que el 100% de los escritos de los hombres igualan o sobrepasan el 50% frente al 89% de las producciones femeninas: la mayoría de los varones ha aportado una cantidad más elevada de palabras semánticas, aunque sea por poco. Esto se ve bien en el gráfico: el diagrama que representa los resultados de las mujeres se extiende hasta el extremo inferior a partir del Q1 (que concuerda con el límite del diagrama masculino), donde se hallan índices más pequeños.

(ii) IAT

Los datos del intervalo de aparición de palabras nocionales se dividen en los siguientes datos.

Descriptivos	Variante	
	Mujer	Hombre
Media	1,83	1,86
Mediana	1,82	1,82
Mínimo	1,52	1,69
Máximo	2,33	2,00

Tabla 146. Estadísticos descriptivos del IAT para la variable “sexo”.

Otra vez, las mujeres ofrecen simultáneamente el mejor (1,52) y el peor (2,33) resultado. Las medianas coinciden entre los dos muestreos (1,82) y las medias cambian, de poco, favoreciendo las encuestadas ya que el promedio del IAT es levemente más pequeño del índice de los compañeros.

El gráfico ayuda a organizar la información y la comprensión de estos datos.

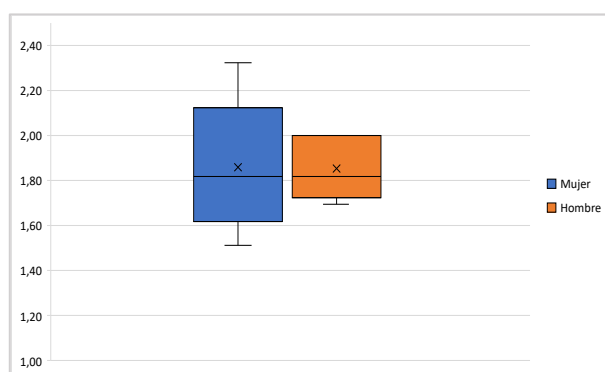


Gráfico 47. Diagrama de cajas del IAT para la variable “sexo”.

Se ve que media y mediana manifiestan simetría entre las dos secciones. La caja naranja se solapa con la parte central de la azul poniendo de relieve la desviación interior menor de los varones (de 1,69 a 2) y mayor de las alumnas que logran cifras más variadas (de 1,52 a 2,33).

En todo caso, el grado de riqueza léxica no es alto porque los resultados de la fórmula se acercan más a 2 que a 1, el valor ideal¹¹⁰. Esto se entiende bien observando el gráfico: los intervalos que van del Q2 (la línea de la mediana) al Q3 y suben hasta el

¹¹⁰ Hemos tomado esta medida de corte, tal cual hemos planteado en el análisis del índice de hápax.

Q4, son más amplios comparados con la distancia comprendida entre la mediana y el Q1, en particular en la caja correspondiente a los escritos de las mujeres.

2.3.1.3 Conclusiones parciales

Si tenemos en cuenta exclusivamente los promedios, parece que el sexo no influye en la determinación de la riqueza léxica en ninguno de los índices estudiados.

Igualmente, si desglosamos los datos de cada texto, se confirma que el sexo no tiene influencia en la competencia de un sujeto. Mujeres y hombres han alcanzado los mismos resultados de diversidad léxica, el 100% de los relatos entregados iguala o sobrepasa el 50% de variación y el 0,50 de TTR. El mismo porcentaje de encuestados (85% de cada grupo) ha logrado un buen índice de hápax. Al contrario, los cómputos de densidad léxica han señalado una alteración de esta tendencia ya que el 100% de los hombres supera el 50%, frente al 89% de las mujeres, lo que supone un ligero influjo de este factor sociolingüístico.

Por último, los descriptores estadísticos subrayan que las informantes femeninas presentan una mayor desviación intragrupal en todos los índices, quizás esto se debe a la presencia masiva de las mujeres frente a los hombres al momento de suministrar las pruebas. Los chicos han ofrecido resultados más compactos con menor dispersión interna, lo cual puede significar que el nivel de competencia léxica está compartido por todos los componentes de esta agrupación.

2.3.2 Variable “nivel de ELE”

La repartición de los informantes aquí presenta dos secciones homogéneas, formadas por 50 sujetos cada una. De una primera ojeada, se nota que la media de vocablos manifiesta una situación similar, casi idéntica.

Informantes	<i>Tokens</i>	<i>Types</i>	Promedio de <i>tokens</i>	Promedio de <i>types</i>
B1 (50)	5.000	722	100	62
B2 (50)	5.000	779	100	63

Tabla 147. Datos generales de RL según la variable “nivel de ELE”.

Las tablas 148 y 149 muestran los datos que nos han permitido cotejar la competencia de la muestra mediante el estudio de la variedad, la densidad léxica y el

grado de desviación inter e intragrupal para averiguar si el bagaje léxico se enriquece según el nivel lingüístico aumenta.

Nivel B1					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,64	64%	1,23	48%	2,08
2	0,53	53%	1,51	55%	1,82
3	0,66	66%	1,43	56%	1,79
4	0,62	62%	1,24	50%	2,00
5	0,61	61%	1,27	59%	1,69
6	0,57	57%	1,50	55%	1,82
7	0,60	60%	1,43	59%	1,69
8	0,60	60%	1,43	55%	1,82
9	0,60	60%	1,43	55%	1,82
10	0,64	64%	1,36	57%	1,75
11	0,62	62%	1,32	59%	1,69
12	0,65	65%	1,48	60%	1,67
13	0,63	63%	1,40	52%	1,92
14	0,59	59%	1,51	55%	1,82
15	0,68	68%	1,39	63%	1,59
16	0,53	53%	1,77	52%	1,92
17	0,65	65%	1,35	53%	1,89
18	0,61	61%	1,45	43%	2,33
19	0,53	53%	1,47	51%	1,96
20	0,66	66%	1,14	50%	2,00
21	0,59	59%	1,51	53%	1,89
22	0,65	65%	1,27	52%	1,92
23	0,59	59%	1,51	52%	1,92
24	0,66	66%	1,27	55%	1,82
25	0,71	71%	1,18	56%	1,79
26	0,67	67%	1,31	55%	1,82
27	0,59	59%	1,40	55%	1,82
28	0,56	56%	1,65	60%	1,67
29	0,58	58%	1,41	55%	1,82
30	0,64	64%	1,56	60%	1,67
31	0,62	62%	1,07	53%	1,89
32	0,75	75%	1,29	64%	1,56
33	0,64	64%	1,25	53%	1,89
34	0,61	61%	1,39	48%	2,08

35	0,64	64%	1,25	56%	1,79
36	0,69	69%	1,41	66%	1,52
37	0,59	59%	1,55	51%	1,96
38	0,69	69%	1,33	54%	1,85
39	0,63	63%	1,34	50%	2,00
40	0,54	54%	1,69	49%	2,04
41	0,63	63%	1,54	61%	1,64
42	0,70	70%	1,27	58%	1,72
43	0,61	61%	1,42	60%	1,67
44	0,62	62%	1,38	59%	1,69
45	0,57	57%	1,43	54%	1,85
46	0,66	66%	1,35	60%	1,67
47	0,66	66%	1,29	60%	1,67
48	0,64	64%	1,36	57%	1,75
49	0,65	65%	1,25	55%	1,82
50	0,61	61%	1,45	54%	1,85
PROMEDIO	0,62	62%	1,39	55%	1,82

Tabla 148. Índices de RL según la variante “nivel B1”.

Nivel B2					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,67	67%	1,37	61%	1,64
2	0,54	54%	1,64	50%	2,00
3	0,62	62%	1,38	50%	2,00
4	0,60	60%	1,50	54%	1,85
5	0,59	59%	1,31	53%	1,89
6	0,62	62%	1,41	53%	1,89
7	0,55	55%	1,45	48%	2,08
8	0,65	65%	1,38	63%	1,59
9	0,71	71%	1,25	57%	1,75
10	0,66	66%	1,38	56%	1,79
11	0,61	61%	1,39	48%	2,08
12	0,67	67%	1,31	60%	1,67
13	0,62	62%	1,32	53%	1,89
14	0,64	64%	1,33	55%	1,82
15	0,61	61%	1,45	58%	1,72
16	0,63	63%	1,50	55%	1,82
17	0,54	54%	1,59	50%	2,00
18	0,66	66%	1,29	57%	1,75

19	0,68	68%	1,28	59%	1,69
20	0,53	53%	1,56	49%	2,04
21	0,65	65%	1,27	52%	1,92
22	0,64	64%	1,28	52%	1,92
23	0,67	67%	1,34	52%	1,92
24	0,57	57%	1,46	50%	2,00
25	0,73	73%	1,30	56%	1,79
26	0,66	66%	1,47	55%	1,82
27	0,70	70%	1,37	61%	1,64
28	0,58	58%	1,41	46%	2,17
29	0,56	56%	1,65	55%	1,82
30	0,59	59%	1,40	58%	1,72
31	0,72	72%	1,26	59%	1,69
32	0,67	67%	1,34	59%	1,69
33	0,64	64%	1,33	58%	1,72
34	0,67	67%	1,29	52%	1,92
35	0,61	61%	1,45	54%	1,85
36	0,65	65%	1,41	57%	1,75
37	0,63	63%	1,31	49%	2,04
38	0,66	66%	1,27	53%	1,89
39	0,65	65%	1,30	48%	2,08
40	0,61	61%	1,42	53%	1,89
41	0,63	63%	1,29	52%	1,92
42	0,58	58%	1,53	54%	1,85
43	0,60	60%	1,50	57%	1,75
44	0,69	69%	1,35	57%	1,75
45	0,67	67%	1,34	58%	1,72
46	0,66	66%	1,29	55%	1,82
47	0,67	67%	1,29	54%	1,85
48	0,59	59%	1,48	56%	1,79
49	0,67	67%	1,37	57%	1,75
50	0,61	61%	1,39	50%	2,00
PROMEDIO	0,63	63%	1,39	54%	1,85

Tabla 149. Índices de RL según la variante “nivel B2”.

2.3.2.1 Diversidad léxica

(i) TTR

La relación entre palabras repetidas y no repetidas dentro del corpus dividido en B1 y B2 destaca una distribución equivalente de los datos.

Descriptivos	Variante	
	B1	B2
Media	0,62	0,63
Mediana	0,63	0,64
Mínimo	0,53	0,53
Máximo	0,75	0,73

Tabla 150. Estadísticos descriptivos del TTR para la variable “nivel de ELE”.

Los estudiantes más avanzados sobrepasan, de poco, a los compañeros en los valores estadísticos de medio, mientras que el grupo menos avanzado los superan en el extremo máximo, donde alcanzan un 0,75 (contra el 0,73 de los otros). El valor mínimo, por el contrario, es el mismo para los dos niveles.

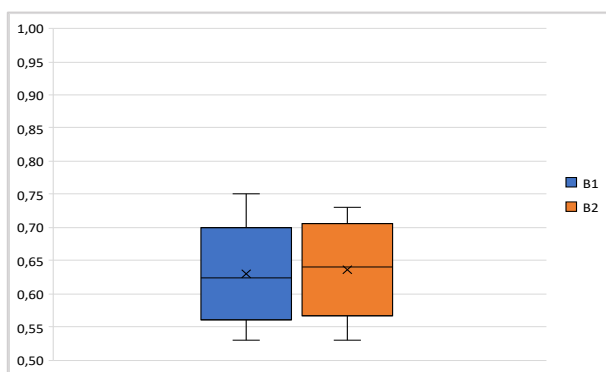


Gráfico 48. Diagrama de cajas del TTR para la variable “nivel de ELE”.

Ambos muestreos presentan una oscilación intragrupal idéntica, parece que a este punto de la ASL el grado de lengua adquirido no cambia la capacidad productiva de los estudiantes en términos de variedad. En este sentido, las cajas presentan los mismos intervalos entre cuartiles.

(ii) Variación léxica

Una vez que se traducen los resultados del TTR en porcentaje obtenemos los datos de la variación léxica.

Descriptivos	Variante	
	B1	B2
Media	62%	63%
Mediana	63%	64%
Mínimo	53%	53%

Máximo	75%	73%
--------	-----	-----

Tabla 151. Estadísticos descriptivos de la variación léxica para la variable “nivel de ELE”.

La tabla refleja la misma realidad que corrobora la ausencia de un efecto de este factor sociolingüístico en la variación del vocabulario conocido. El resultado más bajo (53%) está compartido por las dos agrupaciones, los encuestados de nivel menor tocan el extremo superior (75%) y los otros alcanzan una media (63%) y una mediana (64%) mayores, pero, al fin y al cabo, la variación es mínima.

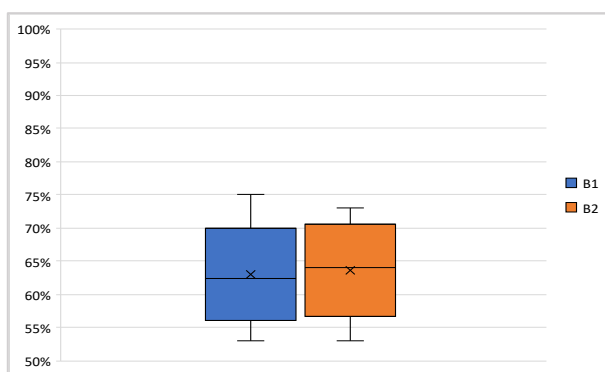


Gráfico 49. Diagrama de cajas de la variación léxica para la variable “nivel de ELE”.

Ya hemos comentado que los datos son satisfactorios en general porque al analizar todos los resultados de TTR y de variación, el entero conjunto textual supera el 50% de diversidad léxica sin distinción de nivel.

Si embargo, el grado de competencia lingüística no parece ejercer un influjo tan fuerte como para alterar los resultados.

(iii) Índice de hápax

Los cálculos del índice de hápax ponen de relieve una gran oscilación interior de los resultados, si excluimos la media.

Descriptivos	Variante	
	B1	B2
Media	1,39	1,39
Mediana	1,39	1,37
Mínimo	1,07	1,25
Máximo	1,77	1,65

Tabla 152. Estadísticos descriptivos del índice de hápax para la variable “nivel de ELE”.

Los alumnos de nivel B1 aportan, a la vez, el mejor valor (1,07) y el peor (1,77). Obtienen el índice más pequeño en el texto 131 que, por supuesto, contiene la mayor cantidad de hápax y el valor más grande en el 116 que denota la escasa presencia de palabras “originales”. Los B2 arrojan datos más cerrados que van de 1,25 a 1,65. La mediana, por su parte, está a favor de estos, aunque de poco.

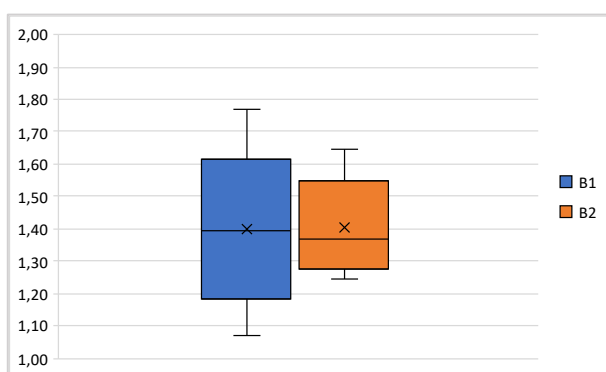


Gráfico 50. Diagrama de cajas del índice de hápax para la variable “nivel de ELE”.

Los resultados de los informantes de nivel B1 se alteran más que los de nivel B2, donde la variación intragrupal es menor, por eso la caja azul es mucho más alargada. La caja naranja, de hecho, es más pequeña como la extensión de los bigotes que apenas superan el Q3 de la azul y no logran el límite del Q1. De todas formas, el 80% de los relatos de los B1 y el 90% de los B2 tienen un índice de hápax de hasta 1,50.

2.3.2.2 Densidad léxica

(i) Densidad léxica

La densidad revela que los resultados de los encuestados de nivel B1, otra vez, varían más en el uso de las palabras semánticas comparado con el empleo que hacen los B2, que tienen una menor desviación.

Descriptivos	Variante	
	B1	B2
Media	55%	54%
Mediana	55%	55%
Mínimo	43%	46%
Máximo	66%	63%

Tabla 153. Estadísticos descriptivos de la densidad léxica para la variable “nivel de ELE”.

Los alumnos intermedios arrojan el valor más elevado (66%), el mínimo se ha detectado en el grupo avanzado (46%). De acuerdo con estos datos, es posible que la variable no influya en la producción de palabras nocionales.

La variación intergrupal mayor se ha detectado en los B1 que llegan a los extremos superior e inferior del corpus, con una oscilación de 23 puntos frente a los 17 de los B2.

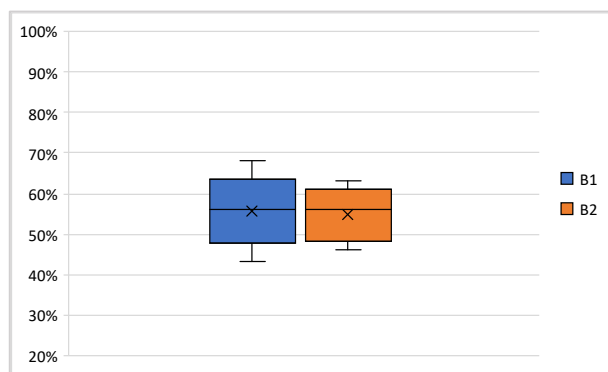


Gráfico 51. Diagrama de cajas de la densidad léxica para la variable “nivel de ELE”.

Sin embargo, la variación intergrupal no es tan significativa para que el nivel de español afecte este índice, como se desprende de las proporciones de los intervalos entre cuartiles de las cajas. Lo que llama la atención es la longitud diferente de los bigotes que marcan los extremos de cada agrupación. Examinando el índice de cada texto, se nota que el 92% de los B1 igualan o sobrepasan el 50% de densidad léxica frente al 88% de los B2. Pues, la mayoría de los alumnos de nivel inferior ha escrito una cantidad mayor de palabras temáticas, aunque sea por poco (4 puntos). El diagrama destaca esta diferencia: la caja azul sobrepasa la naranja alargándose hasta el extremo superior del bigote, donde se colocan los resultados mejores.

(ii) IAT

La tabla 154 muestra los resultados procedentes de la medición del IAT que revela una situación parecida a lo visto hasta ahora.

Descriptivos	B1	B2
Media	1,82	1,85
Mediana	1,82	1,84
Mínimo	1,52	1,59
Máximo	2,33	2,17

Tabla 154. Estadísticos descriptivos del IAT para la variable “nivel de ELE”.

Los datos señalan valores altos de IAT ya que se encuentra una palabra nocional casi cada dos palabras funcionales: en este caso, el grado de riqueza no es alto.

La media matemática y la mediana favorecen, ligeramente, el nivel B1 que alcanza una cifra (1,82) un poco menor del B2 (1,85 y 1,84).

Asimismo, los textos de los aprendices de nivel B1 tocan los polos opuestos (mínimo y máximo), manifestando una mayor desviación intragrupal, desde 1,52 hasta 2,33 contra los estudiantes B2 que va de 1,59 a 2,17.

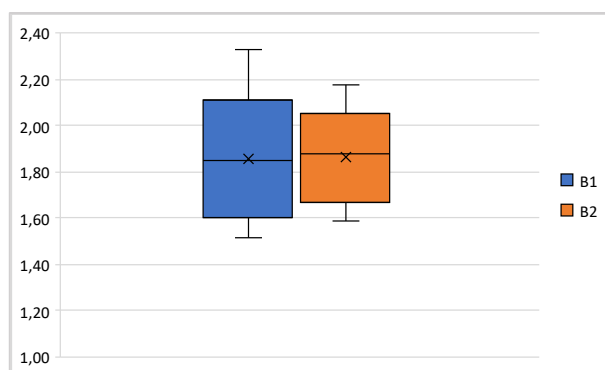


Gráfico 52. Diagrama de cajas del IAT para la variable “nivel de ELE”.

La disparidad de la desviación intragrupal es evidente y radica en la configuración de los diagramas: el tamaño de las cajas, la amplitud intercuartil y la extensión de los bigotes. En los relatos de los alumnos de nivel B1 sigue vigente una oscilación mayor mientras que los participantes del B2 confirman la tendencia surgida en los índices previos arrojando datos más compactos.

2.3.2.3 Conclusiones parciales

La riqueza léxica contabilizada en los textos divididos en función del nivel de español se asemeja a lo largo del corpus, las diferencias entre secciones no son tan significativas: ninguno sobresale notablemente por encima del otro ya que se han computado valores similares.

Parece que, en lo referente a la diversidad de vocabulario, el dominio lingüístico no influye en la riqueza: los textos de ambas muestras igualan o sobrepasan el 50% de variación y el 0,50 de TTR.

Al contrario, el índice de hápax ha destacado un desnivel entre los grupos: el 80% de los B1 y el 90% de los B2 proporcionan un índice hasta 1,50 que marca una leve

superioridad de los informantes más avanzados.

Lo mismo revelan los cómputos de densidad léxica, que señalan un ligero influjo del condicionante puesto que el 92% de los B1 tienen un índice equivalente o superior a 50% frente al 88% de los B2.

El IAT, por su parte, presenta un rendimiento bastante bajo en sendos muestreos e implica un grado carente de riqueza léxica, los resultados superan el 1,50 acercándose más a 2 que a 1.

En resumidas cuentas, los escritos del grupo B1 son más densos de palabras nocionales mientras que los del grupo B2 alcanzan una diversidad léxica superior.

Parece que la mayor influencia de este factor se ubica en el grado de desviación intragrupal de cada variante: el análisis de los textos del conjunto B1 pone de manifiesto una gama más amplia de valores que se traduce en una competencia no bien perfilada como la del grupo opuesto, cuya riqueza léxica es más circunscrita.

2.3.3 Variable “conocimiento de otras LE”

El último apartado dedicado a los resultados por variable está finalizado al análisis de la influencia del conocimiento de otras LE. Se trata de un estudio transversal pionero ya que este factor nunca ha sido objeto de comparación en el ámbito de la riqueza léxica.

Antes del análisis, recordamos que las dos secciones no están equilibradas: contamos con 19 informantes que conocen dos LE y 81 que saben más.

Informantes	<i>Tokens</i>	<i>Types</i>	Promedio de <i>tokens</i>	Promedio de <i>types</i>
Conoce 2 LE (19)	1.900	441	100	61
Conoce + 2 LE (81)	8.100	963	100	63

Tabla 155. Datos generales de RL según la variable “conocimiento de otras LE”.

Las tablas 156 y 157 muestran los resultados peculiares de los textos.

Conoce 2 LE					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,54	54%	1,24	50%	2,00
2	0,62	62%	1,36	50%	2,00
3	0,60	60%	1,40	54%	1,85
4	0,55	55%	1,65	48%	2,08

5	0,56	56%	1,41	55%	1,82
6	0,67	67%	1,39	59%	1,69
7	0,67	67%	1,41	58%	1,72
8	0,66	66%	1,69	55%	1,82
9	0,61	61%	1,54	50%	2,00
10	0,62	62%	1,29	50%	2,00
11	0,64	64%	1,64	57%	1,75
12	0,59	59%	1,38	55%	1,82
13	0,56	56%	1,50	60%	1,67
14	0,58	58%	1,45	55%	1,82
15	0,61	61%	1,65	48%	2,08
16	0,69	69%	1,34	66%	1,52
17	0,54	54%	1,34	49%	2,04
18	0,63	63%	1,29	61%	1,64
19	0,66	66%	1,39	60%	1,67
PROMEDIO	0,61	61%	1,44	55%	1,84

Tabla 156. Índices de RL según la variante “LE =2”.

Conoce más de 2 LE					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,67	67%	1,23	61%	1,64
2	0,59	59%	1,51	53%	1,89
3	0,62	62%	1,43	53%	1,89
4	0,65	65%	1,27	63%	1,59
5	0,71	71%	1,50	57%	1,75
6	0,66	66%	1,43	56%	1,79
7	0,61	61%	1,43	48%	2,08
8	0,67	67%	1,43	60%	1,67
9	0,62	62%	1,32	53%	1,89
10	0,64	64%	1,48	55%	1,82
11	0,61	61%	1,40	58%	1,72
12	0,63	63%	1,51	55%	1,82
13	0,54	54%	1,39	50%	2,00
14	0,66	66%	1,77	57%	1,75
15	0,68	68%	1,35	59%	1,69
16	0,53	53%	1,45	49%	2,04
17	0,65	65%	1,47	52%	1,92
18	0,64	64%	1,14	52%	1,92
19	0,67	67%	1,51	52%	1,92

20	0,57	57%	1,27	50%	2,00
21	0,73	73%	1,51	56%	1,79
22	0,66	66%	1,27	55%	1,82
23	0,70	70%	1,18	61%	1,64
24	0,58	58%	1,31	46%	2,17
25	0,59	59%	1,56	58%	1,72
26	0,72	72%	1,07	59%	1,69
27	0,64	64%	1,29	58%	1,72
28	0,67	67%	1,25	52%	1,92
29	0,61	61%	1,25	54%	1,85
30	0,65	65%	1,55	57%	1,75
31	0,63	63%	1,33	49%	2,04
32	0,66	66%	1,34	53%	1,89
33	0,65	65%	1,27	48%	2,08
34	0,61	61%	1,42	53%	1,89
35	0,63	63%	1,38	52%	1,92
36	0,58	58%	1,43	54%	1,85
37	0,60	60%	1,35	57%	1,75
38	0,69	69%	1,36	57%	1,75
39	0,67	67%	1,25	54%	1,85
40	0,59	59%	1,45	56%	1,79
41	0,67	67%	1,37	57%	1,75
42	0,64	64%	1,31	48%	2,08
43	0,53	53%	1,41	55%	1,82
44	0,66	66%	1,38	56%	1,79
45	0,61	61%	1,25	59%	1,69
46	0,57	57%	1,38	55%	1,82
47	0,60	60%	1,39	59%	1,69
48	0,60	60%	1,31	55%	1,82
49	0,60	60%	1,32	55%	1,82
50	0,62	62%	1,33	59%	1,69
51	0,65	65%	1,45	60%	1,67
52	0,63	63%	1,50	52%	1,92
53	0,59	59%	1,59	55%	1,82
54	0,68	68%	1,29	63%	1,59
55	0,53	53%	1,28	52%	1,92
56	0,65	65%	1,56	53%	1,89
57	0,61	61%	1,27	43%	2,33
58	0,53	53%	1,28	51%	1,96
59	0,66	66%	1,34	50%	2,00

60	0,59	59%	1,46	53%	1,89
61	0,65	65%	1,30	52%	1,92
62	0,59	59%	1,47	52%	1,92
63	0,66	66%	1,37	55%	1,82
64	0,71	71%	1,41	56%	1,79
65	0,67	67%	1,40	55%	1,82
66	0,64	64%	1,26	60%	1,67
67	0,62	62%	1,33	53%	1,89
68	0,75	75%	1,29	64%	1,56
69	0,64	64%	1,45	53%	1,89
70	0,64	64%	1,41	56%	1,79
71	0,59	59%	1,31	51%	1,96
72	0,69	69%	1,27	54%	1,85
73	0,63	63%	1,30	50%	2,00
74	0,70	70%	1,42	58%	1,72
75	0,61	61%	1,29	60%	1,67
76	0,62	62%	1,53	59%	1,69
77	0,57	57%	1,50	54%	1,85
78	0,66	66%	1,35	60%	1,67
79	0,64	64%	1,29	57%	1,75
80	0,65	65%	1,48	55%	1,82
81	0,61	61%	1,37	54%	1,85
PROMEDIO	0,63	63%	1,38	55%	1,83

Tabla 157. Índices de RL según la variante “LE >2”.

2.3.3.1 Diversidad léxica

(i) TTR

La relación entre *tokens* y *types* se desglosa en la siguiente tabla.

Descriptivos	LE =2	LE >2
Media	0,61	0,63
Mediana	0,61	0,64
Mínimo	0,54	0,53
Máximo	0,69	0,75

Tabla 158. Estadísticos descriptivos del TTR para la variable “conocimiento de otras LE”.

En principio, los datos se inclinan a favor de los estudiantes que conocen más de dos idiomas ya que media (0,63) y mediana (0,64) mejores se han localizado en sus

producciones, así como el valor máximo (0,75), pero se halla en este conjunto también el mínimo (0,53). Sin embargo, esto no pesa demasiado en los resultados de este grupo que alcanza un nivel mayor de riqueza comparado con el otro.

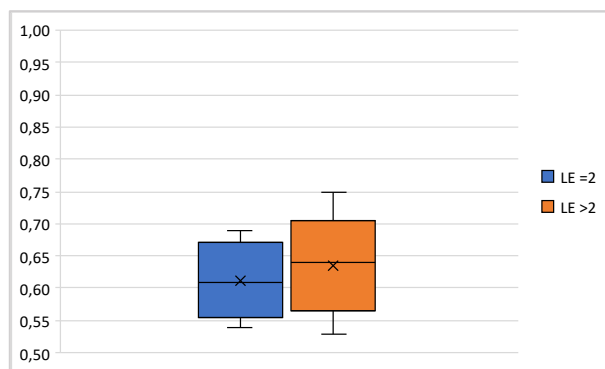


Gráfico 53. Diagrama de cajas del TTR para la variable “conocimiento de otras LE”.

En el gráfico destacan los bigotes del diagrama naranja que se extienden hasta los extremos y, en particular, la cantidad de datos que sobrepasan los resultados obtenidos por los compañeros, quienes no superan el 0,69 de TTR.

De acuerdo con esta distribución, observamos una disimetría en el comportamiento de las dos variantes que marca una superioridad del 9% de los informantes del grupo LE >2.

(ii) Variación léxica

El grado de diversidad del vocabulario utilizado se expresa igualmente mediante la variación léxica resumida en la tabla 159.

Descriptivos	LE =2	LE >2
Media	61%	63%
Mediana	61%	64%
Mínimo	54%	53%
Máximo	69%	75%

Tabla 159. Estadísticos descriptivos de la variación léxica para la variable “conocimiento de otras LE”.

El porcentaje de *types* sobre el total de *tokens* revela un aumento de la riqueza en los textos arrojados por los informantes que saben más de dos LE, que corrobora, otra vez, la hipótesis de Cummins (1979). Los que solo conocen dos idiomas tienen un índice de variación más reducido, si bien adecuado al nivel de ELE, como enseña el gráfico.

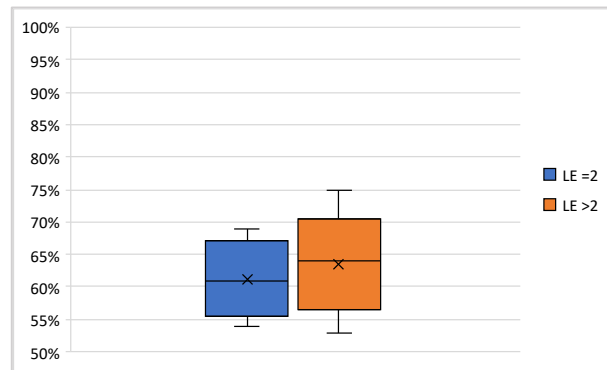


Gráfico 54. Diagrama de cajas de la variación léxica para la variable “conocimiento de otras LE”.

Constatamos que todos han logrado más del 50% de variación, hecho que, sin duda, permite afirmar que el nivel de variedad léxica satisface las expectativas.

Pero, al mismo tiempo, no sorprende el desnivel evidente entre las dos muestras debido a la asimetría entre la posición de las medias y las medianas como entre la distinta prolongación de los bigotes de los dos diagramas: se patentiza el rendimiento mejor de un subconjunto frente al otro (+9%).

(iii) Índice de hápax

La misma tendencia se manifiesta en el análisis del índice de hápax, como se desprende de la tabla.

Descriptivos	LE =2	LE >2
Media	1,44	1,38
Mediana	1,40	1,37
Mínimo	1,24	1,07
Máximo	1,69	1,77

Tabla 160. Estadísticos descriptivos del índice de hápax para la variable “conocimiento de otras LE”.

Los valores oscilan del 65% y ambos extremos se posicionan dentro de los resultados conseguidos por el subconjunto LE >2, tal cual muestra el diagrama naranja. A pesar de que estos estudiantes toquen el 1,77 (+5%), a la vez, alcanzan el mejor índice (1,07) e incrementan su rendimiento del 16% con respecto a sus compañeros (1,24).

Media y medianas de los encuestados que conocen más LE se revelan mejores (-4% y -2% del grupo LE =2): el índice es más pequeño y comporta la mayor presencia de hápax en la producción textual. Efectivamente, los otros informantes proporcionan un índice más elevado que conlleva la menor aparición de palabras “originales”.

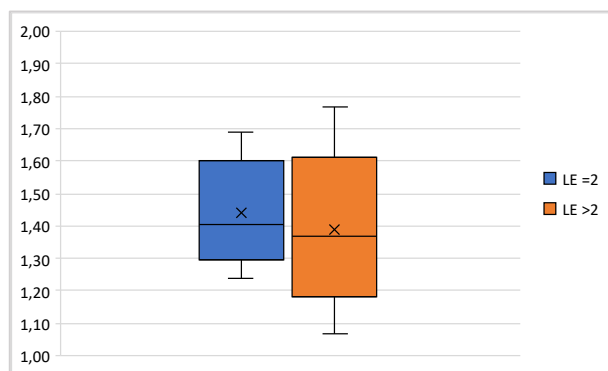


Gráfico 55. Diagrama de cajas del índice de hápax para la variable “conocimiento de otras LE”.

La mayoría de los textos que están posicionados por debajo del Q2, que contiene los datos mejores, son los escritos de los alumnos del grupo LE >2. Al verificar cuáles y cuántos textos por cada agrupación tienen un índice de hápax igual o menor de 1,50 hemos comprobado que se trata del 74% de la variante LE =2 y del 88% de la LE >2, lo que patentiza una producción más grande de hápax en esta sección: comparada con la naranja, la amplitud intercuartil azul Q2-Q3 es reducida.

Se revela una cierta influencia del factor en la productividad de los participantes.

2.3.3.2 Densidad léxica

(i) Densidad léxica

El porcentaje de uso de las palabras léxicas frente a las gramaticales trae a colación un cambio de tendencia en lo que atañe a la influencia de la variable: se manifiesta que los encuestados que conocen dos LE han arrojados resultados más altos.

Descriptivos	LE =2	LE >2
Media	55%	55%
Mediana	55%	55%
Mínimo	48%	43%
Máximo	66%	64%

Tabla 161. Estadísticos descriptivos de la densidad léxica para la variable “conocimiento de otras LE”.

Los datos estadísticos son mayores, pese a que la desviación propia del otro muestreo sea superior (de 43% a 64% *versus* de 48% a 66%).

Por su parte, media y mediana se corresponden en los dos grupos llegando a un porcentaje apreciable del 55%, como ya hemos detectado en el análisis general.

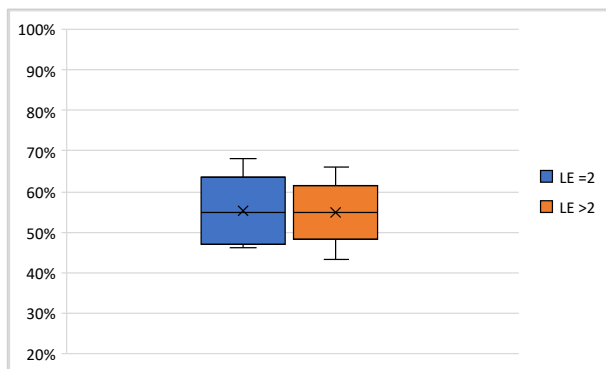


Gráfico 56. Diagrama de cajas de la densidad léxica para la variable “conocimiento de otras LE”.

Se observa que la variante LE =2 presenta un rendimiento mejor, la amplitud intercuartil Q2-Q4 de la caja azul se alarga más, el bigote superior sube hasta el punto máximo (66%) y el inferior no sobrepasa el 48%. Al contrario, el bigote inferior de la variante LE >2 se extiende hasta el 43%, obteniendo el peor índice en absoluto y el superior no logra el mismo nivel (-2%).

Por último, examinando el índice obtenido por cada relato, el 84% de los escritos del primer muestreo igualan o sobrepasan el 50% frente al 91% de las producciones del segundo, los cuales han aportado una más elevada cantidad de palabras de contenido nocional.

(ii) IAT

A partir de los siguientes datos analizamos si el IAT confirma la situación destacada en el apartado previo.

Descriptivos	LE =2	LE >2
Media	1,84	1,83
Mediana	1,82	1,82
Mínimo	1,52	1,56
Máximo	2,08	2,33

Tabla 162. Estadísticos descriptivos del IAT para la variable “conocimiento de otras LE”.

De nuevo, predomina un rendimiento mejor del conjunto textual de los informantes del grupo LE =2 que tienen índices más reducidos. Se observa un menor intervalo, si bien las medianas coinciden entre los dos muestreos (1,82) y las medias cambian solo de 0,01.

En general, ya sabemos que las medidas son altas e implican un escaso grado de

riqueza léxica ya que el entero corpus proporciona un índice de 1,52 como mínimo.

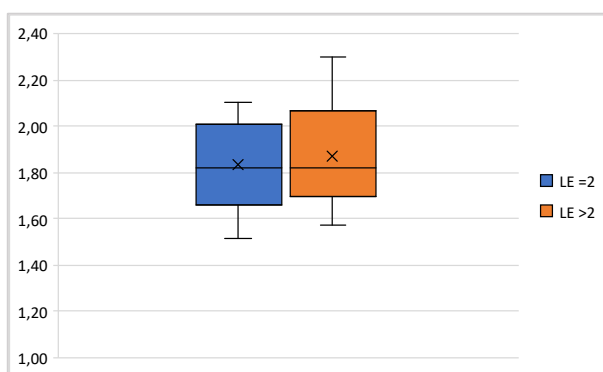


Gráfico 57. Diagrama de cajas del IAT para la variable “conocimiento de otras LE”.

Los resultados del grupo LE >2 destacan, primero, que a partir del Q1 los datos son más elevados y conllevan una menor riqueza léxica. Segundo, el bigote inferior baja a 1,56 pero no logra un índice tan bajo como el obtenido por los compañeros (1,52). Tercero, el intervalo Q3-Q4 sube a 2,33 alcanzado el valor máximo del corpus.

Del otro lado, la amplitud intercuartil y la desviación intragrupal del diagrama azul son menores y oscilan entre valores más pequeños que sugieren una mayor riqueza léxica del subconjunto LE =2.

2.3.3.3 Conclusiones parciales

Hemos demostrado que la variable en cuestión tiene más influencia en la riqueza léxica de los encuestados con respecto a los otros factores contemplados.

Los índices de diversidad han revelado que los informantes que conocen más de dos LE arrojan resultados mejores en los cálculos de TTR y variación del 9%, pese a que hemos detectado que el 100% de los textos ha logrado más del 50% de variación del vocabulario, dato satisfactorio.

Igualmente, estos alumnos alcanzan el mejor índice de hápax en absoluto (1,07) incrementando el rendimiento del 16%. Del mismo modo, se han contabilizado mejores valores de media y mediana. Por último, el 88% tiene un índice de hápax igual o menor de 1,50 frente al 74% de la variante LE =2.

Parece que hay una cierta influencia del factor sociolingüístico que corrobora la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979), cuyo planteamiento asevera que un individuo que ya conoce una lengua extranjera emplea sus competencias como elemento facilitador durante el aprendizaje de otro idioma.

Sorprende un cambio de tendencia al examinar el uso de palabras semánticas: los índices de densidad léxica ponen de manifiesto un rendimiento mejor de los encuestados que saben dos LE. El porcentaje de *tokens* léxicos es mayor en su producción, no obstante los promedios sean idénticos (55%). De igual manera, el IAT destaca la supremacía de este grupo.

En conclusión, cabe señalar la utilidad de incluir el análisis de este factor sociolingüístico en los trabajos sobre riqueza léxica, ya que es en el que más diferencias se observan en la competencia de nuestros informantes, lo cual alienta a seguir investigando sobre los conocimientos lingüísticos de los estudiantes.

2.4 Conclusiones del análisis transversal

Por último, las tablas recapitulativas 163 y 164 contienen los mejores resultados y los promedios obtenidos en cada sección analizada.

Corpus	Variante	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
General		0,75	75%	1,07	66%	1,52
Sexo	M	0,75	75%	1,07	66%	1,52
	H	0,71	71%	1,24	59%	1,69
Nivel de ELE	B1	0,75	75%	1,07	66%	1,52
	B2	0,73	73%	1,25	63%	1,59
Conocimiento de otras LE	LE =2	0,69	69%	1,24	66%	1,52
	LE >2	0,75	75%	1,07	64%	1,56

Tabla 163. Mejores resultados de RL del análisis transversal.

Corpus	Variante	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
General		0,63	63%	1,39	55%	1,84
Sexo	M	0,63	63%	1,39	55%	1,83
	H	0,62	62%	1,39	54%	1,86
Nivel de ELE	B1	0,62	62%	1,39	55%	1,82
	B2	0,63	63%	1,39	54%	1,85
Conocimiento de otras LE	LE =2	0,61	61%	1,44	55%	1,82
	LE >2	0,63	63%	1,38	55%	1,82

Tabla 164. Promedios de RL del análisis transversal.

En primera instancia, hemos analizado la diversidad léxica de nuestro corpus

general, cuyos resultados han manifestado un rendimiento positivo de los participantes. Los textos presentan un buen nivel de *Type/Token Ratio* e igualmente de variación ya que en cada uno se ha detectado más del 50% de vocablos. El índice de hápax también ha mostrado un buen grado de riqueza léxica, la aplicación de la fórmula ha llevado a valores medio-bajos que se aproximan más a 1 que a 2, lo cuales implican un resultado satisfactorio en función del nivel de ELE de nuestros informantes.

En segunda instancia, el análisis de la densidad léxica, es decir del uso de los *tokens* léxicos frente a la totalidad de los *tokens* aportados, ha arrojado resultados divergentes, pese a que los dos índices examinados estén relacionados de manera inversamente proporcional: al aumentar la densidad léxica debería corresponder un IAT reducido. Se ha observado que el 90% de los relatos está formado por más del 50% de palabras temáticas (el promedio es 55%). Sin embargo, el IAT no se ha revelado tan bueno ya que las cifras contabilizadas señalan la aparición de una unidad de contenido semántico cada dos gramaticales (el promedio es 1,84).

Extraña que no obstante el grado de riqueza que se ha manifestado a partir de la diversidad léxica, la densidad no haya aportado el mismo rendimiento.

En la siguiente etapa el análisis de los resultados por variable, realizado conforme a los tres factores sociolingüísticos contemplados (sexo, nivel de ELE, conocimiento de otras LE), permite averiguar si estos surten algún efecto en la capacidad de los encuestados.

El sexo parece no influir en la competencia de los informantes. Mujeres y hombres presentan el mismo nivel de diversidad léxica. Todos los relatos igualan o sobrepasan el 50% de variación y el 85% de cada variante tiene un buen índice de hápax. Pero, los resultados de la densidad léxica modifican esta tendencia, subrayando un ligero influjo del sexo ya que el 100% de los varones supera el 50% del índice frente al 89% de las mujeres.

El nivel de ELE pone de manifiesto que el dominio lingüístico tampoco parece influir en la riqueza léxica. La variedad de vocabulario iguala o sobrepasa el 50% de todos los textos, aunque la medición del índice de hápax destaca una leve superioridad de los informantes avanzados (el 90% de los B2 proporcionan un índice hasta 1,50 frente al 80% de los B1). Del mismo modo, la densidad demuestra un leve influjo del nivel puesto que el 92% de los B1 tienen un índice equivalente o superior a 50% respecto al 88% de los

compañeros. Los textos de la sección B1 son más densos, pero las diferencias no son tan significativas para que este grupo sobresalga por encima del otro.

Asimismo, en este apartado, la desviación de cada muestra patentiza que el subconjunto B1 tiene una escala más amplia de datos que parece conllevar una competencia no bien definida como la del B2, que queda circunscrita en una gama más restringida de valores.

La última variable, conocimiento de otras LE, confirma los planteamientos de Cummins (1979). Los encuestados que conocen más de dos LE aportan resultados mejores de TTR y de variación léxica del 9%. Igualmente, alcanzan el mejor índice de hápax incrementando el rendimiento del 16%. Sorprende un cambio al analizar el uso de palabras nocionales: la densidad léxica y el IAT demuestran un rendimiento mejor de los informantes que saben solo dos LE. Cabe subrayar que este primer análisis de la variable se ha revelado interesante ya que ha manifestado un desnivel entre los sujetos que componen los dos grupos: parece que el conocimiento de otras LE es el factor sociolingüístico que más influye en la competencia de los informantes.

No mencionamos en ningún caso del IAT porque ya hemos comentado que el rendimiento es bajo: en general se supera el 1,50 sin distinción de variante o variable.

3. ANÁLISIS LONGITUDINAL

Realizamos el análisis longitudinal de los relatos comparando los recogidos durante la primera administración de la prueba con los de la segunda, ambas realizadas por los mismos estudiantes de nivel B2 como partimos de la hipótesis de que existe una relación asociativa entre el nivel de ELE y la cantidad de lexías conocidas y activables. De nuevo, aplicamos el mismo protocolo empírico: medimos los índices de TTR, variación, hápax, densidad, IAT y estudiamos los descriptores estadísticos.

Corpus	Textos	<i>Tokens</i>	<i>Types</i>	Promedio de <i>tokens</i>	Promedio de <i>types</i>
B2_a	50	5.000	743	10	61
B2_b	50	5.000	779	100	63

Tabla 165. Datos generales de RL según la fecha de la prueba.

Por su parte, las tablas 166 y 167 contienen los resultados repartidos según la fecha de realización del test, que utilizamos para averiguar si la competencia de los encuestados

se enriquece según se avance en el aprendizaje de la lengua.

I suministración					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,67	67%	1,37	61%	1,64
2	0,54	54%	1,64	50%	2,00
3	0,62	62%	1,38	50%	2,00
4	0,60	60%	1,50	54%	1,85
5	0,59	59%	1,31	53%	1,89
6	0,62	62%	1,41	53%	1,89
7	0,55	55%	1,45	48%	2,08
8	0,65	65%	1,38	63%	1,59
9	0,71	71%	1,25	57%	1,75
10	0,66	66%	1,38	56%	1,79
11	0,61	61%	1,39	48%	2,08
12	0,67	67%	1,31	60%	1,67
13	0,62	62%	1,32	53%	1,89
14	0,64	64%	1,33	55%	1,82
15	0,61	61%	1,45	58%	1,72
16	0,63	63%	1,50	55%	1,82
17	0,54	54%	1,59	50%	2,00
18	0,66	66%	1,29	57%	1,75
19	0,68	68%	1,28	59%	1,69
20	0,53	53%	1,56	49%	2,04
21	0,65	65%	1,27	52%	1,92
22	0,64	64%	1,28	52%	1,92
23	0,67	67%	1,34	52%	1,92
24	0,57	57%	1,46	50%	2,00
25	0,73	73%	1,30	56%	1,79
26	0,66	66%	1,47	55%	1,82
27	0,70	70%	1,37	61%	1,64
28	0,58	58%	1,41	46%	2,17
29	0,56	56%	1,65	55%	1,82
30	0,59	59%	1,40	58%	1,72
31	0,72	72%	1,26	59%	1,69
32	0,67	67%	1,34	59%	1,69
33	0,64	64%	1,33	58%	1,72
34	0,67	67%	1,29	52%	1,92
35	0,61	61%	1,45	54%	1,85

36	0,65	65%	1,41	57%	1,75
37	0,63	63%	1,31	49%	2,04
38	0,66	66%	1,27	53%	1,89
39	0,65	65%	1,30	48%	2,08
40	0,61	61%	1,42	53%	1,89
41	0,63	63%	1,29	52%	1,92
42	0,58	58%	1,53	54%	1,85
43	0,60	60%	1,50	57%	1,75
44	0,69	69%	1,35	57%	1,75
45	0,67	67%	1,34	58%	1,72
46	0,66	66%	1,29	55%	1,82
47	0,67	67%	1,29	54%	1,85
48	0,59	59%	1,48	56%	1,79
49	0,67	67%	1,37	57%	1,75
50	0,61	61%	1,39	50%	2,00
PROMEDIO	0,63	63%	1,39	54%	1,85

Tabla 166. Índices de RL en la I suministración de la prueba.

II suministración					
Texto	TTR	Variación léxica	Índice de hápax	Densidad léxica	IAT
1	0,57	57%	1,63	58%	1,72
2	0,46	46%	1,59	42%	2,38
3	0,67	67%	1,22	52%	1,92
4	0,69	69%	1,25	56%	1,79
5	0,59	59%	1,37	54%	1,85
6	0,60	60%	1,67	60%	1,67
7	0,59	59%	1,48	55%	1,82
8	0,62	62%	1,35	51%	1,96
9	0,55	55%	1,57	54%	1,85
10	0,63	63%	1,29	53%	1,89
11	0,62	62%	1,44	51%	1,96
12	0,62	62%	1,35	52%	1,92
13	0,61	61%	1,39	54%	1,85
14	0,66	66%	1,40	58%	1,72
15	0,58	58%	1,41	57%	1,75
16	0,61	61%	1,36	62%	1,61
17	0,70	70%	1,27	58%	1,72
18	0,51	51%	1,16	53%	1,89
19	0,62	62%	1,35	53%	1,89

20	0,60	60%	1,36	56%	1,79
21	0,60	60%	1,36	51%	1,96
22	0,67	67%	1,20	56%	1,79
23	0,62	62%	1,35	55%	1,82
24	0,60	60%	1,54	53%	1,89
25	0,60	60%	1,50	55%	1,82
26	0,58	58%	1,45	55%	1,82
27	0,68	68%	1,21	55%	1,82
28	0,62	62%	1,51	55%	1,82
29	0,67	67%	1,26	55%	1,82
30	0,66	66%	1,22	56%	1,79
31	0,62	62%	1,41	50%	2,00
32	0,57	57%	1,46	47%	2,13
33	0,60	60%	1,40	56%	1,79
34	0,60	60%	1,43	58%	1,72
35	0,62	62%	1,41	58%	1,72
36	0,56	56%	1,51	50%	2,00
37	0,68	68%	1,21	54%	1,85
38	0,61	61%	1,49	60%	1,67
39	0,60	60%	1,43	53%	1,89
40	0,67	67%	1,34	56%	1,79
41	0,60	60%	1,30	58%	1,72
42	0,62	62%	1,44	55%	1,82
43	0,58	58%	1,29	52%	1,92
44	0,55	55%	1,57	54%	1,85
45	0,63	63%	1,31	57%	1,75
46	0,60	60%	1,46	57%	1,75
47	0,61	61%	1,36	61%	1,64
48	0,60	60%	1,43	57%	1,75
49	0,60	60%	1,25	57%	1,75
50	0,63	63%	1,34	57%	1,75
PROMEDIO	0,61	61%	1,39	55%	1,83

Tabla 167. Índices de RL en la II suministración de la prueba.

3.1 Diversidad léxica

(i) TTR

Empezamos por los datos que exponen la *Type/Token Ratio*.

Descriptivos	Suministración	
	I	II
Media	0,63	0,61
Mediana	0,64	0,61
Mínimo	0,53	0,46
Máximo	0,73	0,70

Tabla 168. Estadísticos descriptivos del TTR según la fecha de la prueba.

Los alumnos han obtenido un rendimiento mejor en la primera prueba, es decir a comienzo del año académico. Esto extraña ya que, como pensábamos y es esperable, debería darse el resultado opuesto.

En la primera encuesta todos los valores son más altos, incluso la media (0,63) y la mediana (0,64), que superan los de la segunda (0,61).

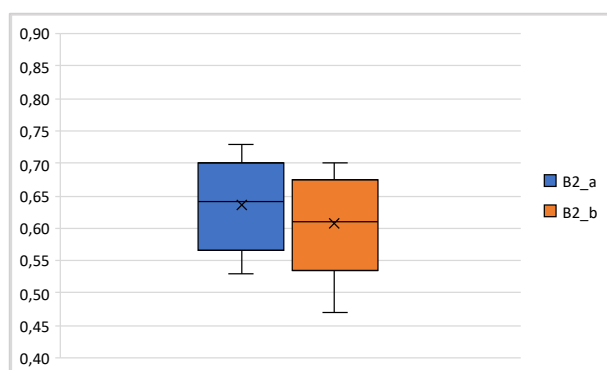


Gráfico 58. Diagrama de cajas del TTR según la fecha de la prueba.

Sobresale la oscilación intragrupal mayor de los datos del corpus B2_b, cuyos valores pasan de 0,70 a 0,46. En particular, el texto 2 es el único que tiene este resultado escaso porque en los restantes relatos de ambas agrupaciones los *types* superan o igualan la mitad del texto. El corpus B2_a que presenta una variación de 0,53 a 0,70 revela una menor desviación interna y aporta valores satisfactorios, aun mejores con respecto a los conseguidos en la segunda prueba, como enseñan la colocación y el tamaño de los diagramas.

(ii) Variación léxica

Lo mismo pasa en los cálculos del índice de variación léxica, cuyos datos expresan los valores del TTR en tantos por ciento.

Descriptivos	Suministración	
	I	II
Media	63%	61%
Mediana	64%	61%
Mínimo	53%	46%
Máximo	73%	70%

Tabla 169. Estadísticos descriptivos de la variación léxica según la fecha de la prueba.

No cabe duda de que los resultados mejores se han conseguido en la primera suministración de la prueba: los datos son más altos, sobre todo el mínimo ha bajado del 7%. Esto confirma un empeoramiento de la riqueza léxica expresada y choca con las expectativas sobre el proceso de ASL.

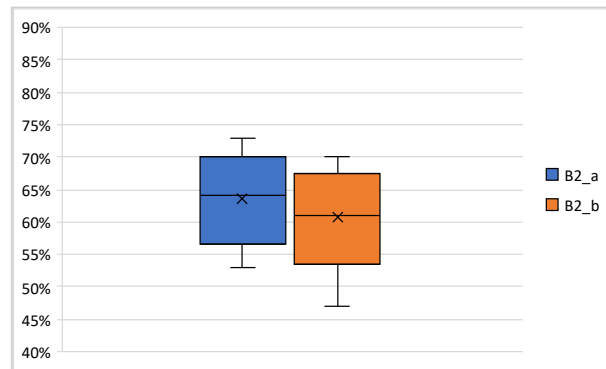


Gráfico 59. Diagrama de cajas de la variación léxica según la fecha de la prueba.

Los diagramas corroboran el bajón sufrido por los informantes al final del año académico. El resultado más escaso (46%) se ve claramente en la caja naranja, como los promedios (61%) que están por debajo de los otros y el extremo superior (70%) que coincide con el Q3 de la caja azul, la cual sube hasta el valor máximo (73%).

(iii) Índice de hápax

Por su parte, las palabras hápax aportan índices muy parecidos en las dos agrupaciones.

Descriptivos	Suministración	
	I	II
Media	1,39	1,39
Mediana	1,37	1,38
Mínimo	1,25	1,16

Máximo	1,65	1,67
--------	------	------

Tabla 170. Estadísticos descriptivos del índice de hápax según la fecha de la prueba.

La media matemática es de 1,39 en ambos subconjuntos y la mediana cambia solo de 0,01. En general, la cantidad de unidades “originales” es bastante buena ya que en el corpus B2_a hemos contabilizado el 90% de escritos con un índice igual o superior a 1,50 mientras en el B2_b el porcentaje aumenta hasta el 94%.

Los datos son medio-bajos y revelan un grado aceptable de riqueza ya que la mayoría se aproxima más a 1 que a 2. Por tanto, el índice es bueno, pero todavía mejorable debido a que debería acercarse más a 1 para que el vocabulario resultase aun más rico.

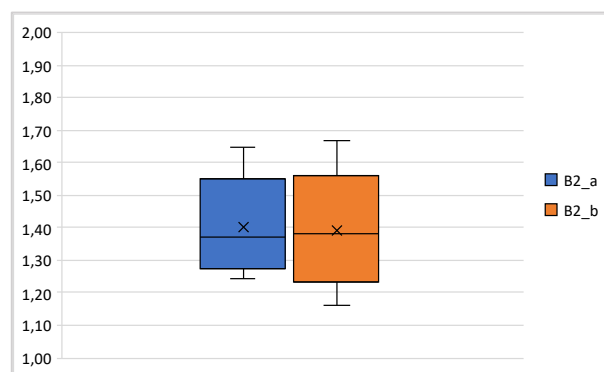


Gráfico 60. Diagrama de cajas del índice de hápax según la fecha de la prueba.

Observamos un dato relevante: la desviación intragrupal del diagrama naranja manifiesta que en la segunda prueba hay una oscilación mayor, los informantes tocan los dos extremos, sea en positivo sea en negativo, aunque el valor máximo aumenta solo del 0,02. El valor mínimo favorece este mismo muestreo, como se desprende de la extensión de su bigote inferior, que baja hasta el 1,16 incrementando el rendimiento del 0,09.

El texto 18 del corpus B2_b tiene 44 hápax y 51 *types*, mientras que el resultado mejor del corpus B2_a ve 57 hápax en el texto 9 donde los vocablos son 71. Al contrario, los relatos más pobres son, respectivamente, el 6, en el que hemos contado 36 hápax y 60 *types* y el 28 que contiene 34 palabras “originales” sobre 56 vocablos.

3.2 Densidad léxica

(i) Densidad léxica

El análisis de la densidad léxica destaca un nivel de riqueza léxica muy similar.

Descriptivos	Suministración	
	I	II
Media	54%	55%
Mediana	55%	55%
Mínimo	46%	42%
Máximo	63%	62%

Tabla 171. Estadísticos descriptivos de la densidad léxica según la fecha de la prueba.

Los valores de medio son satisfactorios y están repartidos uniformemente en los dos corpus, la mayoría de los textos tiene una cantidad de palabras llenas que supera el 50%, sugiriendo un buen grado de riqueza. Los extremos ponen de manifiesto, de nuevo, el mejor rendimiento en la primera prueba: el valor máximo se diferencia de un punto y el mínimo de cuatro, hecho que extraña si tenemos en cuenta que los alumnos han asistido a un curso de ELE entero, confirmando la tendencia detectada antes.

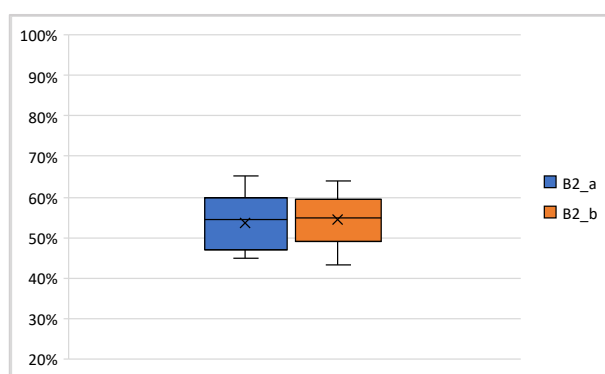


Gráfico 61. Diagrama de cajas de la densidad léxica según la fecha de la prueba.

La desviación inter e intragrupal es parecida, la caja azul solo presenta más oscilación en los intervalos intercuartiles Q1-Q2. En efecto, el 88% de los textos del corpus B2_a y el 96% del B2_b alcanzan o superan el 50% de densidad léxica.

Pese a esta diferencia, si desglosamos los índices obtenidos por cada relato hallamos con cifras que se asemejan mucho, el uso de unidades semánticas es bastante productivo ya que la cantidad de *tokens* léxicos es buena frente al total de *tokens*.

(ii) IAT

Terminamos el análisis longitudinal con los resultados obtenidos de los cálculos del IAT, mostrados en la tabla siguiente.

Descriptivos	Suministración	
	I	II
Media	1,85	1,83
Mediana	1,84	1,82
Mínimo	1,59	1,61
Máximo	2,17	2,37

Tabla 172. Estadísticos descriptivos del IAT según la fecha de la prueba.

Otra vez, si bien la media y la mediana difieren solo de 0,02 se ha detectado un rendimiento superior en la primera encuesta, cuyos datos comprenden simultáneamente el valor mínimo (1,59) y el máximo (2,17).

En la segunda suministración las cifras son más altas: los resultados empeoran. En particular, destaca que índice crece de +0,20 consiguiendo un IAT muy alto que denota la reducción de la riqueza léxica (2,37).

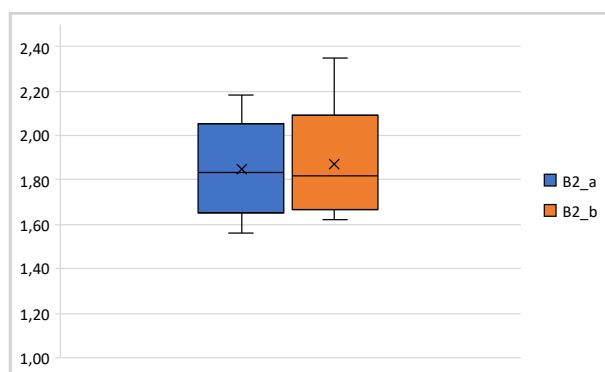


Gráfico 62. Diagrama de cajas del IAT según la fecha de la prueba.

Sobresale el desnivel entre los puntos extremos tocados por los bigotes ya que, si examinamos el tamaño de las cajas se observa cierta simetría de los resultados entre las producciones. A más y mejor, al analizar individualmente el índice de cada texto constatamos que los IAT presentados en corpus B2_b son un poco más bajos y mejores.

Se confirma que el nivel de riqueza léxica no es alto desde el momento en el que la fórmula ha proporcionado números que se acercan más a 2, ningún texto presenta un índice inferior a 1,50: el mínimo alcanzado es 1,59.

3.3 Conclusiones del análisis longitudinal

La diversidad léxica individuada mediante los índices de TTR y de variación ha

revelado que los alumnos han conseguido resultados mejores en la primera prueba: todos los datos son más altos y, por tanto, mejores.

El índice de hápax, en cambio, destaca resultados similares en las dos suministraciones: la media es la misma y las medianas varían solo de 0,01. A lo largo de ambos corpus la cantidad de hápax es apreciable ya que hemos computado el 90% en el B2_a y el 94% en el B2_b de los textos que igualan o están por debajo del 1,50.

La densidad léxica, igualmente, aporta resultados similares: el índice supera el 50% en la mayoría de los escritos, proporcionando un buen nivel de riqueza en el 88% de la primera muestra y el 96% de la segunda. Si bien haya esta discrepancia de 8 puntos, en general, hemos encontrado cifras que se asemejan mucho y conllevan un uso de las palabras temáticas bastante productivo.

Por último, el análisis del IAT también ha desvelado cierta simetría. Pese al rendimiento superior de la primera encuesta, al desglosar los textos hemos visto que, en realidad, los resultados presentados por el corpus B2_b son un poco más bajos. De todas formas, ya sabemos que esta medida no contribuye al alcance de un alto nivel de riqueza: ningún relato presenta un índice inferior a 1,50.

Si tenemos en cuenta los promedios y los descriptores estadísticos no se detecta una gran diferencia en las producciones en los dos momentos de recogida de los datos. Los resultados difieren de poco entre una y otra suministración.

Todo esto extraña ya que se esperaba que la asistencia al curso hubiera surtido efecto en la competencia de los informantes. Por el contrario, sobresalen los textos escritos al comienzo de las actividades didácticas en ciertos índices, en detrimento de los entregados al final del año académico.

Esto supone que no se ha desarrollado la capacidad de los participantes como se había hipotizado ya que se confiaba en un incremento de los índices: casi parece que la habilidad del grupo ha empeorado.

4. ANÁLISIS COMPARATIVO

Cotejamos ahora nuestros datos con los resultados de otras investigaciones sobre la riqueza léxica de aprendices no nativos de español con el objetivo de destacar correspondencias y diferencias entre los grupos.

Examinamos los índices de TTR e IAT ya que los otros investigadores no han

empleado todas las fórmulas aplicadas aquí¹¹¹.

No tenemos la pretensión de ser exhaustivos porque somos conscientes de que son varios los factores que influyen en el aprendizaje y dependen mucho del entorno en el que se desarrolla. Asimismo, tratamos textos que abarcan varias temáticas y se basan en patrones diferentes. De todos modos, planteamos este análisis ya que consideramos los informantes de cada estudio modelos estándar del estudiante de nivel intermedio de ELE.

Tras revisar la literatura sobre la riqueza léxica, hemos optado por utilizar los resultados obtenidos en los siguientes trabajos al estar dedicados a aprendices de distintas procedencias que tienen el mismo dominio del español de nuestros encuestados.

Cuba Vega y Cuba (2004) mide la riqueza léxica de estudiantes brasileños producidas en textos con extensión fija de 100 *tokens*. Berton (2014) examina la producción de aprendientes suecos que estudian español como L3. Recoge 180 relatos de diferente tamaño, de los cuales hemos tenido en cuenta los que más se acercaban a las características de nuestro corpus, es decir los que llegan a un máximo de 117 *tokens*. Wang (2016), siguiendo la metodología de Cuba Vega y Cuba (2004), trabaja con discentes sinohablantes. Basso (2017) estudia un corpus de textos escritos por aprendices italófonos universitarios de diferentes longitudes del que hemos extraído los que no exceden los 100 *tokens*.

La tabla expone las características principales de las investigaciones cotejadas.

Investigación	País	Máx. extensión textos	Nivel educativo
Nalesso (2019)	Italia	100	Universidad
Cuba Vega y Cuba (2004)	Brasil	100	Universidad
Berton (2014)	Suecia	117	Bachillerato
Wang (2016)	China	100	Universidad
Basso (2017)	Italia	100	Universidad

Tabla 173. Datos generales de RL de las investigaciones cotejadas.

Si excluimos a Berton (2014), todos los trabajos están centrados en la medición

¹¹¹ Entre las investigaciones contempladas solo una no ha aplicado el IAT, o mejor ningún índice que computase la densidad léxica. Tal cual explica su autor: “La medida de la densidad léxica no parece útil para este estudio, ya que en las producciones de aprendices principiantes e intermedios como los informantes del presente estudio el uso de palabras funcionales, como por ejemplo los marcadores del discurso, es muy reducido” (Berton, 2014: 28). Efectivamente, el trabajo se refiere al grado de diversidad léxica en la producción escrita de los informantes.

de la riqueza léxica de aprendices de español LE universitarios de nivel intermedio y analizan redacciones constituidas por el mismo número de tokens, por lo que la comparación resulta equilibrada y coherente.

4.1 Análisis descriptivo

Empezamos analizando el TTR, la tabla contiene los promedios extraídos de las distintas investigaciones que son nuestro punto de partida para la comparación.

Investigación	TTR
Nalesso (2019)	0,63
Cuba Vega y Cuba (2004)	0,64
Berton (2014)	0,70
Wang (2016)	0,60
Basso (2017)	0,60
PROMEDIO	0,63

Tabla 174. Promedios del TTR de las investigaciones cotejadas.

La media matemática contabilizada es de 0,63, la cual revela que la diversidad léxica es buena para todos los aprendices ya que supera el 50% del texto: la presencia de *types* prevalece sobre la totalidad de los *tokens*.

Dentro de este conjunto, Berton (2014) alcanza las cifras más elevadas que le han permitido contabilizar una media de 0,70. Siguen los informantes de Cuba Vega y Cuba (2004) que obtienen un TTR de 0,64 y los nuestros que consiguen un 0,63. Las últimas posiciones están ocupadas por los encuestados chinos de Wang (2016) y los italianos de Basso (2017) que presentan un promedio de 0,60.

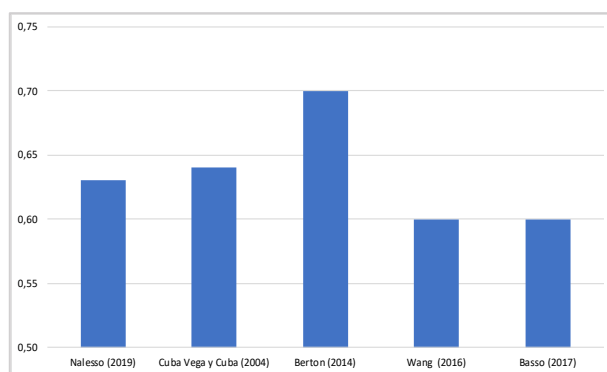


Gráfico 63. Promedios del TTR de las investigaciones cotejadas.

La desviación máxima es de 0,10 entre los que han arrojado resultados más altos

y los más bajos, la variedad léxica producida no es exactamente la misma para todos, aunque el desnivel entre uno y otro no es tan notable.

Del mismo modo, el contraste entre los índices del IAT, donde el número de estudios comparables se reduce a cuatro, pone de manifiesto que la densidad de vocabulario tampoco consigue resultados muy diferentes.

Investigación	IAT
Nalesso (2019)	1,84
Cuba Vega y Cuba (2004)	1,81
Wang (2016)	1,80
Basso (2017)	2,35
PROMEDIO	1,95

Tabla 175. Promedios del IAT de las investigaciones cotejadas.

El promedio es de 1,95, cifra que se aleja del valor ideal de este índice (recuérdese, 1) e implica un bajo grado de riqueza léxica, ya que supone que las léxias con mayor presencia sean gramaticales con respecto a las nocionales. Igualmente, los promedios de cada investigación se manifiestan elevados.

Los sinohablantes aportan la media mejor, si bien elevada (1,80). Después, Cuba Vega y Cuba (2004) han calculado un IAT igualmente alto (1,81) seguido por el nuestro que es de 1,84. Por último, Basso (2017) registra el peor índice en absoluto (2,35) que supone la aparición de una palabra temática cada más de dos atemáticas en la producción de sus aprendices, lo cual sería +0,55 al compararlo con Wang (2016).

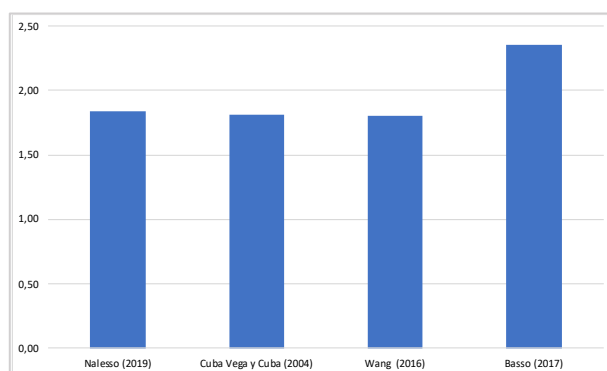


Gráfico 64. Promedios del IAT de las investigaciones cotejadas.

Al excluir el resultado de Basso (2017), los demás índices varían muy poco: este indicador sugiere que los aprendices que tienen este nivel lingüístico han desarrollado la misma capacidad de utilizar las palabras nocionales en la producción de eductos escritos.

4.2 Conclusiones del análisis comparativo

Antes de terminar, recordamos que la modalidad de ejecución del análisis y la composición de la muestra no corresponden perfectamente en los estudios objeto de la comparación, con lo cual puede que las conclusiones se vean afectadas por las limitaciones que estos cambios les imponen.

En todo caso, la homogeneidad detectada en los datos de TTR confirman los buenos resultados obtenidos si consideramos el nivel de ELE de los sujetos participantes.

Sorprenden los datos de Berton (2014), quien trabaja con alumnos de bachillerato que estudian español como L3 que superan el rendimiento de los otros aprendices, universitarios que estudian español como L2, de los que se esperaría un rendimiento mejor.

Asimismo, extrañan los escasos resultados expuestos en Wang (2016), cuyos encuestados han arrojado uno de los TTR más pobres (0,60) si consideramos que en el análisis de la disponibilidad léxica los sinohablantes (Hidalgo, 2017b) han obtenido resultados sobresalientes.

Por otro lado, el IAT proporciona cifras altas que implican una menor riqueza porque cuanto más pequeño es su valor, más denso de palabras semánticas es el vocabulario. Sin embargo, esto confirma la teoría de Johansson (2008) que afirma que una buena variedad léxica no implica una igualmente buena densidad.

Aunque los dos índices cotejados han aportado informaciones diferentes, parece que todos los estudiantes han entregado composiciones aceptables y saben desempeñar una tarea de producción escrita simple y no extensa.

Quizá ampliando la longitud de los textos la variación del vocabulario disminuya, como afirma Berton (2014: 47): “[se] confirma los hallazgos de estudios anteriores (Arnaud, 1984; Richards, 1987; Malvern *et al.* 2004), es decir, la reducción del valor del TTR al aumentar la extensión de la muestra de lengua considerada.” Basso (2017) y Nalesso (2019b) han corroborado una vez más que la *Type/Token Ratio* decrece de manera progresiva a medida que aumenta el número de *tokens* textuales, esto es, cuanto más largo es un texto, menos rico resulta el vocabulario¹¹².

¹¹² Berton (2014) ha computado un decrecimiento del TTR del 30,26%:
- Textos de 1 a 39 *tokens*: TTR 0,76;

5. CONCLUSIONES DEL ANÁLISIS CUANTITATIVO DE LA RIQUEZA LÉXICA

Hemos dividido el análisis cuantitativo de la riqueza léxica en tres secciones: en la primera hemos realizado el estudio transversal del corpus, en la segunda hemos desarrollado el análisis longitudinal, en la tercera hemos cotejado nuestros resultados generales con los de cuatro investigaciones dedicadas a otros aprendices de ELE.

El análisis transversal ha revelado que el índice de la diversidad léxica es positivo ya que todos los informantes han presentado una adecuada relación entre los *types* utilizados en sus textos conforme al número de *tokens* (en promedio 0,63). La variación también es buena porque las palabras diferentes cubren más del 50% de cada relato (en promedio 63%). Igualmente, el índice de hápax destaca un aceptable nivel de riqueza léxica: la correlación entre los vocablos y las palabras “originales” ha permitido contabilizar cifras medio-bajas que son positivas en este caso (en promedio 1,39).

La densidad léxica, al contrario, ha evidenciado un rendimiento inferior. Por un lado, el 90% del grupo ha escrito en sus textos más del 50% de palabras temáticas (en promedio 55%), pero, del otro lado, el IAT es alto: va de 1,52 a 2,33 (en promedio 1,84), señalando la presencia de una palabra de contenido nocional cada dos gramaticales.

Como este indicador, el intervalo de aparición de palabras nocionales, se ha revelado alto en todo tipo de análisis no vamos a discutirlo más.

Estos datos comprueban que un texto con un buen índice de variedad léxica no necesariamente tiene el mismo resultado en lo que se refiere al uso de las unidades nocionales y se deduce que la aplicación de diferentes medidas de cálculo es esencial para obtener datos lo más completos y fiable posible a la hora de analizar un conjunto textual (Read, 2000; Johansson, 2008; Jarvis, 2013).

El análisis por variable ha puesto de manifiesto que estas no son los causantes de

-
- Textos de 40 a 78 *tokens*: TTR 0,65;
 - Textos de 79 a 117 *tokens*: TTR 0,53.

Basso (2017) ha calculado los siguientes valores, que suponen una disminución cada vez más importante del índice, hasta llegar a un -26,67%:

- Textos de 100 *tokens*: TTR 0,60;
- Textos de 150 *tokens*: TTR 0,53;
- Textos de 200 *tokens*: TTR 0,48;
- Textos de 250 *tokens*: TTR 0,44.

Nalesso (2019b) ha notado la misma tendencia, aunque en menor medida (-8,77%):

- Textos de 108 a 144 *tokens*: TTR 0,57;
- Textos de 151 a 200 *tokens*: TTR 0,55;
- Textos de 203 a 246 *tokens*: TTR 0,52.

una mayor o menor riqueza léxica. En particular, el género de los alumnos y el dominio lingüístico no parecen influir en la competencia del aprendiz.

(i) Mujeres y hombres tienen el mismo nivel de diversidad léxica ya que los *types* utilizados igualan o superan el 50% de cada uno de los textos, así como el índice de hápax, bueno en sendos casos.

Aunque el desglose de los resultados inherentes a la densidad muestra una situación distinta que evidencia un ligero influjo del factor, que favorece los varones: en sus relatos más del 50% de palabras es nocional, mientras que este porcentaje se detecta en el 89% de los textos de sus compañeras.

(ii) El nivel de conocimiento de la LO tampoco incide en la riqueza de los encuestados. Se ha observado que los índices no experimentan ni un aumento ni un decrecimiento. La variedad léxica sobrepasa el 50% en los relatos de ambos subconjuntos, si bien el índice de hápax favorece levemente el avanzado.

La densidad invierte la tendencia ya que el 92% de los informantes B1 y el 88% de los B2 presenta un índice que iguala o rebasa el 50%. La única diferencia destacable es la mayor oscilación intragrupal de la variante B1, cuya escala de valores es más amplia, tanto que parece implicar un conocimiento léxico no bien definido de los estudiantes. Por otra parte, la desviación interna de la variante B2 se circunscribe en una serie más restringida de valores que supone conocimientos más compactos. De todas formas, no vemos la predominancia de ningún grupo: los índices no cambian significativamente, parece casi que el desarrollo de la competencia de un nivel al otro se ralentiza, contrariamente a lo esperado.

Pese a que estos no sean los resultados en los que se confiaba, confirman la hipótesis de que para los aprendices italianos de español la etapa más compleja del aprendizaje es el nivel intermedio a causa de la afinidad entre LM y LE: ya tienen desarrollados los conocimientos que les permiten comunicarse y el avance hacia un dominio mayor de la lengua (es decir, progresar de un nivel umbral a uno avanzado, según las indicaciones del MCER) le solicita más esfuerzo, tanto que la progresión lingüística es menor¹¹³.

¹¹³ Esto corrobora la teoría de Calvi (1995) según la cual esta fase del aprendizaje de español es un obstáculo para los discentes italófonos. Para solucionar este *impasse*, que puede llevar a la fosilización del nivel alcanzado, la autora sugiere el diseño de una propuesta didáctica basada en la contrastividad, asumiendo que a partir de ese momento los aprendices han de enfrentarse con estructuras y reglas que se alejan de lo que han estudiado hasta la fecha y les ha parecido “demasiado fácil”.

(iii) El conocimiento de otras lenguas parece la variable que más influye en la riqueza de los informantes. El desnivel entre las dos variantes es ligeramente más pronunciado aquí respecto a los factores que acabamos de discutir.

Primero, los estudiantes que conocen más de dos LE presentan una diversidad léxica superior (sobrepasan del 9% los que conocen solo dos LE en los cómputos de variación léxica y del 16% en el índice de hápax).

Segundo, se ha detectado un cambio al examinar la densidad puesto que los informantes del grupo LE =2 han empleado una cantidad mayor de palabras nocionales.

Por su parte, los resultados del estudio longitudinal no han manifestado un gran desnivel en las producciones de los participantes, los valores difieren de poco entre una y otra administración de la prueba.

La *Type/Token Ratio* y la variación léxica han sido mejores en la primera. Los índices de hápax y de la densidad léxica se asemejan en las dos muestras: se ha conseguido un buen uso de las unidades semánticas en ambas fechas de administración.

El IAT presenta cifras un poco mejores en los relatos escritos al final del curso: los datos del corpus B2_b son más pequeños con respecto a los del B2_a.

Esto confirma una detención en el avance del aprendizaje, contrariamente al aumento esperado. Es más, parece que la capacidad expresiva ha empeorado desde el momento en el que algunos índices favorecen los textos entregados al principio del año académico.

En cuanto a la correlación de los resultados de diferentes investigaciones sobre la riqueza léxica en ELE podemos afirmar que, considerando el dominio lingüístico, todos los informantes han escrito textos del mismo nivel, cuyas características son adecuadas a sus capacidades.

Los indicadores satisfacen las expectativas, pese a los escasos datos aportados por los promedios del IAT, confirmando lo planteado en Berton (2014) y averiguando que esta tipología de aprendiz obtiene un buen resultado al escribir un texto simple y no extenso, pero puede que un aumento de los *tokens* baje tal rendimiento (Berton, 2014; Basso, 2017; Nalesso, 2019b).

A modo de conclusión, traemos a colación la importancia de seguir planteando proyectos dedicado a los hablantes no nativos de español en este ámbito investigativo y

fijar un protocolo empírico común que permita la realización de comparaciones más detalladas. La medición de la riqueza léxica, efectivamente, ofrece posibilidades interesantes de observación de la competencia léxica, y comunicativa, de los participantes como lo ha demostrado este estudio con sus carencias y limitaciones.

CAPÍTULO VII.

APROXIMACIÓN A UN ANÁLISIS CUALITATIVO DE LA RIQUEZA LÉXICA

1. INTRODUCCIÓN

Sabemos que los índices cuantitativos de la riqueza léxica conllevan otras informaciones además de las numerales, recordamos la “calidad de la escritura” de López Morales (2011), lo cual quiere decir que el estudio cuantitativo, implica consideraciones cualitativas de los textos escritos por nuestros informantes. Sin embargo, en este capítulo proponemos un acercamiento a un análisis cualitativo a partir de las palabras actualizadas en el conjunto textual. Para ello, someteremos el corpus de trabajo a un análisis finalizado a examinar morfológicamente las unidades léxicas más utilizadas.

Nos centraremos, en particular, en las palabras de contenido nocional dividiéndolas en cuatro categorías: sustantivos, verbos, adjetivos y adverbios. Indagamos, primero, las listas de frecuencia recopiladas con todos los *types* y solo con los *types léxicos* extraídas del listado de la frecuencia general; segundo, el registro de las palabras clave que han surgido de las producciones. Como medida de corte hemos decidido tener en cuenta los 25 vocablos más frecuentes de cada lista.

Recordamos, antes de la discusión de los resultados, que el primer paso que hemos realizado antes del conteo ha sido la lematización de los *tokens* de manera que las formas flexionadas pudiesen ser contadas como un único lema, por lo que en las siguientes tablas se encontrarán unidades léxicas no marcadas.

Cabe señalar que se trata de un análisis piloto en este ámbito ya que nunca se ha propuesto un estudio de esta naturaleza en los trabajos dedicados a la riqueza léxica. Esperamos que la metodología empleada y las modestas conclusiones que sacaremos puedan constituir un punto de partida para ulteriores investigaciones.

2. LISTAS DE FRECUENCIA

A continuación, la repartición de las unidades léxicas se ilustra mediante histogramas, cuyas barras muestran los porcentajes de ocurrencia de las distintas clases de palabras, y tablas organizadas según rango y frecuencia de aparición.

2.1 Análisis general

En el corpus analizado, resultante de los 100 textos entregados, los *tokens* se elevan a 10.000 y el número de los vocablos, incluidas las formas flexionadas de una misma palabra, es de 1.072. Estos se dividen con una notable disparidad entre palabras funcionales del discurso (7%) y palabras de contenido nocional (93%), datos que reflejan los valores de la densidad léxica ya observados y discutidos.

El gráfico ilustra claramente dicho desnivel.

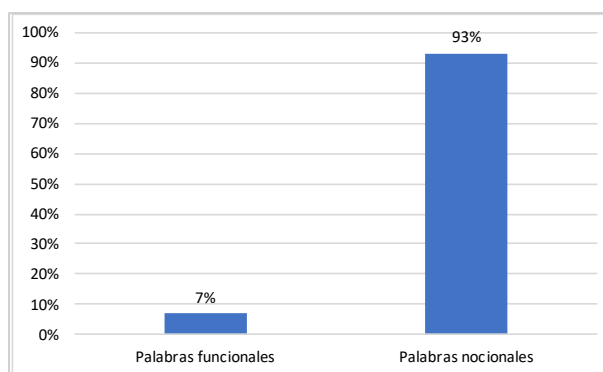


Gráfico 65. Distribución de las palabras en la lista de frecuencia del corpus general de RL.

Pese a esta distribución, los ítems gramaticales son los que se colocan en las primeras posiciones de la lista de las palabras más frecuentes del entero corpus.

La tabla 176 indica que se detectan artículos, conjunciones, preposiciones y pronombres como las unidades más utilizadas (11 sobre 25 en total), como era esperable, si excluimos el verbo *ser* que hemos considerado como nocional pese a ser auxiliar y copulativo debido al bajo nivel de ELE de los encuestados.

El verbo *ir* se ubica en la posición 10 en medio de las partículas que más aparecen en los relatos; mientras que desde la posición 14 encontramos solamente palabras temáticas entre las que sobresale la presencia de otros verbos (*hacer, gustar, estar, tener, visitar*), por encima de la aparición de adverbios, adjetivos y sustantivos.

Rango	Vocablo	Frecuencia
1	el	870
2	y	436
3	de	415
4	ser	305
5	en	289

6	que	287
7	un	273
8	a	231
9	mi	185
10	ir	153
11	me	152
12	con	136
13	por	126
14	ciudad	115
15	mucho	109
16	viaje	107
17	hacer	106
18	todo	95
19	más	95
20	día	94
21	gustar	93
22	muy	92
23	estar	83
24	tener	81
25	visitar	80

Tabla 176. Lista de las 25 palabras más frecuentes del corpus general de RL.

2.2 Análisis de las palabras nocionales

Al excluir del análisis las unidades funcionales, la lista de las palabras semánticas más frecuentes presenta la siguiente distribución: la gran mayoría pertenece a la clase de los sustantivos (48%); un número menor de verbos (24%); una cifra más pequeña de adjetivos (20%) y una cantidad escasa de adverbios (8%).

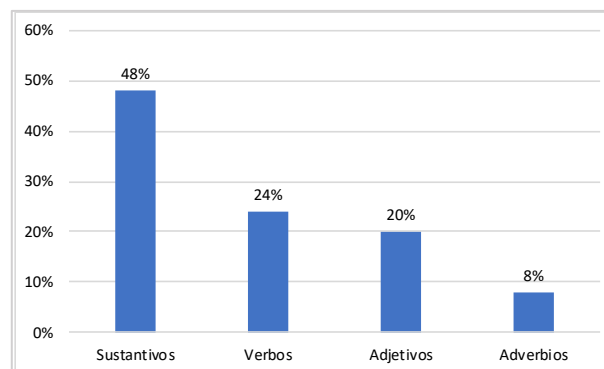


Gráfico 66. Distribución de las palabras nocionales en el corpus general de RL.

Por su parte, la tabla 177 contiene los 25 vocablos incluidos en estos porcentajes,

esto es, los que se han manifestado más veces en el muestreo sin distinción de categoría morfológica.

Destaca el hecho de que la mayor parte de los lemas está estrechamente vinculada a la temática propuesta para la redacción del texto (recordamos, un viaje que le había gustado a los informantes), con lo cual ya a raíz de esta selección se comprendería el asunto del que tratan los relatos.

Contamos con verbos (*ir, visitar, ver, llegar*) y sustantivos (*ciudad, viaje, lugar*) representativos del tema y, es más, al observar los *types* más frecuentes se añaden otros lexemas emblemáticos como *tomar, avión, maravilloso, isla, museo*.

Rango	Vocablo	Frecuencia
1	ser	305
2	ir	153
3	ciudad	115
4	mucho	109
5	viaje	107
6	hacer	106
7	todo	95
8	día	94
9	gustar	93
10	muy	92
11	estar	83
12	tener	81
13	visitar	80
14	año	78
15	amigo	62
16	allí	58
17	también	53
18	ver	47
19	pasado	42
20	como	41
21	semana	39
22	llegar	37
23	haber	37
24	lugar	35
25	familia	32

Tabla 177. Lista de las 25 palabras nocionales más frecuentes del corpus general de RL.

2.2.1 Sustantivos

Dentro del conjunto de las palabras de contenido nocional, los sustantivos componen casi la mitad de las unidades léxicas más frecuentes, el 48%. Los lexemas que tienen el mayor número de apariciones son *ciudad* con 115 y *viaje* con 107, seguidas por *día*, *año* y *amigo* que superan las 50 ocurrencias.

Si analizamos el repositorio de sustantivos empleados en el corpus, hay 212 palabras hápax que constituyen el 46% del total, entre los que contamos con algunos nombres de ciudad o localidad turística, ya que hemos aceptado los topónimos en español (*Bolonia*, *Londres*, *Magaluf*, *Mallorca*), si bien hemos descartado todos los demás nombres propios, así como las palabras en otros idiomas.

Rango	Vocablo	Frecuencia
1	ciudad	115
2	viaje	107
3	día	95
4	año	78
5	amigo	62
6	semana	39
7	lugar	35
8	vez	33
9	familia	32
10	Londres	29
11	mañana	28
12	gente	25
13	avión	23
14	casa	23
15	isla	21
16	novio	20
17	museo	20
18	noche	20
19	playa	20
20	compañero	19
21	cosa	19
22	escuela	18
23	hermano	18
24	aeropuerto	17
25	experiencia	16

Tabla 178. Lista de los 25 sustantivos más frecuentes del corpus general de RL.

2.2.2 Verbos

El 24% de las palabras más frecuentes corresponde a la categoría verbal. Dentro de esta cifra, al excluir el auxiliar *ser* (292 apariciones), encontramos una amplia gama de verbos entre los que *ir* es el que más se ha utilizado (166 apariciones) antes de *hacer*, *gustar*, *tener*, *estar*, *visitar*, *pasar* que presentan más de 50 ocurrencias respectivamente. Otros vocablos, si bien considerados significativos en este ámbito temático (por ejemplo, *llegar*, *salir*, *viajar*) se han manifestado en un número reducido de ocurrencias.

Hemos apreciado la actualización de algunas unidades pluriverbales repetidas con el verbo *tomar* como *tomar el avión*, *tomar el tren*, *tomar el autobús*, *tomar el sol* (16 ocurrencias sobre un total de 26) y apariciones de la perífrasis verbal *ir + a + infinitivo*, como *ir a bailar*, *ir a nadar*, *ir a visitar*, aunque no se haya podido incluirlas en la tabla debido al número de apariciones.

Para terminar, señalamos que los hápax contabilizados en esta sección son el 50% sobre el total de tipos de verbos aportados.

Rango	Vocablo	Frecuencia
1	ser	292
2	ir	166
3	hacer	110
4	gustar	93
5	tener	83
6	estar	83
7	visitar	82
8	pasar	60
9	ver	47
10	llegar	38
11	conocer	29
12	poder	28
13	tomar	26
14	quedar	24
15	vivir	21
16	comer	20
17	encantar	20
18	encontrar	17
19	salir	15
20	divertir	14
21	viajar	14

22	haber	14
23	decidir	12
24	dar	12
25	hablar	11

Tabla 179. Lista de los 25 verbos más frecuentes del corpus general de RL.

2.2.3 Adjetivos

Los adjetivos cubren el 20% de la lista de las palabras temáticas más frecuentes. Dentro de la agrupación *primero* es la unidad que más se ha reiterado, sobre todo en la expresión *la primera vez* (repetida 19 veces en 34 ocurrencias totales del ítem).

Los otros lexemas más utilizados son *maravilloso*, *grande*, *pequeño* que tienen más de 20 apariciones.

Como muestra la tabla 180, y como es lógico teniendo en cuenta la tipología de texto examinada, se observa que la mayoría de los adjetivos son calificativos ya que expresan cualidades o propiedades del objeto designado, incluso si alargamos el análisis al entero listado.

Los hápax se presentan en cantidad notable (56%) y manifiestan, al mismo tiempo, un alto grado de variedad de vocabulario en el uso de esta categoría morfológica.

Rango	Vocablo	Frecuencia
1	primero	34
2	maravilloso	24
3	grande	22
4	pequeño	20
5	típico	19
6	mejor	18
7	diferente	17
8	nuevo	16
9	estupendo	15
10	último	15
11	bueno	14
12	español	13
13	junto	12
14	famoso	11
15	siguiente	11
16	solo	10
17	importante	9

18	lleno	9
19	particular	9
20	amable	7
21	fantástico	6
22	largo	6
23	verde	5
24	único	5
25	vario	5

Tabla 180. Lista de los 25 adjetivos más frecuentes del corpus general de RL.

2.2.4 Adverbios

La clase de los adverbios es la que menos aparece entre las palabras semánticas (8%). Los de mayor uso son *más*, *muy* y *mucho* (también en la forma superlativa *muchísimo*) que tienen, de manera respectiva, 95, 92 y 79 repeticiones en el muestreo. Los demás *types* incluidos en la tabla de los 25 más frecuentes presentan un número bastante inferior de ocurrencias, hasta llegar a 4 (*ahora*).

Constatamos la considerable actualización de las siguientes locuciones de tiempo: *por la mañana*, *por la tarde*, *por la noche*.

Los adverbios en *-mente* creados a partir de adjetivos son los siguientes, en orden alfabético: *afortunadamente*, *contrariamente*, *desafortunadamente*, *estupendamente*, *exactamente*, *exageradamente*, *inmediatamente*, *magníficamente*, *mediamente*, *normalmente*, *obviamente*, *particularmente*, *perfectamente*, *prácticamente*, *precisamente*, *seguramente*, *solamente*, *temporáneamente*, *totalmente*, *tranquilamente*, *únicamente*, *verdaderamente*. Revelan una frecuencia va de un máximo de 3 a un mínimo de 1, esto es, se trata en gran mayoría de palabras hápax. A este propósito, el porcentaje de los lexemas que aparecen sólo una vez en la totalidad de adverbios contemplados es del 37%.

Rango	Vocablo	Frecuencia
1	más	95
2	muy	92
3	mucho	79
4	allí	58
5	también	53
6	no	47
7	después	28
8	cerca	18

9	siempre	18
10	nunca	15
11	así	14
12	además	12
13	bien	10
14	ya	8
15	luego	7
16	casi	6
17	solo	6
18	aquí	6
19	antes	6
20	poco	5
21	bastante	5
22	tanto	5
23	lejos	5
24	entonces	5
25	ahora	4

Tabla 181. Lista de los 25 adverbios más frecuentes del corpus general de RL.

3. PALABRAS CLAVE

En esta sección, observamos cuáles son las palabras clave del corpus.

La tabla 182 encierra las unidades léxicas extraídas mediante la herramienta *Keywords* de *Sketch Engine*, que permite definir cuál es el tema principal de los textos analizando sus lexías¹¹⁴ y comparándolas con las que forman parte del corpus general, el *Spanish Web corpus 2018*, tomado como referencia.

Sobresale la presencia masiva de sustantivos, seguidos por verbos y adjetivos, que concurren a establecer la temática abarcada en los textos: se deduce que se trata el asunto de los viajes, en particular gracias a las unidades *vacación, viaje, visitar, avión, monumento, típico, aeropuerto, playa, isla, museo, tren*. Detectamos, además, la presencia de lexemas que implican acciones, lugares, medios de transporte, destinos y sensaciones ligadas a la experiencia del viaje.

Se señala que, aunque ciertos vocablos han alcanzado un índice de frecuencia bajo, se encuentran en posiciones más altas de la lista con respecto a las que habían

¹¹⁴ El *software* permite también la extracción de términos clave (entendidos como unidades multipalabra), pero, debido al hecho de que no hemos detectado una cantidad significativa en nuestro corpus, hemos contemplado solamente la unidad palabra.

presentado en las listas previamente analizadas porque son ítems léxicos que efectivamente representan el tema abordado.

Notamos lo mismo al comparar las frecuencias contenidas en esta tabla con el uso que se suele hacer de las mismas palabras en contextos comunicativos generales.

Rango	Vocablo	Frecuencia en Nalesso (2019)	Frecuencia en corpus de referencia
1	vacación	12	2.403
2	estupendo	15	113.919
3	viaje	107	1.198.008
4	Londres	29	317.393
5	maravilloso	24	330.841
6	visitar	82	1.264.694
7	pasear	10	175.041
8	avión	23	433.906
9	monumento	12	227.479
10	típico	19	378.160
11	novio	20	432.929
12	gustar	93	241.1475
13	coger	10	258.519
14	aeropuerto	17	458.292
15	allí	58	1.613.761
16	encantar	20	587.578
17	divertir	14	455.357
18	paisaje	10	342.996
19	playa	20	725.652
20	isla	21	773.341
21	museo	20	755.141
22	tren	10	374.200
23	bar	10	397.360
24	verano	16	657.445
25	amigo	62	2.631.598

Tabla 182. Lista de las palabras clave del corpus general de RL.

Ahora bien, aprovechando del instrumento *Word Sketch*, analizamos los distintos usos de algunos de estos vocablos en el discurso de los informantes mediante las siguientes figuras que ilustran cómo se han combinado con otras unidades¹¹⁵.

¹¹⁵ Es oportuno recordar que aparecen sin flexión morfológica debido al proceso de lematización que hemos realizado antes del tratamiento informático de los textos.

Elegimos dentro del listado de las palabras clave, un ítem para cada una de las categorías morfológicas que hemos tenido en cuenta: un sustantivo, un verbo, un adjetivo y un adverbio.

El sustantivo *vacación* encabeza la lista, pues, resulta la primera palabra clave del corpus, aunque presenta solo 12 ocurrencias en total. De la imagen se ve que se han creado distintos tipos de relaciones en combinación con las unidades que siguen:

- (i) Preposiciones de lugar, tiempo, modo y compañía;
- (ii) El posesivo *nuestro*;
- (iii) Nombres comunes con los que forma unidades multipalabra;
- (iv) Verbos que lo tienen como objeto;
- (v) Adjetivos que funcionan como atributos.



Figura 3. Uso del vocablo *vacación* en el corpus general de RL.

En segunda instancia, *visitar* es el primer verbo que aparece en la lista, se ha acompañado con:

- (i) Sustantivos (nombres comunes y propios) con la función de objeto;
- (ii) Adverbios demostrativos, de modo y temporales;
- (iii) La preposición de modo e instrumento *con*;
- (iv) La preposición *a* que desempeña la función de acusativo personal.



Figura 4. Uso del vocablo *visitar* en el corpus general de RL.

Estupendo se coloca primero en la lista de los adjetivos, ubicado en el rango 2. Los encuestados lo han utilizado con las palabras que vemos a continuación:

- (i) Verbos copulativos (*ser* y *parecer*), con los que forma predicados nominales, teniendo la función de atributo;
- (ii) Nombres comunes para los cuales es un adjetivo calificativo.



Figura 5. Uso del vocablo *estupendo* en el corpus general de RL.

Para terminar, observamos en la lista de las palabras clave *allí*, adverbio de tiempo y de lugar. Vemos que se ha empleado en los textos junto a:

- (i) Verbos, cuyo significado completa ejerciendo el papel de complemento.
- (ii) Otros adverbios, temporales y de modo.



Figura 6. Uso del vocablo *allí* en el corpus general de RL.

4. CONCLUSIONES DEL ANÁLISIS CUALITATIVO DE LA RIQUEZA LÉXICA

Este primer planteamiento de análisis cualitativo de la riqueza léxica nos permite afirmar que, dejando de lado índices y números, es posible realizar una investigación de distinta naturaleza sobre los resultados, finalizado al estudio de la forma más que a la cantidad de los datos, si bien nos hemos fundado en el criterio de frecuencia de las palabras.

En la sección dedicada al análisis de los listados de frecuencia, se ha observado que, en general, el 93% de los vocablos que componen la lista de las palabras más frecuentes está formado por palabras de contenido nocional, no obstante en los primeros rangos se detecta una mayoría de unidades funcionales del discurso.

En lo que atañe a las unidades nocionales

- (i) los sustantivos son la categoría más presente. *Ciudad* y *viaje* son los lexemas más utilizados en los textos, seguidos por otros nombres comunes y topónimos;
- (ii) los verbos *ser* e *ir* son los que han tenido más ocurrencias y, a continuación, constatamos el uso de verbos predicativos y de unidades pluriverbales con *tomar* y de la perífrasis *ir + a + infinitivo*;
- (iii) los adjetivos más frecuentes son calificativos, señalan una cualidad del sustantivo, ya sea concreta o abstracta;
- (iv) los adverbios son de vario tipo (de tiempo, lugar, modo, cantidad) y se presentan en forma simple, mientras que los compuestos (en *-mente*) tienen menos ocurrencias y, por tanto, se hallan en posiciones inferiores.

En el siguiente apartado, la extracción de las palabras clave nos ha permitido averiguar cuáles son las unidades léxicas que definen el tema abordado en el corpus. Dentro de esta lista hemos destacado cuatro vocablos (*vacación, visitar, estupendo, allí*) representativos para cada clase gramatical contemplada (sustantivos, verbos, adjetivos, adverbios) y hemos procedido al análisis de los usos más comunes que han hecho los estudiantes en sus relatos.

Todo esto, nos ha llevado a una nueva perspectiva de estudio de la riqueza léxica: hemos conseguido, a nuestro parecer, resultados interesantes a partir de los cuales esperamos despertar el interés de otros investigadores.

Finalizamos apuntando que se trata de un trabajo breve e introductorio que deja el camino abierto, quizá, a una ampliación de las propuestas y de los estudios empíricos.

CONCLUSIONES

La presente Tesis consiste en un experimento transversal y, por primera vez, longitudinal encaminado al estudio de la competencia léxica de un grupo representativo de los estudiantes italianos de ELE con la finalidad de abrir el camino a aspectos no tratados con profundidad, o no tratados en absoluto, en otros proyectos enmarcados en el ámbito de los estudios sobre el aprendizaje del léxico en español lengua extranjera. Por un lado, hemos realizado el análisis cuantitativo, cualitativo y comparativo del léxico disponible de la muestra mediante el estudio de la *disponibilidad léxica*; por otro lado, hemos utilizado la metodología de la *riqueza léxica* con el propósito de examinar el vocabulario activo de los encuestados mediante el análisis de un corpus de textos escritos.

No es el primer proyecto que aborda el aprendizaje del léxico de ELE, pero se ha realizado con un enfoque original: la correlación de los datos procedentes de las dos metodologías citadas con la intención de evaluar la competencia de los participantes y su desarrollo en el tiempo.

El volumen del corpus, adecuado para un trabajo de estas características, nos ha permitido extraer resultados concluyentes en cuanto a la competencia léxica del conjunto de aprendices contemplado, tanto en el estudio y la comparación intramuestral (en función de las tres variables sociolingüísticas y de la fecha de suministración de la prueba), como del contraste intermuestral (con otros discentes de ELE) y con los corpus CREA y CAES.

Todo esto ha posibilitado el alcance de los objetivos investigadores que nos fijamos en un principio:

- (i) hemos analizado el léxico disponible y activo de una muestra de estudiantes universitarios italianos de español;
- (ii) hemos evaluado la incidencia de los factores sociolingüísticos sexo, nivel de ELE y conocimiento de otras lenguas extranjeras en el bagaje léxico de los encuestados;
- (iii) hemos examinado la evolución en el tiempo de la disponibilidad y la riqueza de vocabulario del grupo;
- (iv) hemos comparado nuestros datos con los obtenidos en otros proyectos

dedicados a estudiantes de ELE;

(v) hemos elaborado un diccionario del léxico disponible.

La Tesis empieza con la exposición del marco teórico y el estado de la cuestión de los que partimos, los primeros capítulos abarcan una revisión teórica e historiográfica de elementos relacionados con la didáctica del vocabulario de las LE, la disponibilidad y la riqueza léxica.

Hemos reseñado la influencia de la lingüística aplicada en la enseñanza de las lenguas extranjeras, cuyas teorías (conductista, nativista, cognitivista, ambientalista, interaccionista) permiten estudiar los complejos procesos que subyacen al desarrollo lingüístico, esto es, explican cómo funcionan los mecanismos responsables de la adquisición de una LE y según las cuales se han planteado los distintos métodos y enfoques didácticos en busca del *método ideal*.

Estas propuestas teóricas se han modificado notablemente desde las primeras hipótesis hasta llegar a una explicación de los procesos adquisicionales que ponen en su centro el aprendiz tanto desde una perspectiva individual como social. Esto ha repercutido en la práctica didáctica diaria en función de cada uno de los enfoques metodológicos aplicados a la enseñanza de las LE y en los materiales diseñados con fines pedagógicos.

En la siguiente sección se ha descrito el tratamiento que el componente léxico ha tenido en los métodos y enfoques según tres etapas en las que el foco de la enseñanza pasa de las estructuras lingüísticas, a una revaloración gradual del léxico y culmina en los aspectos cualitativos del conocimiento del vocabulario (Morante Vallejo, 2005).

La primera etapa llega hasta los años setenta del siglo pasado (Método Tradicional, Movimiento de Reforma, Método Directo, Métodos Estructurales, Método Audiolingual) cuando el léxico era concebido como elemento secundario, supeditado a la gramática y la sintaxis. Las palabras se aprendían de memoria mediante listas de frecuencia. Además, el aprendizaje lingüístico tenía lugar exclusivamente por medio de la creación de hábitos que se desarrollaban al transferir formas y significados de la LM a la LO (de ahí que se haya planteado la *Hipótesis del Análisis Contrastivo*).

La elaboración del concepto de *competencia comunicativa* (Hymes, 1972) caracteriza la segunda etapa (Método Nocio-funcional, Enfoque Comunicativo, Enfoque Natural) llevando a una nueva valoración del léxico. Aprender el vocabulario significa adquirir significados, no memorizar palabras, desde el momento en el que el objetivo de

la ASL es la comunicación. Esto se debe al nuevo interés que se otorga a la interacción y la sociedad, implicando que la acción docente y los materiales didácticos han de planificarse y crearse en virtud de las necesidades y expectativas comunicativas del alumnado. El potenciamiento y la integración de todas las destrezas lingüísticas son el objeto principal de las actividades pedagógicas, diseñadas mediante un *input* comprensible, significativo y auténtico. Cobra protagonismo el contexto de uso de las palabras y la frecuencia deja de ser el único criterio de selección de los contenidos.

A mediados de los años ochenta se inaugura una nueva etapa en la que se presenta una ulterior revaloración del léxico. Se aboga por un aprendizaje cualitativo del vocabulario y se determina la noción de *competencia léxica* (Canale y Swain, 1980; Marconi, 1999; Jiménez Catalán, 2002). Hacen hincapié en el panorama de la enseñanza de las LE propuestas metodológicas de corte lexicalista (Enfoque por tareas, Enfoque Léxico) que ponen el léxico en el centro de la didáctica y del proceso adquisitivo ya que comunicar implica transmitir un significado: se promueve la primacía del vocabulario.

En lo referente a las tendencias actuales, desde la década final del siglo XX y de la publicación del MCER (2001) podemos hablar de una “era post-método” que coincide con una prolongación de la etapa antecedente. No se aplican en el aula metodologías delimitadas, sino que la enseñanza está caracterizada por un eclecticismo basado en las propuestas anteriores (Enfoque Ecléctico, Enfoque Accional, Enfoque Competencial) porque se comparte la idea que un solo *método ideal* no es la solución más eficaz a la pregunta *¿cómo se enseña una LE?*

A la luz de estos hallazgos, esto es, que el léxico ha pasado de ser un componente marginal a uno nuclear, se ha abierto un nuevo interrogante *¿qué léxico enseñar?* debido a la necesidad de organizar los contenidos a través de un nuevo proceso de selección, que sea más objetivo. Hasta donde conocemos, los criterios de selección del vocabulario se dividen entre (i) básicos (frecuencia, rango, productividad, rentabilidad); (ii) ligados al contexto de aprendizaje (las necesidades lingüísticas y los intereses del alumnado, la regularidad y la facilidad de aprendizaje, el componente cultural, el tipo de curso); (iii) las pautas del MCER y del PCIC; (iv) la disponibilidad léxica.

Hemos dedicado el Capítulo II a los orígenes y el desarrollo actual de los estudios de disponibilidad y de riqueza léxica, partiendo de sus antecedentes hasta llegar a las propuestas más recientes, finalizadas al perfeccionamiento de la metodología empírica y

a la profundización de aspectos descuidados en las primeras investigaciones.

Ambas metodologías tienen una doble posibilidad de aprovechamiento investigativo, sea en el campo didáctico-empírico sea en el teórico, porque, además de ser útiles para la valoración de la competencia léxica, permiten otras aplicaciones en varios ámbitos. Efectivamente, ni la disponibilidad ni la riqueza léxica están limitadas a los estudios de lingüística y didáctica, sino que pueden ser aprovechadas por otras disciplinas como la sociolingüística, la dialectología, la psicolingüística.

En primera instancia, la obtención de los índices de disponibilidad léxica tiene como finalidad la recogida y la medición del léxico disponible de una comunidad de habla o de un grupo de personas, es decir de aquel conjunto de palabras que los hablantes tienen en su lexicón mental, cuyo uso está sometido a una temática concreta: el objetivo es descubrir cuántas y cuáles unidades léxicas es capaz de usar un sujeto en determinados temas de comunicación.

En segunda instancia, cabe recordar que los diccionarios de frecuencia se utilizaban como único recurso lexicográfico, hasta mediados del siglo XX, para la selección del léxico en la elaboración de materiales didácticos, orientados hacia el desarrollo de un conocimiento “natural” de la LO, próximo al de un nativo. Pero, pronto, la literatura ha destacado que el mero criterio de frecuencia no era proficuo para el aprendiz extranjero (Gougenheim *et al.*, 1964: 31-35) porque el recuento del léxico en función de la mayor o menor aparición de las palabras era un método que, en realidad, no se ha revelado eficaz: algunas de las lexías que presentaban índices de frecuencia bajos o muy bajos (que, por tanto, no habían sido incluidas en los diccionarios) eran, al contrario, comúnmente usadas por los nativos. Basarse solo en este criterio no resultaba tan provechoso como para ofrecer al discente el caudal y el tipo de lexías necesarias para el adecuado aprendizaje de un idioma. A partir de los años cincuenta se ha realizado una revisión metodológica que ha puesto los cimientos de los análisis de la disponibilidad léxica gracias a la preocupación surgida por una selección léxica adecuada al establecimiento del vocabulario que debía aprenderse de una LE.

El estudio pionero, iniciado para mejorar la enseñanza del francés como lengua extranjera, ha desarrollado el concepto de disponibilidad léxica y el primer experimento basado en ella llevó a la publicación del *Français Fundamental* en 1964. Se diseñó una prueba de tipo asociativo en la que se sometieron los participantes a estímulos explícitos

que les permitiesen arrojar la mayor cantidad posible de palabras vinculada a dichos estímulos.

La primera prueba se llevó a cabo con un grupo de escolares francófonos, elegidos porque en aquella edad los niños ya tienen un vocabulario bien desarrollado, pero libre de todo tipo de especialización sectorial (Gougenheim *et al.*, 1964). El test se realizó en cuadernillos en los que se encontraba un cuestionario sociológico basado en preguntas personales que servían para analizar la incidencia de algunas variables sociales en los conocimientos lingüísticos; el siguiente apartado explicaba las instrucciones de la prueba; en las páginas a continuación aparecían las fichas por rellenar. Tras la recogida de los datos, el comité científico llevó a cabo un proceso de edición de las respuestas que permitió llegar a la etapa final: la confección del *Français Fondamental* (1964). Se inauguraba así una línea de investigación, muy fructífera, que ha introducido un notable cambio en el ámbito de la lingüística y la didáctica de LE.

A este propósito, destacamos la formulación de los dieciséis estímulos temáticos, los *centros de interés*, que eran, y siguen siendo, los elementos nucleares de este tipo de encuesta. Dichos centros se han planteado en relación con temas que el comité consideraba universales, pero su universalidad se ha visto pronto cuestionada. Muchos investigadores los han adaptado o sustituido según sus propósitos, evidentemente, se trata del punto de partida para el diseño de la prueba que permite extraer el léxico disponible del grupo encuestado.

Años más tarde, la disponibilidad léxica llega al mundo hispánico gracias a López Morales que ha impulsado la creación de numerosos proyectos en los campos de la dialectología, la psicolingüística, la etnolingüística, la sociolingüística y la didáctica que culminan en el *Proyecto Panhispánico de disponibilidad léxica*, encaminado a la recopilación de un diccionario global del léxico disponible de los hablantes nativos de español (para el cual se ha fijado un protocolo estándar que garantizase la repetitividad del experimento y su comparabilidad con otros).

La lengua española es la que más ha contribuido al desarrollo de esta metodología en el ámbito de los estudios sobre la LM y la LE. En efecto, tras realizarse encuestas en casi todas España e Hispanoamérica, la disponibilidad léxica se ha aplicado al campo de ELE, cuyo exponente pionero ha sido Carcedo (1998). Este autor junto a Samper Hernández (2002) ha contribuido a modificar las directrices metodológicas de los

experimentos (las variables sociolingüísticas, la edición de los datos y el tratamiento de los errores) conforme a los nuevos informantes. A raíz de estas contribuciones, muchos investigadores en distintas partes del mundo han elaborado estudios con aprendices extranjeros de español.

Por su parte, a partir de la segunda mitad del siglo XX, la medición de la riqueza léxica permite obtener informaciones sobre la habilidad en utilizar las palabras, esto es, sobre la capacidad expresiva de un grupo de sujetos. Esta metodología analiza textos orales o escritos producidos por la muestra en función de índices que miden la variedad y la densidad del vocabulario utilizado, proporcionando un estudio ulterior de la competencia léxica. Por eso, además del léxico disponible, es útil examinar el léxico activo que el aprendiz utiliza en su educto en la LO.

Guiraud (1954) basándose en la distinción entre *palabra* y *vocablo*, realizó la investigación pionera sobre la riqueza de un corpus de textos literarios, escritos por francófonos nativos, con el objeto de sentar las bases de un nuevo protocolo analítico. Para ello, desarrolló las primeras fórmulas, aun vigentes, a partir del total de palabras del texto y de su longitud.

De nuevo, López Morales (1984) importa esta metodología en el panorama hispano, proponiendo importantes modificaciones al protocolo empírico: nuevos procedimientos, operaciones e índices que prevén los cómputos de (i) el total de palabras y vocablos de los que consta el texto; (ii) las palabras con y sin contenido semántico.

A raíz de este trabajo se han realizado muchos estudios sobre el español desde múltiples perspectivas (en particular, de la sociolingüística) y se han desarrollado nuevas medidas que concurren a definir si un texto es rico: la densidad léxica, la variación léxica, la sofisticación léxica, la originalidad léxica, el número de errores. Sin embargo, en el ámbito del español LE no se ha promovido todavía un número suficientes de proyectos.

A estas alturas, resumimos las posibles explotaciones de las dos metodologías descritas en nuestro marco investigativo:

- (i) analizar el léxico disponible y activo de los aprendices;
- (ii) definir las etapas de aprendizaje;
- (iii) estudiar las interferencias de la LM y de otras LE conocidas;
- (iv) cotejar el vocabulario de los participantes con otros grupos de estudiantes y con los nativos;

- (v) establecer cuáles son las unidades poco rentables o los déficits léxicos;
- (vi) proporcionar comparaciones culturales;
- (vii) examinar la adecuación de los manuales y de otros materiales pedagógicos;
- (viii) eliminar el léxico superfluo de los recursos didácticos y realizar una selección de los contenidos apropiada;
- (ix) crear actividades, unidades didácticas y otros materiales finalizados a incrementar la competencia léxica de los discentes;
- (x) ampliar la cobertura de los análisis hacia la terminología (español para fines específicos).

En este sentido, la aplicación de las dos metodologías citadas permite comprender mejor la competencia léxica de los aprendices. La disponibilidad léxica se coloca en el plano paradigmático del léxico y analiza no solo de cuáles palabras dispone un aprendiz, sino cómo se organizan conceptualmente en su lexicón mental; el análisis longitudinal de la misma proporciona las claves necesarias para observar su evolución temporal y, por tanto, si la acción didáctica a la que se somete el grupo es efectiva. Por su parte, la riqueza léxica se sitúa en un plano sintagmático y examina cuántas palabras distintas es capaz un aprendiz de emplear en un texto y cómo lo hace. Por tanto, nuestro objetivo es aportar un estudio más preciso y coherente mediante la correlación de los datos.

Después de este bloque estrictamente teórico, el Capítulo III se centra en la metodología aplicada para la investigación, realizada a la luz de los resultados del test de medición de la competencia léxica que hemos suministrado a cien alumnos del *Dipartimento di Studi Linguistici e Letterari* de la Universidad de Padua, aprendices de ELE de nivel B1 y B2, organizando la recogida para poder comparar el conocimiento del estudiantado en cada uno de los dos niveles y en función de las variables contempladas.

Hemos descrito los pasos de edición del material y de su tratamiento informático y estadístico. A tal propósito, durante la fase de edición no hemos detectado muchos errores, sino que se han manifestado en mayor medida faltas ortográficas de acentuación que no perjudican el conocimiento del vocabulario de los informantes.

Aunque no cabe en nuestra intención proponer un fichero de errores colectivos – reto que dejamos para el futuro – podríamos hablar de errores de selección o de colocación falsa conforme al criterio lingüístico de Vázquez (1999), probablemente causados por una distracción ya que se han encontrado solo en algunas ocasiones mientras que en la

mayoría de los casos las palabras eran correctas.

Asimismo, hemos observado la presencia de errores intralingüales e interlingüales (extranjerismos, préstamos, falsos amigos y palabras inventadas) a causa de la interferencia de la lengua materna y de otras lenguas conocidas, en particular del inglés y del francés al ser las dos LE más estudiadas junto al español en este entorno académico. Según el criterio etiológico (Torijano, 2004) la organización del lexicón mental de un aprendiz extranjero favorecería la recuperación de material lingüístico de su idioma nativo o de cualquier otro como estrategia de aprendizaje (transferencia). Análogamente, la lingüística contrastivo-perceptiva considera que los procesos cognitivos del aprendiz sumados a su experiencia personal se reflejarían en el lexicón de la LO (Di Gesù, 2016). De todos modos, planteamos en una de las hipótesis de partida que la afinidad entre el italiano y el español habría llevado a una incidencia mayor de la LM en las actualizaciones de los participantes por la poca distancia léxica que caracteriza los lexemas (Calvi, 1995: 86-90), que en realidad no se ha manifestado en gran medida.

Al contrario, se ha corroborado otra hipótesis inicial: localizamos una cantidad apreciable de palabras del registro coloquial y de la variedad diatópica hispanoamericana. De la misma manera, en lo que atañe a las categorías morfosintácticas de las respuestas activadas, hemos comprobado la presencia abrumadora de sustantivos y de pocas unidades multipalabras, como se ha puesto de relieve en otras investigaciones ya publicadas.

El Capítulo IV inaugura la segunda parte de la Tesis dedicada al estudio empírico. Analizamos desde la perspectiva cuantitativa el vocabulario disponible de nuestros encuestados.

Los resultados generales de las pruebas asociativas muestran que el corpus de trabajo (cien informantes) contiene un total de 12.579 palabras y 1.490 vocablos, cifras que se distribuyen en una media por sujeto de 125,79 palabras y de 14,90 vocablos.

Los dieciséis centros de interés se pueden repartir en cuatro franjas, dentro de las cuales constatamos cuatro campos que superan las 1.000 unidades (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI10, “la ciudad”; CI14, “los animales”), manifestándose como los más productivos y un grupo que no alcanza las 400 unidades léxicas (CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”), quizás debido al poco interés suscitado por las

temáticas y la escasa, o nula, atención que les otorgan los programas curriculares y los materiales de ELE.

La variedad léxica, esto es, la presencia de vocablos, demuestra que no existe una relación directa con la productividad de palabras, contamos con una distribución distinta de los CI más ricos ya que predominan seis áreas que coinciden en parte con las que acabamos de exponer (CI05, “alimentos y bebidas”; CI16, “profesiones y oficios”, CI10, “la ciudad”; CI11, “el campo”; CI15, “juegos y distracciones”; CI14, “los animales”).

Los índices de cohesión y de densidad léxica destacan que algunos estímulos se presentan más cerrados (CI01, “partes del cuerpo”; CI03, “partes de la casa (sin muebles)”; CI04, “los muebles de la casa”; CI12, “medios de transporte”) y otros, al revés, son más abiertos y difusos (CI11, “el campo”; CI13, “trabajos del campo y del jardín”; CI16, “profesiones y oficios”). Podríamos afirmar que una mayor concreción implica un aprendizaje sistematizado a partir del cual los estudiantes no han ampliado sus conocimientos por encima de los propuestos en los cursos. En el caso opuesto, donde las respuestas son más heterogéneas, es posible que cada sujeto haya escrito ítems distintos de los compañeros ya que se trata de argumentos que han aprendido o profundizado por su cuenta.

Si consideramos los resultados en función de los condicionantes sociolingüísticos, la población investigada se distribuye como sigue: (i) variable “sexo”: 87 mujeres; 13 hombres; (ii) variable “nivel de ELE”: 50 aprendices de nivel B1; 50 aprendices de nivel B2; (iii) variable “conocimiento de otras LE”: 19 estudiantes conocen 2 LE; 81 estudiantes conocen más de 2 LE.

Los datos sugieren que el factor sexo no tiene una influencia significativa en la cantidad de actualizaciones del grupo, las mujeres han arrojado el 9,61% más de palabras con respecto al conjunto masculino, pero destaca la supremacía de los hombres en la producción de vocablos ya que sobrepasan las compañeras del 66,35%. El nivel de cohesión revela que las respuestas de las mujeres son más dispersas con respecto a los varones que aportan un conjunto de lexías más cerrado.

En lo referente al nivel de ELE, como esperábamos, constatamos un aumento de los resultados en casi todos los centros (si excluimos el CI15, “juegos y distracciones”, y el CI08, “la escuela: muebles y materiales”) del grupo de nivel B2, que produce el 7,51% más de palabras y el 14,05% de vocablos. Aunque se ha revelado un incremento del

bagaje léxico, este no es tan significativo para sugerir un desarrollo sobresaliente de los conocimientos del alumnado. Al examinar la concreción interna de cada variante de la variable, los estudiantes de nivel B1 alcanzan +8,47% en el índice de cohesión y un +4,25% en la densidad léxica, con lo cual su léxico resulta levemente más compacto.

El conocimiento de otras LE se ha demostrado influyente en la capacidad de activar más palabras para el grupo que conoce más de dos LE, pues ha arrojado una cantidad mayor de palabras del 10,54%. Aunque el número de vocablos favorece los alumnos que conocen dos ya que, en promedio, han escrito el 50,84% más. Además, consiguen un más alto grado de cohesión y, por tanto, tienen una mayor asociación conceptual con respecto al lexicón de los compañeros.

El análisis longitudinal ha manifestado que la competencia léxica ha mejorado, se ha contabilizado una mayor producción y riqueza en los estudiantes después de la asistencia al curso. Los datos procedentes de la segunda administración indican un aumento del 30,29% en el total de palabras en todos los estímulos temáticos (pasan de 6.517 a 8.491) y del 7,50% en el total de vocablos (pasan de 1.161 a 1.244). La acción docente ha sido eficaz, incluso porque los índices de cohesión y de densidad revelan un conjunto de palabras más cerrado al final del año académico (la cohesión ha crecido en media del 23,75%).

El grueso del siguiente apartado consiste en cotejar el léxico disponible de nuestros informantes con trabajos realizados por distintos investigadores con otros estudiantes de ELE con la finalidad de detectar similitudes y diferencias en el vocabulario de los grupos. En particular, el objetivo de este contraste era demostrar que la competencia léxica de nuestros informantes es mayor que la de los estudiantes de distinta procedencia debido a la afinidad lingüística entre italiano y español.

Esta hipótesis no ha sido confirmada, las cantidades de palabras y vocablos de los italianos se han revelado escasas. En lo que se refiere a la media de palabras, en orden decreciente se colocan Sánchez-Saus (2011) con 16,57; Hidalgo (2017b) con 16,01; Šifrar Kalan (2014b) con 15,73; Samper Hernández (2002) con 13,99; Carcedo (2000c) con 12,80; Del Barrio y Vann (2018) con 11,16; Nalesso (2019) con 9,94. Los promedios de vocablos revelan lo siguiente: Samper Hernández (2002) con 4,83; Carcedo (2000c) con 3,39; Šifrar Kalan (2014b) con 2,50; Sánchez-Saus (2011) con 2,48; Hidalgo (2017b) con 1,95; Del Barrio y Vann (2018) con 1,91; Nalesso (2019) con 1,11.

El índice de cohesión pone de relieve que las respuestas de nuestros informantes presentan, en media, un valor de 0,88 colocándose entre los que obtienen una concreción mayor (Samper Hernández, 2002; Carcedo, 2000c; Del Barrio y Vann, 2018) y los que han contabilizado cifras inferiores (Šifrar Kalan, 2014b; Sánchez-Saus, 2011; Hidalgo, 2017b).

Si contemplamos la naturaleza asociativa de los centros de interés, se verifica lo afirmado en Echeverría *et al.* (1987: 67) al defender que algunos CI consiguen un alto grado de cohesión semántica porque se trata de estímulos que posibilitan limitadas asociaciones mentales, mientras que un bajo índice conlleva una gran variedad temática que activa múltiples conexiones.

Si nos referimos al producto de la acción docente, lo que estos resultados sugieren es que los estudiantes que alcanzan un nivel mayor de cohesión han desarrollado un conocimiento léxico que es el producto de un aprendizaje bien organizado sobre un argumento que ha sido trabajado, mientras que los otros quizás no le han dedicado la misma atención y el índice es más abierto.

En suma, se han localizado carencias y lagunas en la competencia de nuestros encuestados posiblemente debidas a (i) el grado de atención que le otorgan los programas curriculares al componente léxico; (ii) el método didáctico empleado; (iii) el contexto de aprendizaje (país de origen *versus* país hispanohablante), con lo cual sería interesante tener en cuenta también el tipo de enseñanza como variable.

Todo esto confirmaría, una vez más, la importancia de plantear actividades explícitas para la adquisición del léxico, porque el aprendizaje incidental no parece suficiente (al menos en nuestra Universidad, el profesorado confirma su mayor concentración en la gramática y en la contrastividad entre italiano y español, en detrimento de otros elementos). Centrarse en el vocabulario de la LO posibilita la comprensión y la expresión de los discentes que, al contrario, sería difícil sin una sólida base léxica. Puede que en Italia se descuide la enseñanza de este componente quizá por la sensación de facilidad que causa la poca distancia tipológica entre LM y LE y la consiguiente afinidad léxica. Esto justificaría el inferior rendimiento de los informantes.

Si dejamos de un lado los valores en sí mismos, todas las agrupaciones se han comportado de manera semejante: aportan más palabras en el CI05, “alimentos y bebidas”, y menos en el CI12, “medios de transporte” y el CI02, “la ropa”. El mayor

número de vocablos se registra en el CI05, “alimentos y bebidas”; el CI10, “la ciudad”; el CI16, “profesiones y oficios”; el CI15, “juegos y distracciones”, y el menor en el CI01, “partes del cuerpo”; el CI02, “la ropa”; el CI12, “medios de transporte”.

El objetivo del Capítulo V ha sido el estudio cualitativo del corpus a partir de los índices de disponibilidad léxica extraídos de *Dispogen* con dos objetivos: averiguar las temáticas predominantes y las categorías gramaticales más difundidas en las respuestas.

La oscilación de vocablos más disponibles (con $ID \geq 0,1$) va de un mínimo de tres a un máximo de veinticinco ítems por CI. Se califican como más productivos: el CI05, “alimentos y bebidas”; el CI01, “partes del cuerpo”; el CI10, “la ciudad”; los que tienen el número menor coinciden con aquellos que se han manifestado como los menos rentables en el análisis cuantitativo: CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”. Al analizar los datos por variable, sobresale la supremacía de las mujeres sobre los hombres, de los estudiantes de nivel B2 sobre los B1 y, por último, de los que conocen más de dos LE sobre los que solo saben o están estudiando una.

Ha surgido, además, que los sustantivos son la clase preponderante en toda la muestra, los verbos aparecen en porcentajes escasos, los adjetivos se presentan aun menos y no hemos localizado ninguna otra categoría morfológica.

En lo que se refiere a las asociaciones mentales activadas, los vocablos se encajan perfectamente en los ámbitos prototípicos de cada estímulo. Solo en posiciones más bajas o en centros poco rentables (CI04, “los muebles de la casa”; CI11, “el campo”; CI12, “medios de transporte”; CI13, “trabajos del campo y del jardín”) los informantes han escrito lexías referentes a asociaciones secundarias.

A raíz de estos datos, la cardinalidad del conjunto nos ha permitido observar que en ciertos ámbitos se ha alcanzado el 100% (en el CI12, “los medios de transporte”, según el factor sexo; en CI08, “la escuela: muebles y materiales”; CI09, “iluminación y calefacción”; CI12, “medios de transporte”, según el nivel de ELE; en CI06, “objetos colocados en la mesa para la comida”; CI09, “iluminación y calefacción”, según el conocimiento de otras LE) y en otros, al contrario, ha bajado hasta el 0%, en el caso de que no había ninguna coincidencia entre las respuestas (0% en el CI09, “iluminación y calefacción”, según el género; 30% en el CI07, “la cocina y sus utensilios”, en la comparación entre niveles; 37,50% en el CI07, “la cocina y sus utensilios”, en función de las LE conocidas).

Igualmente, en el análisis longitudinal los porcentajes de compatibilidad varían del 30% en el CI09, “iluminación y calefacción”, pasando por valores elevados, hasta llegar al 100% en el CI16, “profesiones y oficios”.

Al igual que en los otros apartados, los sustantivos son la clase dominante ya que solo hemos detectado verbos en: el CI08, “la escuela: muebles y materiales” (10%); el CI13, “trabajos del campo y del jardín” (55%); el CI16, “profesiones y oficios” (31%).

El estudio longitudinal de los datos confirma un desarrollo de la competencia léxica: en general, los encuestados han aumentado su bagaje léxico (más disponible), en la segunda prueba tienen a disposición una cantidad mayor de vocablos con $ID \geq 0,1$.

En la segunda parte del capítulo la comparación entre el total de vocablos recopilados en nuestro corpus con el CREA ha demostrado que un cierto porcentaje de las respuestas pertenece al léxico más frecuente del español (el 20% de estas unidades coincide con aquellas de la muestra que presentan un $ID \geq 0,1$). Asimismo, el contraste con el CAES ha revelado una correspondencia con sus entradas (las palabras más frecuentes de los aprendices de ELE) y el vocabulario disponible de nuestros informantes.

Hay que tener en consideración que ambos corpus de referencia contienen también unidades funcionales del discurso, declinadas, conjugadas, y nombres propios, lo que implica que los vocablos coincidentes habrían aumentado en gran medida si no hubiésemos editado el material recogido.

Los resultados de los análisis del léxico disponible corroboran una de las hipótesis de partida, y lo dicho ya por otros autores en estudios sobre la disponibilidad léxica en aprendices de ELE, que los centros de interés producen cantidades mayores de palabras y de vocablos en función de la rentabilidad que tienen para el alumnado, así pues, algunos de estos estímulos se revelan poco adecuados para cumplir con las necesidades y los intereses culturales y comunicativos del siglo XXI propios de una sociedad digitalizada y fuertemente intercultural. Es bien sabido que existen centros poco productivos, como el CI09, “iluminación y calefacción”; el CI11, “el campo”, y el CI13, “trabajos del campo y del jardín”. Consideramos que esto podría derivar del escaso interés que los temas suscitan en los discentes o de la poca atención que se le dedica debido a la realidad socioambiental en la que viven los informantes.

Sánchez-Saus (2011) aboga por utilizar la lista de contenidos que el MCER recomienda para el nivel A1 (Robles Ávila, 2006) en la selección de los centros de interés

y Santos Díaz (2017) sugiere que la nómina debería basarse en las nociones específicas del PCIC, desde el momento en el que se trata de los patrones que guían la elaboración de materiales y la evaluación para los Diplomas de ELE del Instituto Cervantes.

Esto justificaría la necesidad de tomar nuevas decisiones que solventen la inadecuación de algunos de los centros de interés, establecidos como *input* para la elicitación hace más de setenta años: se trataría de eliminar los pocos productivos, ampliar o unir otros e incluir nuevos que impulsen la actualización de otras categorías morfosintácticas y de unidades complejas porque, además, otra de las cuestiones ya denunciadas es la limitación de los datos recogidos a la clase de los sustantivos y la unidad palabra, entendida como unidad monoverbal. Asimismo, esto respaldaría lo criticado ya en un estadio más temprano en este ámbito investigativo, el riesgo de elegir con demasiada subjetividad los CI por parte de los investigadores (Juilland, 1970; Mackey, 1971a; Benítez Pérez, 1994; López Morales, 1999; Bombarelli, 2005; Sánchez-Saus, 2011; Moreno Fernández, 2012; González Fernández, 2014; Paredes García, 2014; Tomé Cornejo, 2015).

Por eso, se echa en falta la utilidad de corregir y perfeccionar tanto las encuestas como los protocolos de ejecución de los experimentos mediante una implementación que permita solucionar los puntos críticos, como se ha hecho en varias reuniones científicas sobre el asunto (Bilbao, en 1999; San Millán de la Cogolla, en 2003; Salamanca, en 2011 y en 2016) cuyo objetivo era llegar a acuerdos metodológicos y establecer un *modus operandi* común.

El restante bloque está constituido por el estudio de la riqueza léxica.

En el Capítulo VI hemos llevado a cabo el estudio cuantitativo del corpus. Los resultados se han discutido en tres secciones que han abordado los análisis (i) transversal; (ii) longitudinal; (iii) comparativo.

El análisis general de la variedad léxica muestra que todos los relatos alcanzan un buen nivel de riqueza, en cada uno se detecta más del 50% de palabras diferentes (el promedio del TTR es 0,63 y 63% de variación). El índice de hápax también es bueno ya que presenta valores medio-bajos (en promedio 1,39) que se aproximan al ideal. Como planteaba una de nuestras hipótesis, considerado el nivel de ELE, los indicadores revelan una apreciable variación y el desarrollo de una adecuada competencia léxica que conlleva un buen manejo de la lengua española.

La densidad léxica, por su parte, revela resultados divergentes: el 90% de los textos se compone de una cantidad satisfactoria de *tokens* léxicos (la media es de 55%), pero los cálculos del IAT parten de un mínimo de 1,52 hasta llegar a un máximo de 2,33. El promedio es de 1,84 lo que supone la aparición de una palabra semántica cada dos gramaticales.

Es oportuno señalar que, en absoluto, el intervalo de aparición de palabras nocionales se ha revelado bastante alto en todo tipo de análisis, con lo cual no vamos a discutirlo más.

Esto corrobora que un texto con una abundante variedad léxica no necesariamente presenta el mismo grado de densidad, por eso se aboga por la aplicación de diferentes medidas finalizadas a obtener datos lo más completos posible (Read, 2000; Johansson, 2008; Jarvis, 2013).

A la hora de repartir los datos según la variable, parece que el sexo no influye en la determinación de la riqueza léxica en ninguno de los índices estudiados. Los relatos de ambas variantes igualan o sobrepasan el 50% de variedad léxica y el 85% de cada grupo presenta un buen índice de hápax. La densidad léxica marca una leve alteración de esta correspondencia, pero no de manera significativa, poniendo manifiesto un ligero influjo del factor ya que en toda la producción de los hombres el índice cubre o supera la mitad del texto mientras que el mismo resultado se detecta en el 89% de los escritos de la sección femenina.

La riqueza léxica contabilizada en función del nivel de español no supone una gran influencia de este condicionante, los valores son semejantes. El desnivel entre los relatos no es representativo estadísticamente: todos los textos igualan o sobrepasan el 50% de variación y el 0,50 de TTR. El índice de hápax ha manifestado una pequeña diferencia: el 80% de los B1 y el 90% de los B2 proporcionan valores que suben hasta 1,50 pero tampoco esta disparidad es discriminante. Lo mismo revela la densidad léxica ya que el 92% de los estudiantes de nivel intermedio tienen un índice equivalente o superior a 50% frente al 88% de los discentes avanzados. De ahí que no destaque la predominancia de ningún grupo: si bien el dominio lingüístico aumenta, los índices no ascienden ni bajan significativamente, parece casi que el desarrollo de la competencia léxica de un nivel al otro se ha ralentizado.

El conocimiento de otras LE se ha demostrado como la variable que tiene más

influencia con respecto a los otras dos. La variedad léxica ha revelado que los informantes que conocen más de dos LE obtienen resultados mejores del 9%. De la misma forma, estos aprendientes alcanzan el mejor índice de hápax (en media 1,07) con un aumento del 16%. Al parecer, se corrobora la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979), según la cual un individuo que ya conoce una LE emplea sus competencias como elemento facilitador durante el aprendizaje de otra. Por eso, nos ha extrañado un cambio de tenencia al estudiar la densidad léxica porque se han computado índices mejores de los informantes que saben dos.

El análisis longitudinal da cuenta de la evolución del conocimiento de los aprendices tras la asistencia a los cursos de español de la Universidad. Junto al estudio de la variable nivel de ELE contribuye a verificar o rechazar que el dominio lingüístico surte algún efecto en el grupo ya que es esperable un aumento de la competencia léxica. Contradiendo nuestra hipótesis inicial, llama la atención el índice de la diversidad desde el momento en el que los alumnos han conseguido un rendimiento notable a principio del año académico: los datos de TTR y variación son más altos y, por tanto, mejores. En cambio, el índice de hápax y densidad se asemejan en ambas pruebas. Hemos contabilizado un buen uso de las palabras semánticas (más del 50%) en el 88% de la primera muestra y el 96% de la segunda. Sin embargo, los resultados difieren de poco entre los dos momentos de la administración de la encuesta. Hecho este que sorprende porque parece demostrar que no se ha desarrollado el conocimiento de los participantes como se había hipotizado: es más, casi se ha producido un bloqueo o incluso un bajón.

En cuanto a la correlación de los resultados de diferentes investigaciones se ha validado que aprendices no nativos de distinta procedencia que comparten el nivel de español producen en sus textos una riqueza semejante.

El grado de variedad léxica es homogéneo. En orden decreciente se ha detectado el siguiente TTR: 0,70 en Berton (2014); 0,64 en Cuba Vega y Cuba (2004); 0,63 en Nalesso (2019); 0,60 en Wang (2016); 0,60 en Basso (2017).

Igualmente, el IAT pone de manifiesto que la densidad tampoco consigue resultados diferentes, si excluimos el trabajo de Basso (2017) que llega a 2,35: Wang (2016) contabiliza un promedio de 1,80; Cuba Vega y Cuba (2004) presentan 1,81; Nalesso (2019) consigue una media de 1,84.

Por último, hemos planteado el Capítulo VII como acercamiento a un análisis de

tipo cualitativo de la riqueza léxica por lo que las conclusiones no son definitivas. Se trata de un primer experimento de explotación del material desde un nuevo punto de vista dirigido a promover investigaciones más profundas.

Los resultados más destacables atañen a la alternancia y capacidad de uso de las palabras nocionales y funcionales: se ha procurado demostrar que, en su mayoría, los vocablos más frecuentes del corpus son unidades semánticas, no obstante en los primeros rangos del listado se detectan muchas unidades funcionales del discurso compartidas en los textos.

Merece la pena subrayar la importancia de seguir planteando proyectos de esta naturaleza con el objetivo de fijar pautas comunes para la realización de nuevos análisis que otorguen la elaboración de comparaciones más detalladas. La evaluación de la riqueza léxica ofrece posibilidades interesantes de estudio de la competencia léxica, y comunicativa, como lo ha demostrado esta investigación con sus carencias y limitaciones.

En definitiva, lo que sugieren los datos de la disponibilidad léxica indica que los resultados de nuestros informantes son buenos, no sobresalientes como se esperaba debido a la afinidad que une el italiano al español, en particular si tenemos en cuenta su escaso rendimiento en el contraste con otros aprendientes. De todos modos, hemos podido averiguar un aumento del léxico disponible de un nivel a otro, aunque no mucho, y entre las dos administraciones del test.

La medición de la riqueza léxica, por su parte, prueba que no necesariamente la capacidad expresiva se acompaña con el avance del nivel de español. En efecto, en términos generales, los resultados han sido apreciables considerando el grado de dominio lingüístico (incluso en el análisis comparativo), pero no se ha producido el mismo desarrollo de la competencia entre los dos niveles encuestados y tampoco en el análisis longitudinal. Esto contradice, en parte, la idea de que existe una relación asociativa entre el nivel de ELE y la cantidad de lexías activadas y utilizadas en las dos partes de la prueba a la que se ha sometido el alumnado.

A nuestro entender, los hallazgos descritos nos permiten señalar las siguientes conclusiones: aunque resulta del análisis de la disponibilidad que el conocimiento del léxico disponible aumenta, no sucede lo mismo en lo que se refiere al léxico activo (al discurso) estudiado mediante el análisis de la riqueza léxica. Esto demostraría la utilidad de combinar los estudios sobre léxico disponible y léxico activo, cuya correlación crítica

posibilita una detallada descripción de la competencia léxica.

Las causas de esto pueden ser múltiples, tantas como los agentes que condicionan el proceso de aprendizaje. Por encima de los factores emotivos planteados a partir del *filtro afectivo* de Dulay y Burt (1977) y del *monitor* de Krashen (1982), imputamos la responsabilidad primaria al impacto del parentesco lingüístico del italiano con el español que lleva al descuido de la enseñanza del vocabulario, quizá por la sensación de facilidad causada por la poca distancia tipológica que, en realidad, provoca interferencias e intraferencias (Torijano, 2004). La afinidad es causa de conflictos y dificultades para el alcance de la competencia lingüística. De hecho, se podría considerar como rasgo del proceso de aprendizaje de lenguas cognadas la elevada facilidad inicial para la comprensión, obstaculizada por una tendencia hacia fenómenos de contaminación y una propensión hacia el distanciamiento de los discentes (Calvi, 1995; Bailini, 2016).

Sin embargo, los capítulos analíticos han corroborado o refutado las hipótesis planteadas a principios de la investigación, como ya hemos señalado a lo largo de esta sección conclusiva. Consideramos conseguido el objetivo de la Tesis y esperamos haber contribuido a poner los cimientos de un nuevo sistema capaz de mejorar la ejecución de los estudios sobre la competencia léxica de los aprendices de ELE, basados en la correlación de los datos procedentes de los análisis de disponibilidad léxica (parámetros de tipo paradigmático) y de riqueza léxica (parámetros de tipo sintagmático).

En efecto, una cuestión a la que todavía no se ha encontrado una respuesta unívoca es *¿cómo se puede valorar la competencia léxica?* (Morante Vallejo, 2005: 50-53). Sin duda, (i) los análisis de disponibilidad resultan indispensables para diagnosticar tanto la cantidad y la calidad de vocabulario que conocen los informantes, como el grado de asociación cognitiva de lo aprendido; (ii) los análisis de riqueza permiten examinar sus eductos y averiguar si saben una palabra y cómo la utilizan, esto es, si del conocimiento pasivo-receptivo ha entrado a formar parte del activo-productivo.

Como colofón, sabemos que ninguna metodología está exenta de críticas y que cada investigación surge de la anterior para mejorarla, de ahí que para finalizar reiteramos que nuestro principal objetivo no perseguía aportar respuestas prescriptivas a los interrogantes marcados, sino acercarnos a explicar el complejo fenómeno de la competencia léxica mediante un aporte original a los estudios de este tipo.

Por esta razón, como ya se ha sugerido en varias ocasiones, entre las que

recordamos la reciente conferencia inaugural de Sánchez-Saus de las *II Jornadas de investigación en Disponibilidad léxica* que han tenido lugar en Lleida en 2019, sería útil promover la cooperación entre autores cuyo trabajo se relacione con el estudio de la competencia léxica en ELE ya que persisten problemas vinculados con la comunicación y distribución de los resultados. Se trata de información útil para la comunidad investigadora, pero desaprovechada. Con el apoyo de las humanidades digitales y de la lingüística de corpus, deberíamos apostar por la creación de una base de datos compartida y constantemente actualizada que fomente el manejo y la comparación del material, la presentación de innovaciones y nuevas propuestas, abriendo el corpus a diferentes aplicaciones y abarcando también los análisis de riqueza léxica, que sentarían las bases para nuevas investigaciones que aun no han sido abordadas. Se crearía un espacio para la reflexión y el debate entre los expertos en la materia y otros investigadores que pueden exponer sus trabajos o proyectos más recientes.

Tras todo ello, tenemos el propósito de seguir el camino abierto por esta Tesis, ya que las posibilidades que brinda esta línea de trabajo son abundantes y prometedoras: a partir de este importante conjunto de datos es posible (i) examinar la interlengua de los aprendices; (ii) elaborar un corpus cacográfico (Trigo Ibáñez, Romero Oliva y Santos Díaz, 2018); (iii) realizar otros análisis siguiendo a los numerosos proyectos dedicados a la contrastividad entre italiano y español.

A este propósito, la investigación en el ámbito de la lingüística contrastiva ha sido y sigue siendo muy fecunda. Emblemático en este marco es el proyecto *Contrastiva – Portal de lingüística contrastiva español-italiano* dirigido por San Vicente fruto de la labor del Grupo de investigación LICEI (Lingüística contrastiva español-italiano). En este sentido, en atención a los consolidados estudios cognitivos sobre la revalorización de la lengua materna en la ASL, se podría concluir defendiendo la necesidad de cotejar los datos de disponibilidad y riqueza léxica, además del español, de la lengua nativa de los aprendices y de sus otras segundas lenguas, que permitan fijar un parámetro de comparación dentro de la misma muestra. Por todo ello, junto a Castillo hemos planteado un proyecto cuya acción prioritaria es la implementación del test de medición de la competencia léxica en ELE con otras pruebas realizadas en la LM y en otros idiomas conocidos, del cual se espera, como resultado principal, el diseño de un modelo de encuesta que permita el análisis interlingüístico. No se trata del primer proyecto que

aborda más idiomas simultáneamente ya que disponemos de los trabajos de Hernández Muñoz (2014; 2018), Šifrar Kalan (2016) y Hernández Muñoz y Tomé Cornejo (2017), quienes se han focalizado en los puntos de encuentro entre primera y segunda lengua, pero sí sería el primero realizado con informantes italófonos.

Todo esto, se supone, dará cuenta, describirá y analizará críticamente la correlación entre los datos cuantitativos y cualitativos de la disponibilidad y de la riqueza, que confluyen en el estudio de la competencia léxica, desde una nueva perspectiva.

REFERENCIAS BIBLIOGRÁFICAS

- Aabidi, L. (2019). La disponibilidad léxica en español como lengua extranjera: dos décadas de investigación científica. *MarcoELE*, 28. Recuperado de https://marcoele.com/descargas/28/aabidi-disponibilidad_lexica.pdf
- Agustin Llach, M. D. P. (2013). Dinámica del aprendizaje incidental de léxico en lengua extranjera. *Revista Nebrija de Lingüística Aplicada*, 14. Recuperado de https://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_526a6afcdce77.pdf
- Ahumada Lara, I. (2006). *El léxico disponible de los estudiantes preuniversitarios de la provincia de Jaén*. Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Aitchison, J. (1994). *Words in the mind*. London: Blackwell.
- Alba, O. (1995). *Léxico disponible de la República Dominicana*. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.
- Almela, R. et al. (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. Madrid: Editorial Universitas.
- Alonso Ramos, M. (dir.). (2004). *Diccionario de colocaciones del español (DICE)*. Universidad de La Coruña. URL <http://www.dicesp.com>
- Alvar Ezquerro, M. (dir.). (1995). *Diccionario para la enseñanza de la lengua española (español para extranjeros)*. Barcelona: Vox-Bibliograph/Universidad de Alcalá.
- Alvar Ezquerro, M. (2003). *La enseñanza del léxico y el uso del diccionario*. Madrid: Arco/Libros, S. L.
- Ambadiang, T. y García Parejo, I. (2006). La cultura lingüística y el componente cultural en la enseñanza de lenguas no maternas: observaciones sobre algunos paradigmas de la competencia cultural. *Didáctica. Lengua y literatura*, 18, pp. 61-92. Recuperado de <https://revistas.ucm.es/index.php/DIDA/article/view/DIDA0606110061A/19152>
- Andrés Pérez, B. (1997). *Riqueza léxica en textos escritos de tres niveles de EGB* (Memoria de licenciatura, Universidad de Alcalá).
- Anthony, L. (2018) [programa informático]. *AntConc (Versión 3.5.2)*. Tokyo: Waseda University. URL <http://www.laurenceanthony.net/software/antconc/>

- Arbulu Barturen, M^a. B. (1998). *Contribución al estudio del sistema vocálico de hablantes extranjeros que aprenden español* (Tesis doctoral, Universidad del País Vasco - Euskal Herriko Unibertsitatea).
- Arbulu Barturen, M^a. B. (2000). La variación estilística en el análisis de la interlengua. En M^a. A. Martín Zorraquino y C. Díez Pelegrín (eds.), *¿Qué español enseñar?: norma y variación lingüísticas en la enseñanza del español a extranjeros. Actas del XI Congreso Internacional ASELE, Zaragoza 13-16 de septiembre de 2000* (pp. 155-164). Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=608250>
- Arnaud, P. J. L. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. En T. Culhane, C. Klein Bradley y D. K. Stevenson (eds.), *Practice and problems in language testing: papers from the International Symposium on Language Testing* (pp. 14-28). Colchester: University of Essex.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Ávila, R. (1986). Léxico infantil de México: palabras, tipos, vocablos. En J. D. Moreno de Alba (ed.), *Actas del Congreso del II Congreso Internacional sobre el español de América* (pp. 510-517). México: Universidad Nacional Autónoma de México.
- Ávila, R. (octubre, 2001). *Los medios de comunicación masiva y el español internacional*. Comunicación presentada en el II Congreso Internacional de la lengua española, Valladolid 16-19 de octubre de 2001. Recuperado de http://congresosdelalengua.es/valladolid/ponencias/unidad_diversidad_del_espanol/1_la_norma_hispanica/avila_r.htm
- Ávila Martín, M. C. (2010). Estadística y lingüística de corpus: implicaciones pedagógicas en la enseñanza y el aprendizaje del léxico. *Cauce. Revista internacional de Filología, Comunicación y sus Didácticas*, 32-33, pp. 163-175. Recuperado de https://cvc.cervantes.es/literatura/cauce/pdf/cauce32-33/cauce_32-33_014.pdf
- Ávila Muñoz, A. M. (2006). *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.
- Ávila Muñoz, A. M. (2016). Can speakers' virtual lexical richness be calculated? Individual and social determining factors. *Spanish in Context*, 13 (2), pp. 285-307.
- Ávila Muñoz, A. M. (2017). The available lexicon: A tool for selecting appropriate vocabulary to teach a foreign language. *Iranian Journal of Language Teaching Research*, 5 (1), pp. 71-91. Recuperado de <http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/5.pdf>
- Ávila Muñoz, A. M. y Sánchez Sáez, J. M^a. (2010). La disponibilidad léxica. Antecedentes y fundamentos. En A. M. Ávila Muñoz y J. A. Villena Ponsoda (eds.), *Variación social del léxico disponible en la ciudad de Málaga* (pp. 35-81). Málaga: Editorial Sarriá.

- Ávila Muñoz, A. M. y Sánchez Sáez, J. M^a. (2011). La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga. *ELUA. Estudios de Lingüística Universidad de Alicante*, 25, pp. 45-74. Recuperado de https://rua.ua.es/dspace/bitstream/10045/21640/1/ELUA_25_02.pdf
- Bailini, S. L. (2016). *La interlengua de lenguas afines. El español de los italianos, el italiano de los españoles*. Milano: LED.
- Bakonyi, H. (1933). *Die gebräuchlichsten Wörter der deutschen Sprache, für den Fremdsprachenunterricht stufenmässig zusammengestellt*. Beiträge zur Methodik des deutschen Sprachunterrichts im Ausland. Goethe-Institut zur Fortbildung ausländischer Deutschlehrer, Deutsche Akademie, Heft 1. München: Reinhardt.
- Baralo, M. (2001). El lexicón no nativo y las reglas de gramática. En S. Pastor y V. Salazar (eds.), *Estudios de Lingüística de la Universidad de Alicante. Tendencias y líneas de investigación en adquisición de segundas lenguas* (pp. 12-40). Alicante: Universidad de Alicante.
- Baralo, M. (2011 [1999]). *La adquisición del español como lengua extranjera*. Madrid: Arco/Libros, S. L.
- Baralo, M. y Estaire, S. (2010). Tendencias metodológicas postcomunicativas. En C. Abelló, C. Ehlers y L. Quintana (eds.), *Escenarios bilingües: el contacto de lenguas en el individuo y la sociedad* (pp. 105-128). Bern: Peter Lang. Recuperado de <http://www.nebrija.com/revista-linguistica/files/articulosPDF/11.BaraloEtaire.pdf>
- Barbasán Ortuño, I. P. (2016). *El error léxico en la interlengua de lenguas afines: italiano y español* (Tesis doctoral, Universitat Politècnica de València). Recuperado de <https://riunet.upv.es/bitstream/handle/10251/63145/-Barbasán%20-%20El%20error%20léxico%20en%20la%20interlengua%20de%20lenguas%20afines%3A%20italiano%20y%20español.pdf?sequence=1>
- Bartol Hernández, J. A. (2001). Reflexiones sobre la disponibilidad léxica. En J. A. Bartol Hernández, S. Crespo, C. Fernández, C. Pensado, E. Prieto y N. Sánchez (eds.), *Nuevas aportaciones al estudio de la Lengua Española. Investigaciones filológicas* (pp. 221-236). Salamanca: Luso-Española de Ediciones.
- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. Castañer Martín y V. Lagüéns Gracia (eds.), *De moneda nunca usada. Estudios filológicos dedicados a José M^a. Enguita Utrilla* (pp. 85-107). Zaragoza: Instituto Fernando el Católico. Recuperado de <http://ifc.dpz.es/recursos/publicaciones/29/95/09bartol.pdf>
- Bartol Hernández, J. A. et al. (s.f.) [programa informático]. *DispoLex*. Universidad de Salamanca. URL <http://www.dispox.com>
- Barrios Rodríguez, M^a. A. (2015). *Las colocaciones en español*. Madrid: Arco/Libros, S.

L.

- Basso, V. (2017). *La riqueza léxica en la producción escrita de aprendices itálofonos de E/LE* (Tesis de máster inédita, Università degli Studi di Padova).
- Bellón Fernández, J. J. (2003). *Léxico disponible de la provincia de Córdoba* (Tesis doctoral, Universidad de Las Palmas de Gran Canaria).
- Benigno, V. y Vedder, I. (2013). La ricorrenza del lessico di base in produzioni scritte di italiano L2 e L1. *Linguistica e Filologia*, 33, pp. 59-84. Recuperado de <https://aisberg.unibg.it/retrieve/handle/10446/30314/15215/VEDDER%20BENIGNO.pdf>
- Benítez Pérez, P. (1992). Disponibilidad léxica en la zona metropolitana de Madrid. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 1 (1), pp. 71-102.
- Benítez Pérez, P. (1994). Léxico real/irreal en los manuales de español para extranjeros. En S. Montesa Peydró y A. Garrido Moraga (eds.), *Actas del Segundo Congreso Nacional de ASELE. Español para extranjeros: didáctica e investigación* (pp. 325-333). Málaga: ASELE. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/02/02_0323.pdf
- Berton, M. (2014). *La riqueza léxica en la producción escrita de estudiantes suecos de ELE* (Tesis de máster, Stockholms Universitet). Recuperado de <https://www.diva-portal.org/smash/get/diva2:729863/FULLTEXT01.pdf>
- Boers, F. y Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillian.
- Bombarelli, Á. (2005). *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE* (Memoria de máster inédita, Universidad de Salamanca).
- Boogards, P. (1996). Lexicon and grammar in second language learning. En P. Jordens y J. Lalleman (eds.), *Investigating Second Language Acquisition. Studies on Language Acquisition* (vol. 12, pp. 357-379). Berlin: Mouton de Gruyter.
- Boogards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 23, pp. 321-343.
- Borrego Nieto, J. (2004). Algunas preguntas en relación con el concepto de léxico disponible. En M^a. V. Galloso Camacho y J. Prado Aragonés (coords.), *Diccionario, léxico y cultura* (pp. 59-69). Huelva: Universidad de Huelva.
- Bosque, I. (dir.). (2004). *Redes: Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- Bosque, I. (dir.). (2006). *Diccionario combinatorio práctico del español contemporáneo*.

Madrid: SM.

- Buchanan, M. A. (1927). *A Graded Spanish Word Book*. Toronto: University of Toronto Press.
- Caggiula, S. (2013). *El español como lengua extranjera: un estudio de disponibilidad léxica y su aplicación a la enseñanza*. Carolina del Norte: Lulu Press.
- Callealta Barroso, F. J. y Gallego Gallego, D. (2016). Medidas de disponibilidad léxica: comparabilidad y normalización. *Boletín de filología*, 51 (1), pp. 39-92. Recuperado de <https://scielo.conicyt.cl/pdf/bfilol/v51n1/art02.pdf>
- Calvi, M. V. (1995). *Didattica di lingue affini. Spagnolo e italiano*. Milano: Guerini scientifica.
- Calvi, M. V., Guasconi, L. y Provoste, N. (1981 [1977]). *Español situacional*. Milano: Unicopli.
- Canale, M. y Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1 (1), pp.1-47. Recuperado de https://www.researchgate.net/profile/Merrill_Swain/publication/31260438_Theoretical_Bases_of_Communicative_Approaches_to_Second_Language_Teaching_and_Testing/links/0c960516b1dadad753000000/Theoretical-Bases-of-Communicative-Approaches-to-Second-Language-Teaching-and-Testing.pdf
- Capsada Blanch, R. y Torruella Casañas, J. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba*, 44, pp. 347-408. Recuperado de <http://www.usc.es/revistas/index.php/verba/article/view/3155/4585>
- Carballo Picazo, A. (1966). *Español conversacional. Ejercicios de vocabulario*. Madrid: CSIC.
- Carcedo González, A. (1998). Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE. *RILCE. Revista de Filología Hispánica*, 14 (2), pp. 205-224. Recuperado de <https://dadun.unav.edu/bitstream/10171/4356/1/ARTÍCULO%203.%20SOBRE%20LAS%20PRUEBAS%20DE%20DISPONIBILIDAD%20LÉXICA%20PARA%20ESTU>
- Carcedo González, A. (1999a). Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística*, 5-6, pp. 75-94. Recuperado de <https://revistas.uca.es/index.php/pragma/article/view/517/451>
- Carcedo González, A. (1999b). Análisis de errores léxicos del español en la interlingua de los finlandeses. En T. Jiménez Juliá, M. C. Losada Aldrey, J. F. Márquez Caneda y S. Sotelo Docío (eds.), *Español como Lengua Extranjera: Enfoque Comunicativo y Gramática. Actas del IX Congreso Internacional de ASELE, Santiago de*

- Compostela 23-26 de septiembre de 1998* (pp. 465-472). Santiago de Compostela: Universidad de Santiago de Compostela. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/09/09_0468.pdf
- Carcedo González, A. (1999c). Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas. *Documentos de Español Actual*, 1, pp. 73-87.
- Carcedo González, A. (2000a). La lengua como manifestación de otredad cultural (o convergencia intercultural). *Espéculo*, Monográfico “Cultura e intercultural en la enseñanza del español como lengua extranjera”. Recuperado de <http://www.ucm.es/info/especulo/ele/carcedo.html>
- Carcedo González, A. (2000b). Índices léxico-estadísticos y graduación del vocabulario en la enseñanza de E/LE. En M. Franco Figueroa, C. Soler Cantos, J. de Cos Ruiz, M. Rivas Zancarrón y F. Ruiz Fernández (eds.), *Nuevas Perspectivas en la Enseñanza del Español como Lengua Extranjera. Actas del X Congreso de ASELE, Cádiz 22-25 de septiembre de 1999* (pp. 175-183). Cádiz: Servicio de Publicaciones Universidad de Cádiz. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/10/10_0173.pdf
- Carcedo González, A. (2000c). *Disponibilidad léxica en español lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Turun Yliopisto.
- Carcedo González, A. (2001). *Léxico disponible de Asturias*. Turku: Universidad de Turku.
- Carroll, J. B. (1938). Diversity of vocabulary and the harmonic series law of word frequency distribution. *The Psychological Record*, 2, pp. 379-386.
- Casadei, F. (2003). *Lessico e semantica*. Roma: Carocci.
- Castillo Peña, C. (2008). Datos sobre el nuevo horizonte para la lengua española en Italia. En F. San Vicente (ed.), *Textos fundamentales de la lexicografía italoespañola (1917-2007)* (pp. 359-389). Monza: Polimetrica.
- Cervero, M. J. y Pichardo Castro, F. (2000). *Aprender y enseñar vocabulario*. Madrid: Edelsa.
- Chomsky, N. (1959). A review of B. F. Skinner’s Verbal Behavior. *Language*, 35 (1), pp. 26-58. Recuperado de http://www.biolingagem.com/ling_cog_cult/chomsky_1958_skinner_verbalbehavior.pdf
- Chomsky, N. (1986). *Knowledge of Language. Its Nature, Origin and Use*. New York: Pantheon Books.

- Cintrón Serrano, F. (1992). *Índices de riqueza léxica en escolares de Barranquitas* (Tesis de maestría, Universidad de Puerto Rico).
- Consejo de Europa (2001). *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación*. (Traducción al español del Instituto Cervantes en 2002). Madrid: MECD-Anaya. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf
- Contreras Izquierdo, N. M. (2008). La enseñanza-aprendizaje de lenguas extranjeras y las TICs: el caso del Español como lengua extranjera (ELE). *Iniciación a la investigación*, 1. Recuperado de <https://revistaselectronicas.ujaen.es/index.php/ininv/article/view/233/214>
- Contreras Izquierdo, N. M. (2014). ¿Usamos el diccionario? (iii) Las variedades del español en la enseñanza de ELE. En N. M. Contreras Izquierdo (ed.), *La enseñanza del español como LE / L2. Actas del XXIV Congreso Internacional ASELE, Jaén 18-21 de septiembre de 2013* (pp. 219-231). Málaga: ASELE. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/24/24_219.pdf
- Contreras Izquierdo, N. M. (2018). Variedades lingüísticas y ELE. La variación diatópica en el léxico coloquial del español: formación del profesor, recursos, orientaciones metodológicas e implicaciones didácticas. En D. G. Níkleva (coord.), *Necesidades y tendencias en la formación del profesorado de español como lengua extranjera* (pp. 283-310). Bern: Peter Lang.
- Contreras Izquierdo, N. M. y Rey Arranz, R. M. (2018). Hacia un modelo de análisis multidimensional de las secuencias formulaicas para su aplicación en ELE/EL2. *RAEL: Revista Electrónica de Lingüística Aplicada*, 17 (1), pp. 20-44. Recuperado de <http://www.aesla.org.es/ojs/index.php/RAEL/article/view/340/231>
- Corder, S. P. (1967). The significance of learners' errors. *IRAL. International review of Applied Linguistics*, 5 (4), pp. 161-170. Recuperado de <http://files.eric.ed.gov/fulltext/ED019903.pdf>
- Crossley, S., Salsbury, T. y McNamara, D. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29 (2), pp. 243-263.
- Cruz Piñol, M. (2015). Léxico y ELE: enseñanza/aprendizaje con tecnologías. *Journal of Spanish Language Teaching*, 2 (2), pp. 165-179.
- Cuba Vega, L. E. (2016). Sexo y vocabulario: relación en escolares habaneros. *Enunciación*, 21 (1), pp. 80-90. Recuperado de <https://revistas.udistrital.edu.co/index.php/enunc/article/view/10139/11753>
- Cuba Vega, L. E. y Cuba, Y. M. (2004). Las pruebas de riqueza léxica y su aplicación en Español como Lengua Extranjera (ELE). En A. Díez Mediavilla (ed.), P. Couto Cantero, F. Vieito Liñares y E. Aradas Carollo (coords.), *Actas VIII Congreso Internacional Sociedad Española de Didáctica de la Lengua y la Literatura, La*

Habana 5-9 de diciembre de 2004: Homenaje a María Zambrano y Alejo Carpentier (pp. 815-830).

- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of educational research*, 49 (2), pp. 222-251.
- De B. Moll, F. (1970 [1954]). *Curso breve de español para extranjeros. Grado Medio*. Palma de Mallorca: Moll.
- Del Barrio de la Rosa, F. (2016). Algunas observaciones sobre la disponibilidad léxica en estudiantes itálofonos de español. En E. Sainz González, I. Solís García, F. Del Barrio de la Rosa e I. Arroyo Hernández (eds.), *Geométrica explosión. Estudios de lengua y literatura en homenaje a René Lenarduzzi* (pp. 127-143). Venezia: Edizioni Ca' Foscari. Recuperado de <https://edizionicafoscari.unive.it/media/pdf/books/978-88-6969-080-8/978-88-6969-080-8-ch-08.pdf>
- Del Barrio de la Rosa, F. (2017a). Los estudiantes itálofonos de ELE y los estudios de disponibilidad léxica en español. *Revista Nebrija de Lingüística Aplicada*, 22, pp. 52-57. Recuperado de https://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_58de02da8bc1d.pdf
- Del Barrio de la Rosa, F. (ed.). (2017b). *Palabras. Vocabulario. Léxico. La lexicología aplicada a la didáctica y a la diacronía*. Venezia: Edizioni Ca' Foscari – Digital Publishing. Recuperado de https://edizionicafoscari.unive.it/media/pdf/books/978-88-6969-170-6/978-88-6969-170-6_3GEHh4o.pdf
- Del Barrio de la Rosa, F. (2018). Pares léxicos en el léxico disponible de estudiantes italianos. *Lingue e linguaggi*, 26, pp. 173-196. Recuperado de https://iris.unive.it/retrieve/handle/10278/3702889/136443/2018_L%26L_Pares.pdf
- Del Barrio de la Rosa, F. y Mae Vann, M. (2018). *Disponibilidad léxica de los estudiantes de español en Italia. Estudio y diccionarios*. Canterano: Aracne.
- Dewaele, J. M. y Pavlenko, A. (2003). Productivity and lexical diversity in native and nonnative speech: A study of cross-cultural effects. En V. Cook (ed.), *The effects of the second language on the first* (pp. 120-141). Clevedon: Multilingual Matters.
- Di Gesù, F. (2016). *Linguistica contrastivo-percettiva di lingue tipologicamente affini: italiano e spagnolo*. Palermo: University press.
- Dimitrijevíc, N. (1969). *Lexical Availability. A new aspect of the lexical availability of secondary school children*. Heidelberg: Julius Gross Verlag.
- Domínguez Vázquez, M^a. J. y Sanmarco Bande, M^a. T. (eds.). (2017). *Lexicografía y didáctica*. Frankfurt am Main: Peter Lang.
- Doughty, C. (2000). La negociación del entorno lingüístico de la L2. En C. Muñoz (ed.),

Segundas lenguas. Adquisición en el aula (pp. 163-194). Barcelona: Ariel Lingüística.

Doughty, C. y Long, M. (eds.). (2002). *The Handbook of Second Language Acquisition*. Oxford: Oxford Blackwell.

Dulay, H. y Burt, M. (1977). Remarks on Creativity in Language Acquisition. En M. Burt, H. Dulay y M. Finocchiaro (eds.), *Viewpoints on English as a Second Language* (pp. 95-126). New York: Regents.

Echeverría, M. S. *et al.* (1987). Disponibilidad léxica en Educación Media. Resultados cuantitativos. *RLA. Revista de lingüística teórica y aplicada*, 25, pp. 55-115.

Echeverría, M. S. *et al.* (1992). Evaluación de la Riqueza Léxica en Estudiantes de Último Año de Enseñanza Media. *Estudios Filológicos*, 27, pp. 59-72.

Echeverría, M. S. *et al.* (2005) [programa informático]. *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción de Chile: Universidad de Concepción.

Echeverría, M. S. *et al.* (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), pp. 81-91.

Ellis, N. C. (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4 (2), pp. 139-155.

Fernández Juncal, C. y Hernández Muñoz, N. (2018). Vías de transformación en la enseñanza de lenguas con mediación tecnológica. *CLAC. Circulo de Lingüística Aplicada a la Comunicación*, 76, pp. 3-12. Recuperado de <https://revistas.ucm.es/index.php/CLAC/article/view/62494/4564456548716>

Fernández Juncal, C. y Hernández Muñoz, N. (2019). Disponibilidad léxica y socionomástica. *Ogigia. Revista electrónica de estudios hispánicos*, 25, pp. 185-210. Recuperado de <https://revistas.uva.es/index.php/ogigia/article/view/2916/2881>

Fernández-Merino Gutiérrez, P. V. (2011). Presencia del léxico disponible de inmigrantes en glosarios específicos de vocabulario. *Cuadernos Comillas*, 2, pp. 1-18. Recuperado de <https://www.fundacioncomillas.es/cuadernos-comillas/cuadernos/lexico/pdf/cuaderno.pdf>

Frey Pereyra, M. L. H. (2006). Disponibilidad léxica y escritura del español como lengua extranjera: propuesta de comparación de dos corpus. *Interlingüística*, 17, pp. 366-373.

Frey Pereyra, M. L. H. (2008). *El léxico disponible en los escritos de alumnos de español*

- como lengua extranjera* (Tesis doctoral, Universidad de Alcalá).
- Gallego Gallego, D. (2014). *Léxico disponible de estudiantes de español como lengua extranjera en la Comunidad de Madrid* (Tesis doctoral, Universidad de Alcalá).
- Galoso Camacho, M^a. V. (2002). *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Ediciones Universidad de Salamanca.
- García Fernández, N. y Sánchez Lobato, J. (1987 [1981]). *Español 2000. Nivel elemental*. Madrid: Coloquio.
- García Hoz, V. (1953). *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid: CSIC.
- García Marcos, A. (2019). *Adquisición del español como segunda lengua. El caso de la integración lingüística de escolares inmigrantes en Almería*. La Cañada de San Urbano, Almería: Editorial Universidad de Almería.
- García Parejo, I. (2018). Comunicación intercultural y enseñanza de ELE: ¿qué competencias para qué desafíos? En D. G. Níkleva (coord.), *Necesidades y tendencias en la formación del profesorado de español como lengua extranjera* (pp. 311-338). Bern: Peter Lang.
- García Rosas, M. P. (1996). *Riqueza léxica en textos de estudiantes de español como lengua extranjera* (Memoria de máster inédita, Universidad de Alcalá).
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Arnold.
- Gass, S. (1997). *Input, Interaction and the Second Language Learner*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Givón, T. (ed.). (1979). *Syntax and Semantics. Vol. 12. Discourse and syntax*. New York: Academic Press.
- Gómez Devís, M^a. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones* (Tesis doctoral, Universitat de València).
- Gómez Devís, M^a. B. (2019). A propósito de las redes semánticas en el léxico disponible de escolares de primero de Educación Primaria. *Ogigia. Revista electrónica de estudios hispánicos*, 25, pp. 165-183. Recuperado de <https://revistas.uva.es/index.php/ogigia/article/view/2917/2414>
- Gómez Molina, J. R. (2004). Los contenidos léxico-semánticos. En J. Sánchez Lobato e I. Santos Gargallo (coords.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2) /lengua extranjera (LE)* (pp. 789-810). Madrid:

SGEL.

- Gómez Molina, J. R. y Gómez Devís, M^a. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos. Estudio de estratificación sociolingüísticas*. Valencia: Universitat de València.
- González Fernández, J. (2013). La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera. *MarcoELE*, 16. Recuperado de http://marcoele.com/descargas/16/gonzalez-disponibilidad_lexica_turcos.pdf
- González Fernández, J. (2014). Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera. *Revista Nebrija de Lingüística Aplicada*, 16, pp. 41-53. Recuperado de <https://revistas.nebrija.com/revista-linguistica/article/view/223/192>
- Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1956). *L'élaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1964). *L'élaboration du français fundamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Gouws, R. H. (2017). La sociedad digital y los diccionarios. En M^a. J. Domínguez Vázquez y M^a. T. Sanmarco Bande (eds.), *Lexicografía y didáctica* (pp. 17-34). Frankfurt am Main: Peter Lang.
- Griffin, K. (2005). *Lingüística aplicada a la enseñanza del español como 2/L*. Madrid: Arco/Libros, S. L.
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire. Essai de méthodologie*. Paris: Presses Universitaires de France.
- Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Paris: Presses Universitaires de France.
- Haché, A. M. (1991). Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna. En H. López Morales (ed.), *La enseñanza del español como lengua materna* (pp. 47-60). Río Piedras: Universidad de Puerto Rico.
- Haensch, G., Wolf, L., Ettinger, S. y Reinhold, W. (1982). *La Lexicografía. De la Lingüística Teórica a la Lexicografía Práctica*. Madrid: Gredos.
- Hallebeek, J. (1986). Las palabras funcionales del español. *Boletín AEPE*, 34-35, pp. 205-216. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/aepe/pdf/boletin_34-35_18_86/boletin_34-35_18_86_24.pdf

- Halliday, M. A. K. (1973). *Explorations in the functions of language*. London: Arnold.
- Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the Development of Language*. London: Arnold.
- Ham Chande, R. (1979). Del 1 al 100 en lexicografía. En L. Fernando Lara, R. Ham Chande y M^a. I. García Hidalgo (eds.), *Investigaciones lingüísticas en lexicografía* (pp. 110-132). México: El Colegio de México.
- Henmon, V. A. C. (1924). *A French Word Book Based on a Count of 400.000 running Words*. Madison (Wisconsin): Bureau of Educational Research, University of Wisconsin.
- Hernández Muñoz, N. (2005). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología. En E. Díez Villoria, B. Zubiauz de Pedro y M^a. Á. Mayor Cinca (coords.), *Estudio sobre la adquisición del lenguaje* (pp. 942-953). Salamanca: Ediciones Universidad de Salamanca.
- Hernández Muñoz, N. (2006). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Ediciones Universidad de Salamanca. Recuperado de https://www.academia.edu/10348393/Hacia_teor%C3%ADa_cognitiva_integrada_de_la_disponibilidad_l%C3%A9xica._El_l%C3%A9xico_disponible_de_los_estudiantes_castellano-manchegos_2006
- Hernández Muñoz, N. (2014). Categorías en el léxico bilingüe perspectivas desde el priming semántico interlenguas y la disponibilidad léxica. *RAEL: Revista Electrónica de Lingüística Aplicada*, 13 (1), pp. 19-38. Recuperado en <http://www.aesla.org.es/ojs/index.php/RAEL/article/view/21/3>
- Hernández Muñoz, N. (2015). La evaluación de la competencia léxica adulta una aproximación a través de la disponibilidad léxica y la especialización académica en preuniversitarios. *Revista de Filología de la Universidad de La Laguna*, 33, pp. 79-99. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=5120308>
- Hernández Muñoz, N. (2018). Encuentros entre el léxico disponible en primera y segunda lengua. En *International conference "Exploring the lexicon of bilingual and plurilingual learners: lexical availability and vocabulary acquisition"*, Universidad de La Rioja, Logroño 4-5 de octubre de 2018 (pp. 15-16). Logroño: Universidad de La Rioja, Servicio de Publicaciones.
- Hernández Muñoz, N., Izura, C. y Tomé Cornejo, C. (2014). Cognitive Factors of Lexical Availability in a Second Language. En R. M^a. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 169-188). New York: Springer.
- Hernández Muñoz, N. y Borrego Nieto, J. (2004). Cuestiones metodológicas sobre los estudios de disponibilidad léxica. En M. Villayandre Llamazares (coord.), *Actas del*

V Congreso de Lingüística General, León 5-8 de marzo de 2002 (vol. 2, pp. 1519-1527). Madrid: Arco/Libros, S. L.

Hernández Muñoz, N. y López García, M. (2014). Análisis de las relaciones semánticas a través de una tarea de libre asociación en español con mapas auto-organizados. *RLA. Revista de Lingüística Teórica y Aplicada*, 52 (2), pp. 189-212. Recuperado de https://scielo.conicyt.cl/pdf/rla/v52n2/art_09.pdf

Hernández Muñoz, N. y Román-Mendoza, E. (2018). Aprende conmigo: exigencias de la era digital para las buenas prácticas en la enseñanza de segundas lenguas. *CLAC. Círculo de lingüística aplicada a la comunicación*, 76, pp. 31-48. Disponible en: <https://revistas.ucm.es/index.php/CLAC/article/view/62496/4564456548717>

Hernández Muñoz, N. y Tomé Cornejo, C. (2017). Léxico disponible en primera y segunda lengua: bases cognitivas. En F. Del Barrio de la Rosa (ed.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía* (pp. 99-122). Venezia: Edizioni Ca' Foscari. Recuperado de https://edizionicafoscari.unive.it/media/pdf/books/978-88-6969-170-6/978-88-6969-170-6-ch-07_nmmSszY.pdf

Hidalgo Gallardo, M. (2017a). Sobre la disponibilidad léxica en ELE. *Boletín de ASELE*, 56, pp. 83-94. Recuperado de https://www.researchgate.net/publication/318133905_Sobre_la_disponibilidad_lexica_en_ELE_Revision_de_la_literatura

Hidalgo Gallardo, M. (2017b). *La disponibilidad léxica como método de detección del vocabulario y de su elección en manuales: aplicación en una muestra de estudiantes sinohablantes de ELE* (Tesis doctoral inédita, Universidad de Jaén).

Higueras García, M. (2006). *Estudio de las colocaciones léxicas y su enseñanza en español como lengua extranjera*. Málaga: ASELE.

Higueras García, M. (2008). El diccionario *Práctico* en la práctica docente del español como lengua extranjera. En D. Azorín Fernández (ed.), *Actas del II Congreso Internacional de Lexicografía Hispánica, Alicante 19-23 de septiembre de 2006. El diccionario como puente entre las lenguas y culturas del mundo* (pp. 436-443). Alicante: Biblioteca Virtual Miguel de Cervantes.

Higueras García, M. (2017). Logros y retos de la enseñanza del léxico. En F. Herrera (ed.), *Enseñar léxico en el aula de español: el poder de las palabras* (pp. 13-19). Barcelona: Difusión.

Higueras García, M. (2019). Del enfoque léxico al enfoque léxico optimizado. En E. Fernández Martín (ed.), *Investigaciones lexicográficas y lexicológicas: estudios actuales sobre el léxico en el ámbito hispánico* (p. 146). Jaén: Editorial de la Universidad de Jaén.

Hymes, D. (1972). On communicative competence. En J. B. Pride y J. Holmes (eds.),

- Sociolinguistics: selected readings* (pp. 269-293). London: Penguin Books.
- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes-Biblioteca Nueva. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/
- Instituto Cervantes. Banco de datos (CAES) [en línea]. *Corpus de aprendices del español*. URL <http://galvan.usc.es/caes>
- Izquierdo Gil, M. C. (2005). *La selección del léxico en la enseñanza del español como lengua extranjera*. Málaga: ASELE.
- Jarvis, S. (2002). Short texts, best-fitting curves, and new measures of lexical diversity. *Language Testing*, 19 (1), pp. 57-84.
- Jarvis, S. (2013). Capturing diversity in lexical diversity. *Language Learning*, 63, pp. 87-106.
- Jiménez Catalán, R. M^a. (2002). El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas. *Atlantis*, XXIV (2), pp. 149-162. Recuperado de https://www.atlantisjournal.org/old/Papers/24_1/jimenez.pdf
- Jiménez Catalán, R. M^a. (2017). Estudios de disponibilidad léxica en español y en inglés: revisión de sus fundamentos empíricos y metodológicos. *Revista Nebrija de Lingüística Aplicada*, 22, pp. 16-31. Recuperado de <https://revistas.nebrija.com/revista-linguistica/article/view/2/4>
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers*, 53, pp. 61-79. Recuperado de <http://journals.lub.lu.se/index.php/LWPL/article/view/2273/1848>
- Juilland, A. y Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. London, The Hague: Mouton.
- Juilland, A. (ed.). (1970). *Frequency dictionary of French Words*. The Hague: Mouton, The Romance Languages and their Structures, First Series.
- Käding, J. W. (1897). *Häufigkeitwörterbuch der Deutschen Sprache*. Steiglitz bei Berlin: der Herausgeber.
- Kilgarriff, A. et al. (2004) [programa informático]. *The Sketch Engine*. URL <http://www.sketchengine.eu>
- Kumaravadivelu, B. (2006a). *Understanding Language Teaching: From Method to Postmethod*. London: Lawrence Erlbaum.
- Kumaravadivelu, B. (2006b). TESOL Methods: Changing Tracks, Challenging Trends. *TESOL Quarterly*, 40 (1), pp. 59-81. Recuperado de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.542.6139&rep=rep1&type=pdf>

- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Krashen, S. (1985). *The input hypothesis*. Oxford: Pergamon Press.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
- Larsen-Freeman, D. (1986). *Techniques and principles in language teaching*. Oxford: Oxford University Press.
- Larsen-Freeman, D. y Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced language learner. *The Modern Language Journal*, 75 (4), pp. 440-448.
- Laufer, B. (2003). The influence of L2 on L1 collocational knowledge and on L1 lexical diversity in free written expression. En V. Cook (ed.), *Effects of the second language on the first* (pp. 19-31). Clevedon: Multilingual Matters.
- Laufer, B. y Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16 (3), pp. 307-22.
- Lemmouh, Z. (2010). *The Relationship among Vocabulary Knowledge, Academic Achievement and the Lexical Richness in Writing in Swedish University Students of English* (Tesis doctoral, Stockholms Universitet). Recuperado de <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-42765>
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. London: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. London: Language Teaching Publications.
- Lewis, M. (2000). *Teaching collocations. Further developments in the Lexical Approach*. London: Language Teaching Publications.
- Liceras Muñoz, J. (1992). *La adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlingua*. Madrid: Visor.
- Lin, J. (2006). *El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera* (Memoria de máster inédita, Universidad de Alcalá).

- Lindqvist, C. (2010). La richesse lexicale dans la production orale de l'apprenant avancé de français. *The Canadian Modern Language Review*, 66 (3), pp. 393-420.
- Linnarud, M. (1986). *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*. Malmö: CWK Gleerup.
- Llobera, M. (coord.). (1995). *Competencia comunicativa: documentos básicos en la enseñanza de lenguas*. Madrid: Edelsa.
- Long, M. (1985). Input and Second Language Acquisition Theory. En S. Gass y C. Madden (eds.), *Input in second language acquisition* (pp. 377-393). Rowley: Newbury House.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. En W. Ritchie y T. Bhatia (eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.
- López Chávez, J. (1992). Alcances panhispánicos del léxico disponible. *Lingüística*, 4, pp. 26-124.
- López Chávez, J. (1993). *El léxico disponible de escolares mexicanos*. México: Editorial Alhambra Mexicana.
- López Chávez, J. (1994). Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica. *REALE: revista de estudios de adquisición de la lengua española*, 1, pp. 67-84.
- López Chávez, J. (1995). Léxico fundamental panhispánico: realidad o utopía. En A. Matus Oliver (ed.), *Actas del III Congreso Internacional sobre el Español de América* (vol. 2, pp. 1006-1014). Santiago de Chile: Universidad Católica de Chile.
- López Chávez, J. y Strassburger Frías, C. (1987). Otro cálculo del índice de disponibilidad léxica. En *Presente y perspectiva de la lingüística computacional en México. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada* (pp. 91-111). México: Universidad Nacional Autónoma de México.
- López Chávez, J. y Strassburger Frías, C. (1991). Un modelo para el cálculo del índice de disponibilidad léxica individual. En H. López Morales (ed.), *La enseñanza del español como lengua materna. Actas del II Seminario Internacional sobre Aportes de la Lingüística a la Enseñanza del Español como Lengua Materna* (pp. 91-112). Río Piedras: Universidad de Puerto Rico.
- López García, Á. (1988). *La psicolingüística*. Madrid: Síntesis.
- López González, A. M. (2010). La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. *RedELE*, 18. Recuperado de <https://www.educacionyfp.gob.es/dam/jcr:4f5777a6-d513-4715-910a-01c047353995/2010-redele-18-01lopez-pdf.pdf>

- López González, A. M. (2014a). Desarrollo de los estudios de disponibilidad léxica en Español Lengua Extranjera (ELE). En N. M. Contreras Izquierdo (ed.), *La enseñanza del español como LE / L2. Actas del XXIV Congreso Internacional ASELE, Jaén 18-21 de septiembre de 2013* (pp. 411-422). Málaga: ASELE. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/24/24_411.pdf
- López González, A. M. (2014b). La estructura interna del léxico disponible en español como lengua extranjera (ELE) de los preuniversitarios polacos. *Studia Romanica Posnaniensa*, 41 (1), pp. 45-61. Recuperado de <https://repozytorium.amu.edu.pl/bitstream/10593/11160/1/04%20srp41-1%20Antonio%20Mar%C3%ADa%20López%20González.pdf>
- López-Mezquita Molina, M^a. T. (2005). *La evaluación de la competencia léxica: tests de vocabulario. Su fiabilidad y validez* (Tesis doctoral, Universidad de Granada).
- López Morales, H. (1973). *Disponibilidad léxica en escolares de San Juan*. MS.
- López Morales, H. (1978). Frecuencia léxica, disponibilidad y programación curricular. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 6 (1), pp. 73-86.
- López Morales, H. (1979). Disponibilidad léxica y estratificación socioeconómica. En H. López Morales (ed.), *Dialectología y sociolingüística. Temas puertorriqueños* (pp. 173-181). Madrid: Hispanova de Ediciones.
- López Morales, H. (1983). Lingüística estadística. En H. López Morales (coord.), *Introducción a la lingüística actual* (pp. 209-225). Madrid: Playor.
- López Morales, H. (1984). *La enseñanza de la lengua materna. Lingüística para maestros de español*. Madrid: Editorial Playor.
- López Morales, H. (1990 [1986]). *Enseñanza de la lengua materna*. Madrid: Ed. Plaza Mayor.
- López Morales, H. (1995). Política y planificación de la enseñanza de la lengua materna. En F. J. García Marcos (coord.), *Actas de las I Jornadas almerienses sobre política lingüística, Almería, 12-30 de noviembre de 1990* (pp. 44-59). Almería: Universidad de Almería, Servicio de Publicaciones.
- López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros, S. L.
- López Morales, H. (2011). Los índices de 'riqueza léxica' y la enseñanza de lenguas. En J. de Santiago Guervós, H. Bongaerts, J. J. Sánchez Iglesias y M. Seseña (eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE. Actas del XXI Congreso Internacional de la ASELE, Salamanca 29 de septiembre-2 de octubre de 2010* (vol. 1, pp. 15-28). Salamanca: Imprenta Kadmos. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0015.pdf

- López Pérez, M. V. (2013). Criterios de selección de voces para la enseñanza del español como L2 en contextos escolares. *Revista de Educación*, Extraordinario 2013, pp. 154-176. Recuperado de http://disde.minedu.gob.pe/bitstream/handle/123456789/2424/2013_LÓPEZ_Criterios%20de%20selección%20de%20voces%20para%20la%20enseñanza%20del%20español%20como%20L2%20en%20contextos%20escolares.pdf?sequence=1&isAllowed=y
- Lorán Santos, R. y López Morales, H. (1983). *Nouveau calcul de l'indice de disponibilité*. MS.
- Lucas Puerta, J. (2006). *La competencia léxica en el discurso escrito en e/le de aprendices francófonos de ascendencia hispanófono* (Memoria de máster, Universidad de Barcelona). Recuperado de <http://www.educacionyfp.gob.es/en/dam/jcr:77789d3e-0806-4a01-902e-7ab443cfd297/2007-bv-07-12lucas-pdf.pdf>
- Mackey, W. F. (1971a). *Le vocabulaire disponible du Français*, 2 vols. Paris: Didier.
- Mackey, W. F. (1971b). *Le sondage dans les enquêtes de disponibilité*. Québec: CIRB.
- Madrigal, M. y Vargas, E. (2016). Índice de riqueza léxica en redacciones escritas por estudiantes universitarios. *Káñina*, 40 (s1), pp. 139-147. Recuperado de <http://www.redalyc.org/jatsRepo/442/44254856010/html/index.html>
- Maldonado González, M^a. C. (dir.). (2002). *Diccionario de español para extranjeros*. Madrid: Ediciones SM.
- Maldonado González, M^a. C. (2017). El diccionario didáctico monolingüe y bilingüe. En M^a. J. Domínguez Vázquez y M^a. T. Sanmarco Bande (eds.), *Lexicografía y didáctica* (pp. 161-175). Frankfurt am Main: Peter Lang.
- Maldonado González, M^a. C. (2019). Las investigaciones de mercado en lexicografía comercial: un aprendizaje para el mundo académico e investigador. *RILE. Revista Internacional de Lenguas Extranjeras*, 10, pp. 101-118. Recuperado de <https://www.raco.cat/index.php/RILE/article/view/354775/446765>
- Malvern, D. y Richards, B. (2009). A new method of measuring rare word diversity: The example of L2 learners of French. En B. Richards, H. Daller, D. Malvern, P. Meara, J. Milton y J. Treffers-Daller (eds.), *Vocabulary Studies in first and second language acquisition* (pp. 164-178). Basingstoke: Palgrave Macmillan.
- Marconi, D. (1999). *La competenza lessicale*. Bari: Editori Laterza.
- Martín García, J. (1999). *El diccionario en la enseñanza del español*. Madrid: Arco/Libros, S. L.
- Martín Peris, E. y Sans Baulenas, N. (1997). *Gente: curso comunicativo basado en el enfoque por tareas. Libro del alumno*. Barcelona: Difusión.

- Martín Peris, E. y Sans Baulenas, N. (2004). *Gente 2. Nueva edición*. Barcelona: Difusión.
- Martín Peris, E. (ed.). (2008) [en línea]. *Diccionario de términos clave de ELE*. Madrid: SGEL. URL https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm
- Martínez Ezquerro, A. (2001). Competencia curricular y diccionario: propuesta teórica. *Contextos educativos. Revista de Educación*, 4, pp. 315-328. Recuperado de <https://publicaciones.unirioja.es/ojs/index.php/contextos/article/view/499/463>
- Martínez Olmos, E. (2007). *Disponibilidad léxica en las comunidades de habla alicantinas* (Tesis doctoral, Universidad de Alicante). Recuperado de https://rua.ua.es/dspace/bitstream/10045/14307/1/tesis_esther_martinez.pdf
- Meara, P. (1983). *Vocabulary in a second language*. London: CILT.
- Melloni, A. y Capanaga, P. (1993). *Itinerarios por el español*. Bologna: Zanichelli.
- Michéa, R. (1950). Vocabulaire et culture. *Les Langues Modernes*, 44, pp. 188-189.
- Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, pp. 338-344.
- Minera Reyna, L. E. (2010). El cuestionario MAALE, técnica para recolección de datos de las variables afectivas motivación y actitudes en el aprendizaje de una lengua extranjera. *RedELE*, 19. Recuperado de <https://www.educacionyfp.gob.es/dam/jcr:dda6745f-400e-48cc-9c7e-fde2ce1d51c4/2010-redele-19-03minera-pdf.pdf>
- Minervini, R. (2008). El diccionario combinatorio para el desarrollo del léxico en aula de E/LE. En A. Nomdedeu Rull y E. Sánchez García (eds.), *Perspectivas de la lexicografía del español en el siglo XXI. Actas del Coloquio Internacional, Nápoles 6-7 de marzo de 2006* (pp. 89-119). Napoli: Università degli Studi di Napoli "L'Orientale".
- Miranda García, A. (1993). Modelo teórico de lexicón mental. *Cauce. Revista Internacional de Filología, Comunicación y sus Didácticas*, 16, pp. 91-100. Recuperado de https://cvc.cervantes.es/literatura/cauce/pdf/cauce16/cauce16_05.pdf
- Morante Vallejo, R. (2005). *El desarrollo del conocimiento léxico en segundas lenguas*. Madrid: Arco/Libros, S. L.
- Moreno Fernández, F. (2012). Disponibilidad léxica: cuestiones metodológicas. A propósito de disponibilidad léxica de los estudiantes hispanos de Redwood City, CA. *Revista Nebrija de Lingüística Aplicada*, 11, pp. 53-61. Recuperado de <https://revistas.nebrija.com/revista-linguistica/article/view/174/145>

- Müller, C. (1968). *Estadística lingüística*. Madrid: Gredos.
- Müller-Spitzer, C. (2017). Investigación sobre el uso de diccionarios en la era digital. En M^a. J. Domínguez Vázquez y M^a. T. Sanmarco Bande (eds.), *Lexicografía y didáctica* (pp. 55-81). Frankfurt am Main: Peter Lang.
- Nalesso, G. (2018a). La eficacia de actividades específicas para el aprendizaje del léxico en ELE. Estudio de caso. En M^a. Bargalló Escrivá, E. Forgas Berdet y A. Nomdedeu Rull (eds.), *Léxico y cultura en LE/L2: corpus y diccionarios. Actas del XXVIII Congreso Internacional ASELE, Tarragona 6-9 de septiembre de 2017* (pp. 505-514).
- Nalesso, G. (2018b). El desarrollo de la competencia léxica de estudiantes italianos universitarios de ELE. *Orillas. Rivista d'Ispanistica*, 7, pp. 381-394. Recuperado de http://orillas.cab.unipd.it/orillas/articoli/numero_7/07Nalesso_Astilleros.pdf
- Nalesso, G. (2019a). Propuesta de investigación en torno al desarrollo de la competencia léxica en estudiantes italianos universitarios de ELE. En E. Fernández Martín (ed.), *Investigaciones lexicográficas y lexicológicas: estudios actuales sobre el léxico en el ámbito hispánico* (p. 159). Jaén: Publicaciones de la Universidad de Jaén.
- Nalesso, G. (2019b). La eficacia de actividades para el aprendizaje del léxico especializado del arte en ELE. Estudio de caso. En C. Ramos Méndez y P. Salamanca Fernández (eds.), *Competencias en el aula de español como lengua extranjera en contextos universitarios y profesionales. Actas del Congreso Internacional UniPro 2018 Múnich* (pp. 100-116). Múnich: Hochschule für Angewandte Sprachen.
- Nation, P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle an Heinle.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nattinger, J. R. y De Carrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Njock, P. E. (1979). *L'univers familier de l'enfant africain*. Québec: CIRB.
- Paolini, A. (2017). *Estudio sobre disponibilidad léxica en alumnos de ELE en la Universidad de Padua* (Tesis de máster, Università degli Studi di Padova). Recuperado de http://tesi.cab.unipd.it/54477/1/ALESSIA_PAOLINI_2017.pdf
- Paredes García, F. (2014). A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE. *Revista Nebrija de Lingüística Aplicada*, 16, pp. 54-59. Recuperado de <https://revistas.nebrija.com/revista-linguistica/article/view/224/193>
- Paredes García, F. (2015). Disponibilidad Léxica y enseñanza de ELE: el léxico

disponible como fuente curricular y como recurso en el aula. *Linred*, 13. Recuperado de http://www.linred.es/monograficos_pdf/LR_monografico13-articulo2.pdf

- Paredes García, F. (2017). La comparabilidad de los trabajos de disponibilidad léxica. *Revista Nebrija de Lingüística Aplicada*, 22, pp. 71-77. Recuperado de <https://revistas.nebrija.com/revista-linguistica/article/view/6/9>
- Pastora Herrero, J. F. (1990). *El vocabulario como agente de aprendizaje*. Madrid: La Muralla.
- Penadés Martínez, I. (1999). *La enseñanza de las unidades fraseológicas*. Madrid: Arco/Libros, S. L.
- Pérez Navarro, J. y Poletini, C. (1989). *¡Claro que sí! Curso de español para italianos. Nivel elemental e intermedio*. Padova: Unipress.
- Pérez Serrano, M. (2009). *Estudio de disponibilidad léxica en estudiantes de E/LE en los centros de interés “medios de transporte” y “profesiones y oficios”* (Memoria de máster, Instituto Cervantes-UIIMP). Recuperado de <http://www.educacionyfp.gob.es/dam/jcr:7edc1722-d2ac-45a2-8d24-384f4794fbab/2011-bv-12-22perez-serrano-pdf.pdf>
- Pérez Serrano, M. (2017). *La enseñanza-aprendizaje del vocabulario en ELE desde los enfoques léxicos*. Madrid: Arco/Libros, S. L.
- Pfeffer, J. A. (1964). *Grundeutsch Basic (spoken) German Word list Grundstufe*. New Jersey: Prentice Hall Inc. Englewood Cliffs.
- Portela, C. (1992). *Índices de riqueza léxica en estudiantes del primer año universitario* (Tesis de maestría, Pontificia Universidad Católica Madre y Maestra).
- Preston, D. R. (1989). *Sociolinguistics and second language acquisition*. Oxford: Blakwell.
- Read, J. A. S. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Real Academia Española. (2005) [en línea]. *Diccionario panhispánico de dudas*. URL <http://www.rae.es/recursos/diccionarios/dpd>
- Real Academia Española. (2009). *Nueva Gramática de la Lengua Española, vol. I: Morfología y Sintaxis*. Madrid: Espasa.
- Real Academia Española. (2010). *Ortografía de la lengua española*. Madrid: Espasa.
- Real Academia Española. (2017) [en línea]. *Diccionario de la Lengua Española*, 23.1ª edición. URL <http://dle.rae.es>
- Real Academia Española. Banco de datos (CREA) [en línea]. *Corpus de referencia del*

español actual. URL <http://corpus.rae.es/creanet.html>

- Rey Castillo, M. (2017). Criterios macroestructurales y microestructurales en los diccionarios monolingües para el aprendizaje del léxico en español como lengua extranjera: entre la teoría y la práctica. *MarcoELE*, 25. Recuperado de <http://marcoele.com/descargas/25/rey-diccionarios.pdf>
- Rey Castillo, M. (2018). *El diccionario en la enseñanza del Español como Lengua Extranjera (ELE). Propuesta de criterios macroestructurales y microestructurales para un Diccionario Monolingüe de Aprendizaje (DMA)* (Tesis doctoral inédita, Universidad Internacional Iberoamericana).
- Reyes Díaz, M. J. (2007). Apuntes para la enseñanza del vocabulario. *Revista de Filología de la Universidad de La Laguna*, 25, pp. 529-538. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2528465>
- Reyes Díaz, M. J. (2010). Riqueza léxica de textos redactados por alumnos de bachillerato de Las Palmas de Gran Canaria. *Anuario de Lingüística Hispánica*, XXIII-XXIV, pp. 147-163. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=3209971>
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10 (1), pp 77-89.
- Richards, J. C. y Rogers T. S. (2001). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- Roberto, J. A., Martí, M. A. y Salamó Llorente, M^a. (2012). Análisis de la riqueza léxica en el contexto de la clasificación de atributos demográficos latentes. *Procesamiento de Lenguaje Natural*, 48, pp. 97-104. Recuperado de <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4493/2661>
- Robles Ávila, S. (coord.). (2006). *La enseñanza del español como lengua extranjera a la luz del Marco Común Europeo de Referencia. Diseño curricular de los cursos para extranjeros de la Universidad de Málaga*. Málaga: Universidad de Málaga.
- Rodríguez Bou, I. (dir.). (1952). *Recuento de vocabulario español*. San Juan de Puerto Rico: Organización de Estados Americanos.
- Romero Gualda, M^a. V. (2008). *Léxico del español como segunda lengua: aprendizaje y enseñanza*. Madrid: Arco/Libros, S. L.
- Rubio Lastra, M. (2018). Disponibilidad léxica de 52 estudiantes taiwaneses universitarios de ELE A1. *MarcoELE*, 27. Recuperado de https://marcoele.com/descargas/27/rubio-disponibilidad_lexica_taiwanenes_a1.pdf
- Rubio Sánchez, R. (2015). Estudio de disponibilidad léxica en aprendices italianos de español: análisis cuantitativo. En A. Gordejuela Senosiáin, D. Izquierdo Alegría, F.

- Jiménez Berrio, A. de Lucas Vicente y M. Casado Velarde (eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 429-441). Pamplona: Servicio de Publicaciones de la Universidad de Navarra. Recuperado de <http://dadun.unav.edu/bitstream/10171/40043/1/35.Rubio%20Sánchez.pdf>
- Rubio Sánchez, R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera. En F. Del Barrio de la Rosa (ed.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía* (pp. 143-161). Venezia: Edizioni Ca' Foscari. Recuperado de <http://edizionicafoscari.unive.it/media/pdf/chapter/978-88-6969-170-6/978-88-6969-170-6-ch-09.pdf>
- Rufat, A. (2018). Enseñanza-aprendizaje del componente léxico-semántico. En M^a. Martínez-Atienza y A. Zamorano Aguilar (eds.), *Teoría y metodología para la enseñanza de ELE. Vol. II. Enseñanza-aprendizaje de los componentes lingüísticos* (pp. 73-108). Madrid: enClave-ELE.
- Ruiz Basto, A. (1987). *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades, Plantel Naucalpan* (Tesis doctoral, Universidad Nacional Autónoma de México).
- Ruiz Martínez, A. (2006). REDES. Diccionario combinatorio del español contemporáneo como herramienta para la enseñanza de léxico en L2. *Linred*, 4. Recuperado de http://www.linred.es/articulos_pdf/LR_articulo_21092006.pdf
- Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.
- Samper Padilla, J. A. (1998). Criterios de edición del léxico disponible: sugerencias. *Lingüística*, 10, pp. 311-333.
- Samper Padilla, J. A., Bellón Fernández, J. J. y Samper Hernández, M. (2003). El proyecto de estudio de la disponibilidad léxica en español. En R. Ávila, J. A. Samper y H. Ueda (eds.), *Pautas y pistas en el análisis del léxico hispano (americano)* (pp. 27-140). Madrid: Vervuert-Iberoamericana.
- Samper Padilla, J. A. y Samper Hernández, M. (2006). Aportaciones recientes de los estudios de disponibilidad léxica. *Lynx: Panorámica de estudios lingüísticos*, 5, pp. 5-95.
- San Vicente, F. (dir.). Banco de datos [en línea]. *Contrastiva. Portal de gramática y de lingüística contrastiva español-italiano*. URL <http://www.contrastiva.it/wp/>
- Sánchez Lobato, J., Moreno García, C. y Santos Gargallo, I. (1999). *Español sin fronteras. Nivel avanzado*. Alcobendas: SGEL.
- Sánchez Lobato, J. y Aguirre, B. (1998). *Léxico fundamental del español (situaciones, temas y nociones)*. *Glosario multilingüe*. Madrid: Sociedad general española de

librería.

- Sánchez Pérez, A. (1992). *Historia de la enseñanza del español como lengua extranjera*. Madrid: SGEL.
- Sánchez-Saus Laserna, M. (2009). La variable “nivel de español” en el léxico disponible de los estudiantes de español como lengua extranjera. *Pragmalingüística*, 17, pp. 140-153. Recuperado de <https://revistas.uca.es/index.php/pragma/article/view/168/171>
- Sánchez-Saus Laserna, M. (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, Universidad de Cádiz). Recuperado de <http://hdl.handle.net/10498/15862>
- Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*. Sevilla: Editorial Universidad de Sevilla.
- Sánchez-Saus Laserna, M. (mayo, 2019). *La disponibilidad léxica en ELE: presente, pasado y retos de futuros*. Comunicación presentada en las II Jornadas de investigación en Disponibilidad léxica, Lleida, 17-18 de mayo de 2019.
- Santos Díaz, I. C. (2017). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum*, 27, pp. 122-139. Recuperado de http://www.ugr.es/~portalin/articulos/PL_numero27/9_Inmaculada.pdf
- Santos Gargallo, I. (2017 [1999]). *Lingüística aplicada a la enseñanza-aprendizaje del español como lengua extranjera*. Madrid: Arco/Libros, S. L.
- Schumann, J. (1978). The acculturation model for second language acquisition. En R. Gingras (ed.), *Second language acquisition research and foreign language teaching* (pp. 27-50). Arlington: Center for Applied Linguistics.
- Šifrar Kalan, M. (2000). ¿Cómo adquirir nuevo vocabulario y memorizarlo? En M^a. A. Martín Zorraquino y C. Díez Pelegrín (eds.), *¿Qué español enseñar?: norma y variación lingüísticas en la enseñanza del español a extranjeros. Actas del XI Congreso Internacional ASELE, Zaragoza 13-16 de septiembre de 2000* (pp. 903-914). Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=608523>
- Šifrar Kalan, M. (2009). Disponibilidad léxica en español lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia. *Verba hispanica*, 17, pp. 165-182. Recuperado de <https://revije.ff.uni-lj.si/VerbaHispanica/article/view/3651/3362>
- Šifrar Kalan, M. (2012). Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM). *MarcoELE*, 15. Recuperado de http://marcoele.com/descargas/15/sifrar-disponibilidad_lexica.pdf

- Šifrar Kalan, M. (2014a). Slovene Student's Lexical Availability in English and Spanish. En R. M^a. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 125-138). New York: Springer.
- Šifrar Kalan, M. (2014b). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, 41 (1), pp. 63-85. Recuperado de <https://pressto.amu.edu.pl/index.php/srp/article/view/36/23>
- Šifrar Kalan, M. (2016). La universalidad de los prototipos semánticos en el léxico disponible de español. *Verba hispanica*, 24, pp. 147-165. Recuperado de <https://revije.ff.uni-lj.si/VerbaHispanica/article/view/7255/6956>
- Šifrar Kalan, M. (2018). La influencia de estancia Erasmus en el léxico disponible de ELE: el caso de los centros de interés sobre España. En *International conference "Exploring the lexicon of bilingual and plurilingual learners: lexical availability and vocabulary acquisition"*, Universidad de La Rioja, Logroño 4-5 de octubre de 2018 (p. 44). Logroño: Universidad de La Rioja, Servicio de Publicaciones.
- Sinclair, J. (ed.). (1987). *Looking Up: An Account of the Cobuild Project in Lexical Computing*. London: Collins.
- Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language*. Oxford: Oxford University press.
- Sinclair, J. (2005). Corpus and text: Basic principles. En M. Wynne (ed.), *Developing Linguistic Corpora: a guide to good practice* (pp. 1-16). Oxford: Oxbow Books. Recuperado de <http://ota.ox.ac.uk/documents/creating/dlc/>
- Šišková, Z. (2012). Lexical Richness in EFL Students' Narratives. *Language Studies Working Papers*, 4, pp. 26-36. Recuperado de https://www.reading.ac.uk/web/files/english-language-and-literature/elal_LSWP_Vol_4_Siskova.pdf
- Skinner, B. (1957). *Verbal behavior*. New York: Appleton Century Crofts.
- Spreafico, L. (2003). Misurare le parole. Analisi lessicale quantitativa di un apprendente di italiano L2. *Linguistica e Filologia*, 17, pp. 93-125. Recuperado de <https://aisberg.unibg.it/retrieve/handle/10446/227/2102/LeF17%282003%29Spreafico.pdf>
- Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
- Tesitelová, J. (1992). *The main areas of quantitative linguistics*. New York: Planum Press.
- Thornbury, S. (2002). *How to teach vocabulary*. Essex: Longman.

- Thorndike, E. L. (1921). *The Teacher's Word Book*. New York: Teachers College, Columbia University.
- Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* (Tesis doctoral, Universidad de Salamanca). Recuperado de https://gedos.usal.es/bitstream/handle/10366/128287/DLE_Tom%E9CornejoC_L_%E9xicodisponible.pdf;jsessionid=755FBEBF157497F5050AF3397A3B3BC86?sequence=1
- Tomé Cornejo, C. (2018). Estrategias de recuperación del léxico disponible en español como lengua materna y como lengua extranjera. En *International conference "Exploring the lexicon of bilingual and plurilingual learners: lexical availability and vocabulary acquisition"*, Universidad de La Rioja, Logroño 4-5 de octubre de 2018 (pp. 42-43). Logroño: Universidad de La Rioja, Servicio de Publicaciones.
- Torrijano Pérez, A. (2004). *Errores de Aprendizaje, Aprendizaje de Los Errores*. Madrid: Arco/Libros, S. L.
- Torres González, A. N. (1999). Incidencia de las variables sociales en los índices producción léxica de estudiantes del último curso de la enseñanza no universitaria. En J. de Las Cuevas y D. Fasla Fernández (eds.), *Contribuciones al estudio de la lingüística aplicada* (pp. 393-401). Castellón: Asociación Española de Lingüística Aplicada.
- Torres González, A. N. (2003a). Riqueza léxica en textos narrativos escritos por estudiantes de Tenerife. En F. Moreno Fernández, F. Gimeno Menéndez, J. A. Samper, M^a. L. Gutiérrez Araus, M. Vaquero y C. Hernández Alonso (coords.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales* (pp. 435-449). Madrid, Arco/Libros, S. L.
- Torres González, A. N. (2003b). La producción léxica de estudiantes no universitarios de Tenerife. En C. Díaz Alayón, M. Morera y G. Ortega (eds.), *Estudios sobre el español de Canarias. Actas del I Congreso Internacional sobre Español de Canarias* (pp. 989-1009). Islas Canarias: Academia Canaria de la Lengua.
- Torruella Casañas, J. (2017). *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*. New York: Peter Lang.
- Trigo Ibáñez, E., Romero Oliva, M. F. y Santos Díaz, I. C. (2018). Elaboración de un corpus cacográfico desde la disponibilidad léxica en estudiantes sevillanos. Un análisis para la enseñanza de la lengua. *Revista de Lingüística y Lenguas Aplicadas*, 13, pp. 119-131. Recuperado de <https://polipapers.upv.es/index.php/rdlyla/article/view/9176/10285>
- Troitño, S. (2017). Implicaciones de crear materiales desde una perspectiva léxica. En F. Herrera (ed.), *Enseñar léxico en el aula de español: el poder de las palabras* (pp. 143-161). Barcelona: Difusión.

- Valencia, A. y Echeverría, M. S. (1999). *Disponibilidad léxica en estudiantes chilenos*. Santiago de Chile: Universidad de Chile y Universidad de Concepción.
- Van der Beke, G. E. (1935). *French Word Book*. New York: The Maximilian Company, Publications of the American and Canadian Committees on Modern Languages, vol. XV.
- Velázquez Puerto, K. (2018). *La enseñanza-aprendizaje de fraseología en ELE*. Madrid: Arco/Libros, S. L.
- VV.AA. (1993). *Viaje al Español. Versión internacional*. Madrid: Santillana.
- VV.AA. (1996). *Diccionario Salamanca de la lengua española*. Madrid: Santillana.
- VV.AA. (2011) [en línea]. *Diccionario de Aprendizaje del Español como Lengua Extranjera*. URL <http://www.iula.upf.edu/rec/dae>
- VV.AA. (2011). *Via rápida. Curso intensivo de español*. Barcelona: Difusión.
- VV.AA. (2013). *Español ELElab B2*. Salamanca: Ediciones Universidad de Salamanca.
- VV.AA. (2016). *Bitácora 2. Nueva edición*. Barcelona: Difusión.
- Vygotsky, L. S. (1962). *Language and Thought*. Cambridge: The MIT Press.
- Wang, X. (2016). *Riqueza léxica en muestras de lengua escrita de estudiantes sinohablantes de pregrado y postgrado* (Tesis de maestría inédita, Universidad de La Habana).
- West, M. (1953). *A General List of English Words*. London: Longman.
- Wilkins, D. (1972). *Linguistics in language teaching*. London: Arnold.
- Wolfe-Quintero, K., Inagaki, S. y Kim, H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Hawaii: University of Hawaii at Manoa.
- Young, R. (1999). Sociolinguistic Approaches to Second Language Acquisition. *Annual Review of Applied Linguistics*, 19, pp. 105-132.
- Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
- Zipf, G. K. (1935). *Psycho-biology of languages*. Cambridge: MIT Press.

ANEXOS

ANEXO I – TEST DE MEDICIÓN DE LA COMPETENCIA LÉXICA

Estimado alumno/a:

Mediante esta prueba se pretende conocer algunos aspectos relativos al desarrollo de su competencia léxica de ELE (Español como Lengua Extranjera) que nos serán muy útiles para nuestra investigación. Las respuestas son anónimas y no serán objeto de evaluación. Además, serán tratadas con discreción y usadas para fines específicamente científicos. Por ello, invitamos a que responda a todas las cuestiones con sinceridad y sin recelo.

Apartado 1a – Datos personales

(Son necesarios y muy importantes para este estudio, se utilizarán para informes estadísticos)

Sexo:

hombre

mujer

Edad: _____

Lengua materna: _____

Lugar de residencia habitual: _____ Provincia _____ (rural urbano)

Carrera universitaria:

Lingue, Letterature e Culture Moderne (II año III año)

Mediazione Linguistica e Culturale (II año III año)

Otro curso _____ (II año III año)

Años llevados estudiando ELE: _____

¿En su opinión qué nivel de ELE tiene? inicial intermedio avanzado

Otras lenguas extranjeras conocidas (además del español):

LE1 _____ Nivel inicial intermedio avanzado

LE2 _____ Nivel inicial intermedio avanzado

LE3 _____ Nivel inicial intermedio avanzado

LE4 _____ Nivel inicial intermedio avanzado

Apartado 1b – Opiniones sobre el aprendizaje de una LE

(Recuerde que no hay respuestas correctas o incorrectas, lo que nos interesa es su opinión)

¿Qué cree que es más importante conocer de una LE?

- gramática
- vocabulario
- gramática y vocabulario

¿Cuánto cree que es importante conocer el vocabulario de una LE?

- mucho
- bastante
- regular
- poco
- nada

¿Cómo cree que aprende más vocabulario?

- ejercicios de léxico
- lectura en español
- películas y canciones en español
- otros _____

El español me parece en general una lengua...

(Marque con una "x" el espacio que mejor represente su opinión)

	1	2	3	4	5	
útil						inútil
difícil						fácil
importante						irrelevante
agradable						desagradable

Apartado 2 – Encuesta de disponibilidad léxica

A continuación, escriba todas las palabras que se le ocurren en relación con los temas indicados en dos minutos para cada uno, si la ficha no es suficiente puede seguir en el reverso de cada folio. Por favor, escriba las palabras, aunque no esté seguro de la grafía correcta o del significado.

Apartado 3 – Encuesta de riqueza léxica

Escriba una redacción de 120 palabras sobre un viaje que le haya gustado.

Si quiere que le devolvamos corregida esta parte, indique aquí y solo aquí su nombre para garantizar que el resto de la prueba es anónima. Nombre y apellido: _____

¡Muchas gracias por su valiosa colaboración!

ANEXO II – LISTA DE PARADA

a	decimotercero	nos	si
al	decimotercio	nosotros	sí
algo	decimoséptimo	noveno	siete
alguien	decimosexto	noventa	sin
algún	decimotercero	nuestro	sino
alguno	decimotercio	nueve	sobre
alterno	del	o	suyo
aquel	diecinueve	ochenta	tal
aquello	dieciocho	ocho	talvez
aun	dieciocho	ochocientos	te
aún	dieciséis	octavo	tercer
aunque	diecisiete	once	tercero
bis	diez	onceno	ti
catorce	diezmilésimo	ora	trece
catorceno	diezmillonésimo	os	trecientos
centeno	diezmilmillonésimo	otro	treinta
centesimo	doble	para	tres
cien	doce	pero	trescientos
cienmilésimo	doceno	por	triple
cienmillonésimo	docientos	porque	tu
cienmilmillonésimo	dos	primer	tú
ciento	duodecimal	primero	tuyo
cinco	duodécimo	primo	u
cincuenta	e	propio	ultra
cinqueno	el	proprio	un
con	él	próximo	undécimo
conmigo	en	que	uno
conque	entre	qué	usted
consigo	ese	quien	veinte
contigo	eso	quién	veintésimo
contra	este	quienquier	veinticinco
cuádruple	esto	quienquiera	veinticuatro
cuarenta	excepto	quince	veintidós
cuarenteno	extra	quinceno	veintidoseno
cuarto	hacia	quinientos	veintinueve
cuaternario	ínter	quinto	veintiocho
cuatro	ínterin	se	veintiséis
cuatrocientos	mas	según	veintisiete
cuyo	más	segundo	veintitrés
de	me	seis	veintiún
décimo	mi	seiscientos	vigésimo
decimooctavo	mí	sendos	vos
decimocuarto	nadie	séptimo	vosotros
decimonono	ni	sesenta	vuestro
decimonoveno	ningún	setecientos	y
decimoquinto	ninguno	setenta	yo
decimoséptimo	no	séptimo	
decimosexto	nono	sexto	

ANEXO III – LISTADO GENERAL DE LÉXICO DISPONIBLE

(CORPUS GENERAL)

CI01, “las partes del cuerpo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	ojos	0,71077543	95	94,99999881	0,070162482
2	mano	0,68998682	91	91,00000262	0,137370753
3	cabeza	0,53324872	65	64,99999762	0,185376662
4	pie	0,52412736	81	81,00000024	0,245199409
5	pierna	0,47413734	69	68,99999976	0,296159527
6	brazo	0,4177554	63	62,99999952	0,342688331
7	boca	0,41181397	65	64,99999762	0,390694239
8	dedo	0,39431173	67	67,00000167	0,440177253
9	nariz	0,37995452	61	61,00000143	0,485228951
10	pelo	0,36899704	61	61,00000143	0,53028065
11	cara	0,33971396	45	44,99999881	0,56351551
12	espalda	0,26327646	50	50	0,600443131
13	orejas	0,22095841	45	44,99999881	0,633677991
14	cueillo	0,21489586	34	34,00000036	0,658788774
15	rodilla	0,21331339	39	38,99999857	0,687592319
16	uñas	0,20997067	42	41,99999869	0,718611521
17	dientes	0,17093405	35	34,9999994	0,744460857
18	lengua	0,16308302	29	28,99999917	0,765878877
19	barriga	0,14163156	24	23,99999946	0,783604136
20	labios	0,09925174	19	18,99999976	0,797636632
21	corazón	0,09316909	20	20,0000003	0,812407681
22	hombro	0,09187716	18	18,00000072	0,825701625
23	estómago	0,08756647	20	20,0000003	0,840472674
24	piel	0,08052495	17	17,00000018	0,853028065
25	garganta	0,07312658	14	14,00000006	0,863367799
26	cejas	0,06678414	16	15,99999964	0,875184638
27	pecho	0,06640178	12	11,99999973	0,884047267
28	cabellos	0,06044529	10	10,00000015	0,891432792
29	codos	0,05182045	10	10,00000015	0,898818316
30	tobillo	0,04391922	8	7,999999821	0,904726736
31	pestañas	0,03943514	8	7,999999821	0,910635155
32	mejilla	0,0381039	7	7,00000003	0,915805022
33	cintura	0,03574285	9	9,000000358	0,922451994
34	frente	0,02417241	8	7,999999821	0,928360414

35	naso	0,01971822	3	2,999999933	0,930576071
36	músculo	0,01701428	4	3,999999911	0,933530281
37	pulso	0,01629602	4	3,999999911	0,93648449
38	cadera	0,01580146	5	5,000000075	0,940177253
39	tripa	0,01542782	3	2,999999933	0,94239291
40	muñeca	0,01443267	4	3,999999911	0,94534712
41	bigote	0,0128414	3	2,999999933	0,947562777
42	cerebro	0,0128209	4	3,999999911	0,950516987
43	sangre	0,01127292	3	2,999999933	0,952732644
44	nalgas	0,01082335	3	2,999999933	0,954948301
45	culo	0,01081932	4	3,999999911	0,957902511
46	rubio	0,01073254	2	1,999999955	0,959379616
47	panza	0,01036588	2	1,999999955	0,960856721
48	tórax	0,0092114	1	0,999999978	0,961595273
49	cuerpo-humano	0,00848499	1	0,999999978	0,962333826
50	tez	0,00781587	1	0,999999978	0,963072378
51	oído	0,00781587	1	0,999999978	0,963810931
52	coronilla	0,00781587	1	0,999999978	0,964549483
53	vientre	0,00721757	2	1,999999955	0,966026588
54	hígado	0,00719952	1	0,999999978	0,96676514
55	pelirrojo	0,00719952	1	0,999999978	0,967503693
56	oír	0,00663176	1	0,999999978	0,968242245
57	seno	0,00649789	2	1,999999955	0,96971935
58	marrón	0,00562705	1	0,999999978	0,970457903
59	cuerpo	0,00562705	1	0,999999978	0,971196455
60	muslo	0,00537332	2	1,999999955	0,97267356
61	azul	0,0051833	1	0,999999978	0,973412112
62	tetas	0,00516145	2	1,999999955	0,97562777
63	sobrecejos	0,00501653	2	1,999999955	0,977104874
64	vagina	0,00440906	2	1,999999955	0,978581979
65	testa	0,00439803	1	0,999999978	0,979320532
66	pene	0,00437949	2	1,999999955	0,980797637
67	salud	0,00425479	2	1,999999955	0,982274742
68	feo	0,0040512	1	0,999999978	0,983013294
69	genitales	0,0040512	1	0,999999978	0,983751846
70	gordo	0,00373173	1	0,999999978	0,984490399
71	respirar	0,00343744	1	0,999999978	0,985228951
72	sien	0,00343744	1	0,999999978	0,985967504
73	carne	0,00343744	1	0,999999978	0,986706056
74	pulmones	0,00343744	1	0,999999978	0,987444609

75	muelas	0,00343744	1	0,999999978	0,988183161
76	intestino	0,00343744	1	0,999999978	0,988921713
77	aparato	0,00343744	1	0,999999978	0,989660266
78	trasero	0,00291667	1	0,999999978	0,990398818
79	hueso	0,00247479	1	0,999999978	0,991137371
80	neurona	0,00247479	1	0,999999978	0,991875923
81	tacto	0,00247479	1	0,999999978	0,992614476
82	pezón	0,00227963	1	0,999999978	0,993353028
83	mentón	0,00193427	1	0,999999978	0,994091581
84	omblogo	0,00193427	1	0,999999978	0,994830133
85	talón	0,0015118	1	0,999999978	0,995568685
86	verde	0,00139258	1	0,999999978	0,996307238
87	guapo	0,00128276	1	0,999999978	0,99704579
88	bajo	0,00128276	1	0,999999978	0,997784343
89	vista	0,0011816	1	0,999999978	0,998522895
90	alto	0,0011816	1	0,999999978	0,999261448
91	ver	0,00100259	1	0,999999978	1

CI02, “la ropa”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	camiseta	0,481934398	70	62,5	0,088495575
2	zapatos	0,439514309	71	63,3928597	0,178255373
3	pantalones	0,409810811	69	61,6071403	0,265486726
4	camisa	0,28956607	44	39,28571343	0,321112516
5	jersey	0,259401023	45	40,17857015	0,378002528
6	chaqueta	0,255687803	47	41,96428657	0,437420986
7	falda	0,252731293	42	37,5	0,490518331
8	vaqueros	0,240881488	41	36,60714328	0,542351454
9	bufanda	0,182538986	34	30,35714328	0,585335019
10	calcetines	0,14323926	29	25,89285672	0,621997472
11	vestido	0,136809915	30	26,78571343	0,659924147
12	abrigo	0,126946315	23	20,53571492	0,689001264
13	traje	0,097627945	17	15,17857164	0,710493047
14	tacones	0,068812624	14	12,5	0,728192162
15	minifalda	0,062010843	11	9,821428359	0,742098609
16	blusa	0,061510492	11	9,821428359	0,756005057
17	bolso	0,058352523	11	9,821428359	0,769911504
18	sudadera	0,057073303	12	10,71428582	0,785082174
19	gafas	0,055289257	11	9,821428359	0,798988622
20	zapatillas	0,041901674	11	9,821428359	0,81289507
21	cinturón	0,039425083	10	8,928571641	0,825537295
22	sombrero	0,038218297	10	8,928571641	0,83817952
23	botas	0,037500799	8	7,142857462	0,8482933
24	guantes	0,03516937	9	8,035714179	0,859671302
25	gorro	0,033433791	10	8,928571641	0,872313527
26	medias	0,030320408	6	5,35714291	0,879898862
27	reloj	0,026498398	6	5,35714291	0,887484197
28	sujetador	0,022082914	8	7,142857462	0,897597977
29	pijama	0,021087589	4	3,571428731	0,902654867
30	corbata	0,020279812	3	2,678571455	0,906447535
31	bañador	0,019839035	5	4,464285821	0,912768647
32	calzoncillos	0,017906409	4	3,571428731	0,917825537
33	pulsera	0,015028615	3	2,678571455	0,921618205
34	anillo	0,01486198	5	4,464285821	0,927939317
35	traje-de-baño	0,012797104	3	2,678571455	0,931731985
36	bikini	0,012454879	4	3,571428731	0,936788875
37	chanclas	0,012055307	3	2,678571455	0,940581542
38	collar	0,012042059	3	2,678571455	0,94437421
39	ponerse	0,012008871	2	1,785714366	0,946902655
40	remera	0,011469782	3	2,678571455	0,950695322

41	mochila	0,009413114	2	1,785714366	0,953223767
42	maleta	0,008928572	1	0,892857183	0,95448799
43	chaleco	0,008928572	1	0,892857183	0,955752212
44	saco	0,008928572	1	0,892857183	0,957016435
45	rebeca	0,007965708	2	1,785714366	0,95954488
46	pañuelo	0,007798741	1	0,892857183	0,960809102
47	ropa-interior	0,007564336	2	1,785714366	0,963337547
48	polera	0,00681188	1	0,892857183	0,96460177
49	vestirse	0,005949898	1	0,892857183	0,965865992
50	anorak	0,005949898	1	0,892857183	0,967130215
51	algodón	0,005332824	2	1,785714366	0,96965866
52	bolsillo	0,005196992	1	0,892857183	0,970922882
53	calzados	0,005196992	1	0,892857183	0,972187105
54	chándal	0,005196992	1	0,892857183	0,973451327
55	lente-de-contacto	0,005196992	1	0,892857183	0,97471555
56	bragas	0,005196992	1	0,892857183	0,975979772
57	seda	0,004658003	2	1,785714366	0,978508217
58	quitarse	0,004539359	1	0,892857183	0,97977244
59	sandalias	0,004180116	2	1,785714366	0,982300885
60	playera	0,003964944	1	0,892857183	0,983565107
61	mantilla	0,003964944	1	0,892857183	0,98482933
62	zapata	0,003964944	1	0,892857183	0,986093552
63	gafas-de-sol	0,00310404	2	1,785714366	0,988621997
64	pendientes	0,003024977	1	0,892857183	0,98988622
65	ropa	0,003024977	1	0,892857183	0,991150442
66	bajero	0,003024977	1	0,892857183	0,992414665
67	zapatilla-de-deporte	0,003024977	1	0,892857183	0,993678887
68	zapatero	0,002307847	1	0,892857183	0,99494311
69	calzas	0,002307847	1	0,892857183	0,996207332
70	tenis	0,00201581	1	0,892857183	0,997471555
71	top	0,001760727	1	0,892857183	0,998735777
72	chancletas	0,001760727	1	0,892857183	1

CI03, “partes de la casa (sin muebles)”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	cocina	0,680141151	94	83,92857313	0,154098361
2	cuarto-de-baño	0,614644945	91	81,25	0,303278689
3	habitación	0,335915476	49	43,75	0,383606557
4	salón	0,265763253	48	42,85714328	0,462295082
5	jardín	0,226448298	45	40,17857015	0,536065574
6	cuarto	0,174641341	27	24,10714328	0,580327869
7	garaje	0,14945288	32	28,57142985	0,632786885
8	escalera	0,094394602	19	16,96428508	0,663934426
9	techo	0,079044521	16	14,28571492	0,690163934
10	puerta	0,076693319	15	13,39285672	0,714754098
11	ventana	0,073338203	16	14,28571492	0,740983607
12	patio	0,062821046	13	11,60714254	0,762295082
13	dormitorio	0,06136163	10	8,928571641	0,778688525
14	comedor	0,058273014	12	10,71428582	0,798360656
15	estudio	0,054905318	13	11,60714254	0,819672131
16	piso	0,046888702	10	8,928571641	0,836065574
17	pasillo	0,043551773	9	8,035714179	0,850819672
18	terrazza	0,042107698	12	10,71428582	0,870491803
19	entrada	0,04129323	11	9,821428359	0,88852459
20	cámara	0,041195538	7	6,25	0,9
21	cuarto-de-estar	0,040128678	8	7,142857462	0,913114754
22	balcón	0,031048205	9	8,035714179	0,927868852
23	pared	0,028426496	4	3,571428731	0,93442623
24	oficina	0,021971243	4	3,571428731	0,940983607
25	despacho	0,016280396	3	2,678571455	0,945901639
26	ingreso	0,010459727	3	2,678571455	0,950819672
27	aseo	0,010068518	2	1,785714366	0,954098361
28	sótano	0,00958765	3	2,678571455	0,959016393
29	corredor	0,008086188	2	1,785714366	0,962295082
30	cantina	0,007488422	2	1,785714366	0,96557377
31	altillo	0,007480752	1	0,892857183	0,967213115
32	mansarda	0,006987588	2	1,785714366	0,970491803
33	muro	0,006987588	2	1,785714366	0,973770492
34	elevador	0,006567967	2	1,785714366	0,97704918
35	cuarto-trasero	0,006274131	2	1,785714366	0,980327869
36	atelier	0,00525136	1	0,892857183	0,981967213
37	cuarto-matrimonial	0,00525136	1	0,892857183	0,983606557
38	portal	0,004399822	1	0,892857183	0,985245902
39	sala	0,004399822	1	0,892857183	0,986885246

40	librería	0,003686365	1	0,892857183	0,98852459
41	pórtico	0,0030886	1	0,892857183	0,990163934
42	escritorio	0,0030886	1	0,892857183	0,991803279
43	piscina	0,0030886	1	0,892857183	0,993442623
44	sobresuelo	0,0030886	1	0,892857183	0,995081967
45	salón-de-baile	0,002587766	1	0,892857183	0,996721311
46	lavandería	0,002168145	1	0,892857183	0,998360656
47	cueva	0,001275199	1	0,892857183	1

CI04, “los muebles de la casa”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	cama	0,510855	76	67,85714	0,115152
2	silla	0,458743	75	66,96429	0,228788
3	mesa	0,440463	73	65,17857	0,339394
4	sofá	0,345038	60	53,57143	0,430303
5	armario	0,313539	54	48,21429	0,512121
6	sillón	0,154345	30	26,78571	0,557576
7	ducha	0,121397	25	22,32143	0,595455
8	lámpara	0,107842	24	21,42857	0,631818
9	tele(visión)	0,106236	20	17,85714	0,662121
10	mesilla-de-noche	0,101045	22	19,64286	0,695455
11	lavadora	0,075863	15	13,39286	0,718182
12	horno	0,075149	18	16,07143	0,745455
13	frigorífico	0,067602	12	10,71429	0,763636
14	espejo	0,062708	13	11,60714	0,783333
15	cajón	0,060212	13	11,60714	0,80303
16	escritorio	0,059886	14	12,5	0,824242
17	nevera	0,053642	13	11,60714	0,843939
18	ventana	0,053459	11	9,821428	0,860606
19	puerta	0,050938	11	9,821428	0,877273
20	estantería	0,049966	9	8,035714	0,890909
21	lavavajillas	0,037307	8	7,142857	0,90303
22	cocina	0,031606	4	3,571429	0,909091
23	guardarropa	0,030218	6	5,357143	0,918182
24	alfombra	0,026155	8	7,142857	0,930303
25	mueble	0,026061	4	3,571429	0,936364
26	refrigerador	0,019333	5	4,464286	0,943939
27	diván	0,01794	4	3,571429	0,95
28	ropero	0,017857	2	1,785714	0,95303
29	reloj	0,016985	4	3,571429	0,959091
30	grifo	0,011519	2	1,785714	0,962121
31	cuadro	0,010759	4	3,571429	0,968182
32	lavabo	0,009954	2	1,785714	0,971212
33	librería	0,009651	2	1,785714	0,974242
34	papelera	0,008929	1	0,892857	0,975758
35	caja	0,007481	1	0,892857	0,977273
36	cortina	0,007481	1	0,892857	0,978788
37	almohada	0,007481	1	0,892857	0,980303
38	radio	0,00632	2	1,785714	0,983333
39	flexo	0,006268	1	0,892857	0,984848

40	retrato	0,0044	1	0,892857	0,986364
41	despensa	0,0044	1	0,892857	0,987879
42	baño	0,003686	1	0,892857	0,989394
43	escalera	0,003089	1	0,892857	0,990909
44	hielera	0,003089	1	0,892857	0,992424
45	cenicero	0,003089	1	0,892857	0,993939
46	bañera	0,003089	1	0,892857	0,995455
47	banco	0,002168	1	0,892857	0,99697
48	despertador	0,002168	1	0,892857	0,998485
49	luz	0,001275	1	0,892857	1

CI05, “alimentos y bebidas”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	agua	0,658388	91	81,25	0,062201
2	cerveza	0,282548	47	41,96429	0,094327
3	tomate	0,274992	50	44,64286	0,128503
4	leche	0,266612	49	43,75	0,161996
5	manzana	0,259139	49	43,75	0,195489
6	naranja	0,241581	47	41,96429	0,227614
7	vino	0,24004	37	33,03571	0,252905
8	paella	0,23787	37	33,03571	0,278195
9	jamón	0,227918	39	34,82143	0,304853
10	pescado	0,225759	44	39,28571	0,334928
11	pan	0,223488	40	35,71429	0,362269
12	arroz	0,191119	35	31,25	0,386193
13	carne	0,182782	34	30,35714	0,409433
14	zumos	0,176995	30	26,78571	0,429938
15	pasta	0,170736	31	27,67857	0,451128
16	café	0,139737	26	23,21429	0,4689
17	bocadillo	0,134283	21	18,75	0,483254
18	azúcar	0,1319	30	26,78571	0,503759
19	huevo	0,125256	26	23,21429	0,521531
20	queso	0,121596	25	22,32143	0,538619
21	patata	0,109742	22	19,64286	0,553657
22	chocolate	0,10957	24	21,42857	0,570062
23	sangría	0,105035	19	16,96429	0,583049
24	plátano	0,104056	20	17,85714	0,596719
25	sal	0,102587	27	24,10714	0,615174
26	aceite	0,092777	24	21,42857	0,631579
27	tapas	0,09256	17	15,17857	0,643199
28	zanahoria	0,090186	20	17,85714	0,656869
29	limón	0,084624	20	17,85714	0,67054
30	gazpacho	0,076027	13	11,60714	0,679426
31	cebolla	0,073468	17	15,17857	0,691046
32	tortilla	0,071329	14	12,5	0,700615
33	pollo	0,066688	13	11,60714	0,709501
34	fresa	0,06035	13	11,60714	0,718387
35	galleta	0,060103	12	10,71429	0,726589
36	mantequilla	0,059106	11	9,821428	0,734108
37	ensalada	0,055134	12	10,71429	0,74231
38	vino-tinto	0,049694	8	7,142857	0,747779

39	harina	0,044054	10	8,928572	0,754614
40	<i>Coca-Cola</i>	0,043163	7	6,25	0,759398
41	melocotón	0,042864	9	8,035714	0,76555
42	pimiento	0,042638	9	8,035714	0,771702
43	fruta	0,041837	9	8,035714	0,777854
44	té	0,041666	8	7,142857	0,783322
45	tarta	0,039138	10	8,928572	0,790157
46	dulce	0,03912	8	7,142857	0,795625
47	churros	0,037384	7	6,25	0,80041
48	ajo	0,035873	9	8,035714	0,806562
49	calabacín	0,034239	7	6,25	0,811347
50	vino-blanco	0,033684	5	4,464286	0,814764
51	mariscos	0,032217	9	8,035714	0,820916
52	ternera	0,031778	7	6,25	0,825701
53	cereza	0,028818	6	5,357143	0,829802
54	lechuga	0,028614	6	5,357143	0,833903
55	vinagre	0,028282	7	6,25	0,838688
56	frijoles	0,026118	5	4,464286	0,842105
57	tequila	0,02576	4	3,571429	0,844839
58	berenjena	0,024784	5	4,464286	0,848257
59	chorizo	0,02471	5	4,464286	0,851675
60	aguacate	0,024287	5	4,464286	0,855092
61	verdura	0,023299	6	5,357143	0,859193
62	pimienta	0,023108	6	5,357143	0,863295
63	papa	0,023037	4	3,571429	0,866029
64	empanada	0,022933	3	2,678571	0,868079
65	piña	0,022331	6	5,357143	0,87218
66	agua-con-gas	0,020574	3	2,678571	0,874231
67	aceituna	0,020091	5	4,464286	0,877649
68	legumbres	0,0196	3	2,678571	0,879699
69	pepino	0,019513	4	3,571429	0,882433
70	cerdo	0,018939	5	4,464286	0,885851
71	banana	0,017701	4	3,571429	0,888585
72	zumodefruta	0,016647	3	2,678571	0,890636
73	zumode-naranja	0,016414	3	2,678571	0,892686
74	frito	0,016011	3	2,678571	0,894737
75	mejillones	0,015789	4	3,571429	0,897471
76	sopa	0,015508	4	3,571429	0,900205
77	ananás	0,014559	3	2,678571	0,902256
78	perejil	0,014445	4	3,571429	0,90499
79	refresco	0,014277	3	2,678571	0,90704

80	garbanzos	0,01424	3	2,678571	0,909091
81	calabaza	0,014096	3	2,678571	0,911141
82	hamburguesa	0,013953	2	1,785714	0,912509
83	dulce-de-leche	0,012754	3	2,678571	0,914559
84	atún	0,011862	3	2,678571	0,91661
85	hierbabuena	0,011839	2	1,785714	0,917977
86	gamba	0,01171	3	2,678571	0,920027
87	agua-sin-gas	0,011556	2	1,785714	0,921394
88	melón	0,011503	2	1,785714	0,922761
89	judías	0,010524	3	2,678571	0,924812
90	ron	0,01031	2	1,785714	0,926179
91	cereales	0,010048	2	1,785714	0,927546
92	aguardiente	0,010045	2	1,785714	0,928913
93	mojito	0,009652	2	1,785714	0,93028
94	caldo	0,009581	3	2,678571	0,932331
95	helado	0,009408	3	2,678571	0,934381
96	alfajor	0,009377	2	1,785714	0,935748
97	alcachofa	0,009377	2	1,785714	0,937116
98	dulce	0,008891	2	1,785714	0,938483
99	salmón	0,008525	2	1,785714	0,93985
100	pulpo	0,008245	2	1,785714	0,941217
101	azafrán	0,008224	1	0,892857	0,9419
102	espagueti	0,008224	1	0,892857	0,942584
103	conejo	0,00788	2	1,785714	0,943951
104	chupito	0,007576	1	0,892857	0,944634
105	burrito	0,007576	1	0,892857	0,945318
106	tinto-de-verano	0,007576	1	0,892857	0,946001
107	hielo	0,007576	1	0,892857	0,946685
108	nata	0,007576	1	0,892857	0,947368
109	pimentón	0,007276	3	2,678571	0,949419
110	vaca	0,007271	2	1,785714	0,950786
111	pera	0,007234	2	1,785714	0,952153
112	postre	0,006978	1	0,892857	0,952837
113	ciruela	0,006978	1	0,892857	0,95352
114	animal	0,006978	1	0,892857	0,954204
115	caña	0,006978	1	0,892857	0,954887
116	quesadilla	0,006978	1	0,892857	0,955571
117	sandía	0,006978	1	0,892857	0,956254
118	taco	0,006978	1	0,892857	0,956938
119	cordero	0,006978	1	0,892857	0,957621
120	calimocho	0,006978	1	0,892857	0,958305

121	jugo	0,006428	1	0,892857	0,958988
122	salchicha	0,006159	2	1,785714	0,960355
123	miel	0,006138	2	1,785714	0,961722
124	lentejas	0,006016	2	1,785714	0,96309
125	palta	0,005921	1	0,892857	0,963773
126	albaricoque	0,005921	1	0,892857	0,964457
127	yogur	0,005896	2	1,785714	0,965824
128	albóndiga	0,005454	1	0,892857	0,966507
129	macarrones	0,005454	1	0,892857	0,967191
130	chontaduro	0,005454	1	0,892857	0,967874
131	naranjada	0,005454	1	0,892857	0,968558
132	licor	0,005454	1	0,892857	0,969241
133	gaseoso	0,005454	1	0,892857	0,969925
134	pechuga-de-pollo	0,005454	1	0,892857	0,970608
135	alcohol	0,005024	1	0,892857	0,971292
136	puerco	0,005024	1	0,892857	0,971975
137	bocata	0,005024	1	0,892857	0,972659
138	camarón	0,004899	2	1,785714	0,974026
139	cóctel	0,004628	1	0,892857	0,97471
140	mango	0,004628	1	0,892857	0,975393
141	ajillo	0,004628	1	0,892857	0,976077
142	arroz-con-leche	0,004263	1	0,892857	0,97676
143	crema	0,004263	1	0,892857	0,977444
144	manteca	0,004263	1	0,892857	0,978127
145	mermelada	0,004071	2	1,785714	0,979494
146	espinaca	0,003927	1	0,892857	0,980178
147	flambeado	0,003927	1	0,892857	0,980861
148	seco	0,003927	1	0,892857	0,981545
149	salado	0,003927	1	0,892857	0,982228
150	uva	0,003927	1	0,892857	0,982912
151	cidra	0,003617	1	0,892857	0,983595
152	asado	0,003617	1	0,892857	0,984279
153	cangrejo	0,003617	1	0,892857	0,984962
154	piña-colada	0,003332	1	0,892857	0,985646
155	montadito	0,003332	1	0,892857	0,986329
156	castaña	0,003332	1	0,892857	0,987013
157	langosta	0,003332	1	0,892857	0,987697
158	solomillo	0,003069	1	0,892857	0,98838
159	apio	0,003069	1	0,892857	0,989064
160	tostada	0,002827	1	0,892857	0,989747

161	amargo	0,002827	1	0,892857	0,990431
162	orégano	0,002827	1	0,892857	0,991114
163	bizcocho	0,002604	1	0,892857	0,991798
164	parrilla	0,002604	1	0,892857	0,992481
165	espárrago	0,002604	1	0,892857	0,993165
166	pavo	0,002399	1	0,892857	0,993848
167	almeja	0,002399	1	0,892857	0,994532
168	embutido	0,00221	1	0,892857	0,995215
169	vegetal	0,00221	1	0,892857	0,995899
170	torrezno	0,00221	1	0,892857	0,996582
171	levadura	0,002035	1	0,892857	0,997266
172	coliflor	0,001875	1	0,892857	0,997949
173	crema-catalana	0,001875	1	0,892857	0,998633
174	bombón	0,001465	1	0,892857	0,999316
175	caballo	0,001243	1	0,892857	1

CI06, “objetos colocados en la mesa para la comida”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	plato	0,692113	94	83,92857	0,177358
2	vaso	0,493172	79	70,53571	0,326415
3	cuchillo	0,453779	68	60,71429	0,454717
4	cuchara	0,349408	59	52,67857	0,566038
5	botella	0,280678	55	49,10714	0,669811
6	tenedor	0,221298	35	31,25	0,735849
7	cucharilla	0,098805	18	16,07143	0,769811
8	servilleta	0,060212	15	13,39286	0,798113
9	aceite	0,038536	10	8,928572	0,816981
10	sal	0,036828	9	8,035714	0,833962
11	mantel	0,034716	9	8,035714	0,850943
12	copa	0,02677	4	3,571429	0,858491
13	taza	0,025734	7	6,25	0,871698
14	ensaladera	0,025116	7	6,25	0,884906
15	jarra	0,022015	6	5,357143	0,896226
16	plato-hondo	0,020105	5	4,464286	0,90566
17	plato-llano	0,020052	5	4,464286	0,915094
18	cucharada	0,020038	3	2,678571	0,920755
19	agua	0,01847	4	3,571429	0,928302
20	sartén	0,015544	3	2,678571	0,933962
21	bebida	0,013658	3	2,678571	0,939623
22	vinagre	0,010019	4	3,571429	0,94717
23	olla	0,008902	3	2,678571	0,95283
24	cuenca	0,007685	3	2,678571	0,958491
25	pan	0,006981	2	1,785714	0,962264
26	comedor	0,006086	1	0,892857	0,964151
27	mesa	0,006086	1	0,892857	0,966038
28	tomar	0,006086	1	0,892857	0,967925
29	pimienta	0,005232	2	1,785714	0,971698
30	pulir	0,005024	1	0,892857	0,973585
31	cubierto	0,005024	1	0,892857	0,975472
32	cacerola	0,005024	1	0,892857	0,977358
33	papel	0,004754	2	1,785714	0,981132
34	cambiar	0,004148	1	0,892857	0,983019
35	alimento	0,004148	1	0,892857	0,984906
36	cucharón	0,003424	1	0,892857	0,986792
37	comida	0,003424	1	0,892857	0,988679
38	comer	0,002827	1	0,892857	0,990566
39	bandeja	0,002827	1	0,892857	0,992453
40	desayuno	0,002334	1	0,892857	0,99434

41	cazo	0,002334	1	0,892857	0,996226
42	vino-tinto	0,001927	1	0,892857	0,998113
43	vino-blanco	0,001591	1	0,892857	1

CI07, “la cocina y sus utensilios”

N°	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	horno	0,389606	53	47,32142985	0,145205479
2	plato	0,172674	24	21,42857164	0,210958904
3	cuchillo	0,155303	27	24,10714328	0,284931507
4	cuchara	0,135013	24	21,42857164	0,350684932
5	sartén	0,110797	17	15,17857164	0,397260274
6	vaso	0,109405	16	14,28571492	0,44109589
7	frigorífico	0,108946	19	16,96428508	0,493150685
8	nevera	0,10836	18	16,07142836	0,542465753
9	lavavajillas	0,090867	15	13,39285672	0,583561644
10	olla	0,084341	15	13,39285672	0,624657534
11	mesa	0,063249	11	9,821428359	0,654794521
12	(horno)microondas	0,056066	9	8,035714179	0,679452055
13	refrigerador	0,04216	6	5,35714291	0,695890411
14	batidor	0,041419	8	7,142857462	0,717808219
15	cuenca	0,040143	6	5,35714291	0,734246575
16	tenedor	0,036962	8	7,142857462	0,756164384
17	ensaladera	0,033472	7	6,25	0,775342466
18	fuego	0,03092	6	5,35714291	0,791780822
19	botella	0,030165	7	6,25	0,810958904
20	bandeja	0,029133	6	5,35714291	0,82739726
21	grifo	0,024003	4	3,571428731	0,838356164
22	paella	0,020767	4	3,571428731	0,849315068
23	cazuela	0,018123	3	2,678571455	0,857534247
24	cucharilla	0,0144	3	2,678571455	0,865753425
25	cazo	0,012022	3	2,678571455	0,873972603
26	cacerola	0,011273	2	1,785714366	0,879452055
27	mezclar	0,010652	2	1,785714366	0,884931507
28	cocinar	0,009921	2	1,785714366	0,890410959
29	bol	0,008929	1	0,892857183	0,893150685
30	envase	0,008929	1	0,892857183	0,895890411
31	taza	0,007421	2	1,785714366	0,901369863
32	guantes	0,007094	1	0,892857183	0,904109589
33	cocina	0,007094	1	0,892857183	0,906849315
34	comedor	0,007094	1	0,892857183	0,909589041
35	congelador	0,007094	1	0,892857183	0,912328767
36	recipiente	0,007094	1	0,892857183	0,915068493

37	enharinar	0,007094	1	0,892857183	0,917808219
38	rodillo	0,007094	1	0,892857183	0,920547945
39	papel	0,005636	1	0,892857183	0,923287671
40	apagar	0,005636	1	0,892857183	0,926027397
41	espumadera	0,005636	1	0,892857183	0,928767123
42	estufa	0,005636	1	0,892857183	0,931506849
43	placa	0,005636	1	0,892857183	0,934246575
44	grasera	0,005636	1	0,892857183	0,936986301
45	balanza	0,005636	1	0,892857183	0,939726027
46	espátula	0,005636	1	0,892857183	0,942465753
47	abrir	0,004478	1	0,892857183	0,945205479
48	lavaplatos	0,004478	1	0,892857183	0,947945205
49	escurridor	0,004478	1	0,892857183	0,950684932
50	balda	0,004478	1	0,892857183	0,953424658
51	trapo	0,003722	2	1,785714366	0,95890411
52	comida	0,003558	1	0,892857183	0,961643836
53	esponja	0,003558	1	0,892857183	0,964383562
54	libro-de-cocina	0,003558	1	0,892857183	0,967123288
55	jabón	0,003558	1	0,892857183	0,969863014
56	cubierto	0,003558	1	0,892857183	0,97260274
57	despensa	0,002827	1	0,892857183	0,975342466
58	hervidor	0,002827	1	0,892857183	0,978082192
59	molde	0,002827	1	0,892857183	0,980821918
60	agua	0,002246	1	0,892857183	0,983561644
61	jarra	0,002246	1	0,892857183	0,98630137
62	cafetera	0,002246	1	0,892857183	0,989041096
63	escudilla	0,001785	1	0,892857183	0,991780822
64	tostador	0,001785	1	0,892857183	0,994520548
65	licuadora	0,001418	1	0,892857183	0,997260274
66	carbón	0,001127	1	0,892857183	1

CI08, “la escuela: muebles y materiales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	libro	0,388528198	76	67,8571403	0,094645081
2	mesa	0,384773791	59	52,67857313	0,168119552
3	silla	0,379523367	62	55,3571403	0,245330012
4	bolí(grafo)	0,36574918	64	57,1428597	0,325031133
5	lápiz	0,348755896	58	51,78571343	0,397260274
6	cuaderno	0,344558567	67	59,82142687	0,480697385
7	pizarra	0,307754308	56	50	0,550435866
8	mochila	0,195003152	38	33,92857015	0,597758406
9	ordenador	0,178551197	38	33,92857015	0,645080946
10	papel	0,140372336	30	26,78571343	0,682440847
11	borrador	0,116734877	24	21,42857164	0,712328767
12	rotulador	0,069617681	16	14,28571492	0,732254047
13	sacapuntas	0,067909956	15	13,39285672	0,750933998
14	estuche	0,057638772	16	14,28571492	0,770859278
15	bolso	0,045729965	9	8,035714179	0,782067248
16	goma	0,044203345	10	8,928571641	0,794520548
17	pupitre	0,040602017	6	5,35714291	0,801992528
18	escritorio	0,040309761	6	5,35714291	0,809464508
19	hoja	0,039687119	10	8,928571641	0,821917808
20	regla	0,038989238	11	9,821428359	0,835616438
21	tijeras	0,037150998	8	7,142857462	0,845579078
22	tiza	0,036431495	9	8,035714179	0,856787049
23	color	0,036044519	8	7,142857462	0,866749689
24	papelera	0,031376302	5	4,464285821	0,872976339
25	cátedra	0,029776724	5	4,464285821	0,879202989
26	diccionario	0,029421568	8	7,142857462	0,889165629
27	marcador	0,027966425	5	4,464285821	0,895392279
28	armario	0,025492486	7	6,25	0,904109589
29	clase	0,02455434	5	4,464285821	0,910336239
30	computador	0,021984624	4	3,571428731	0,915317559
31	ventana	0,021129509	4	3,571428731	0,920298879
32	pastel	0,01851481	3	2,678571455	0,924034869
33	calculador	0,01681616	3	2,678571455	0,927770859
34	mapa	0,014010834	4	3,571428731	0,932752179
35	profesor	0,013524048	3	2,678571455	0,936488169
36	pluma	0,01307642	2	1,785714366	0,938978829
37	aula	0,011832476	3	2,678571455	0,942714819

38	puerta	0,009905555	2	1,785714366	0,945205479
39	apuntes	0,009584014	3	2,678571455	0,948941469
40	reloj	0,00900836	3	2,678571455	0,95267746
41	ir	0,008928572	1	0,892857183	0,95392279
42	perchero	0,008254929	2	1,785714366	0,95641345
43	alumno	0,007887587	2	1,785714366	0,95890411
44	tabla	0,007659319	1	0,892857183	0,96014944
45	volver	0,007659319	1	0,892857183	0,96139477
46	edificio	0,006570499	1	0,892857183	0,9626401
47	celo	0,006570499	1	0,892857183	0,96388543
48	agenda	0,005908215	3	2,678571455	0,96762142
49	goma-de-borrar	0,005236938	2	1,785714366	0,97011208
50	fotocopia	0,005236938	2	1,785714366	0,97260274
51	tableta	0,004835203	1	0,892857183	0,97384807
52	subrayar	0,004147849	1	0,892857183	0,9750934
53	tablero	0,004147849	1	0,892857183	0,97633873
54	boleta	0,004147849	1	0,892857183	0,97758406
55	ejercicio	0,003558206	1	0,892857183	0,97882939
56	sillón	0,003052384	1	0,892857183	0,98007472
57	copiadora	0,003052384	1	0,892857183	0,98132005
58	pincel	0,003052384	1	0,892857183	0,98256538
59	estudiante	0,002618469	1	0,892857183	0,98381071
60	compañero	0,002618469	1	0,892857183	0,98505604
61	pegamento	0,002246237	1	0,892857183	0,98630137
62	registro	0,002246237	1	0,892857183	0,9875467
63	impresora	0,002246237	1	0,892857183	0,98879203
64	pinchar	0,00192692	1	0,892857183	0,99003736
65	deberes	0,00192692	1	0,892857183	0,99128269
66	luz	0,001652996	1	0,892857183	0,99252802
67	proyector	0,001652996	1	0,892857183	0,99377335
68	lámpara	0,001418013	1	0,892857183	0,99501868
69	escribir	0,001418013	1	0,892857183	0,99626401
70	mapamundi	0,001216433	1	0,892857183	0,99750934
71	horario	0,001216433	1	0,892857183	0,99875467
72	cartel	0,00104351	1	0,892857183	1

CI09, “iluminación y calefacción”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	lámpara	0,627458036	80	71,42857313	0,233918129
2	luz	0,422833234	59	52,67857313	0,406432749
3	fuego	0,263790518	39	34,82142985	0,520467836
4	sol	0,071262822	12	10,71428582	0,555555556
5	bombilla	0,06588643	10	8,928571641	0,584795322
6	calor	0,063645363	10	8,928571641	0,614035088
7	aire- acondicionado	0,058199514	9	8,035714179	0,640350877
8	vela	0,048937883	8	7,142857462	0,66374269
9	calorífico	0,042270858	8	7,142857462	0,687134503
10	calentador	0,037709363	6	5,35714291	0,704678363
11	temperatura	0,03762383	5	4,464285821	0,719298246
12	termo(sifón)	0,035461847	5	4,464285821	0,733918129
13	calefacción	0,031161118	6	5,35714291	0,751461988
14	calentar	0,028940191	6	5,35714291	0,769005848
15	chimenea	0,024566591	4	3,571428731	0,780701754
16	estufa	0,024186159	4	3,571428731	0,792397661
17	ventana	0,023553282	5	4,464285821	0,807017544
18	caliente	0,020619711	4	3,571428731	0,81871345
19	frío	0,019967657	4	3,571428731	0,830409357
20	ventilador	0,019542441	3	2,678571455	0,839181287
21	electricidad	0,018422676	4	3,571428731	0,850877193
22	apagar	0,018256757	3	2,678571455	0,859649123
23	farola	0,01629984	2	1,785714366	0,865497076
24	neón	0,015021139	4	3,571428731	0,877192982
25	radiador	0,015014158	2	1,785714366	0,883040936
26	enchufe	0,014742536	2	1,785714366	0,888888889
27	flexo	0,014742536	2	1,785714366	0,894736842
28	horno	0,013952722	2	1,785714366	0,900584795
29	energía- eléctrica	0,011755685	2	1,785714366	0,906432749
30	hogar	0,011519117	2	1,785714366	0,912280702
31	leña	0,011109738	2	1,785714366	0,918128655
32	encender	0,008928572	1	0,892857183	0,921052632
33	corriente- eléctrica	0,008448539	2	1,785714366	0,926900585
34	nevera	0,007371268	1	0,892857183	0,929824561
35	calefactor	0,007371268	1	0,892857183	0,932748538

36	hacer-frío	0,007371268	1	0,892857183	0,935672515
37	gas	0,007371268	1	0,892857183	0,938596491
38	calorífero	0,007371268	1	0,892857183	0,941520468
39	led	0,007371268	1	0,892857183	0,944444444
40	antorcha	0,007371268	1	0,892857183	0,947368421
41	calentamiento	0,006848778	2	1,785714366	0,953216374
42	energía-renovable	0,006085586	1	0,892857183	0,956140351
43	pantalla	0,006085586	1	0,892857183	0,959064327
44	candela	0,006085586	1	0,892857183	0,961988304
45	hacer-calor	0,006085586	1	0,892857183	0,964912281
46	noche	0,006085586	1	0,892857183	0,967836257
47	interruptor	0,006085586	1	0,892857183	0,970760234
48	acondicionador	0,006085586	1	0,892857183	0,973684211
49	central-nuclear	0,00502415	1	0,892857183	0,976608187
50	estrella	0,004147849	1	0,892857183	0,979532164
51	circuito	0,004147849	1	0,892857183	0,98245614
52	madera	0,004147849	1	0,892857183	0,985380117
53	pila	0,003424389	1	0,892857183	0,988304094
54	iluminación	0,002334015	1	0,892857183	0,99122807
55	hielo	0,001590831	1	0,892857183	0,994152047
56	aire	0,001313361	1	0,892857183	0,997076023
57	lamparilla	0,000895168	1	0,892857183	1

CI10, “la ciudad”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	calle	0,487558	61	54,46428657	0,052092229
2	coche	0,314487	52	46,42857015	0,096498719
3	tienda	0,309836	58	51,78571343	0,146029035
4	parque	0,300974	49	43,75	0,187873612
5	escuela	0,272103	48	42,85714328	0,228864219
6	casa	0,268297	42	37,5	0,264730999
7	edificio	0,249073	39	34,82142985	0,298035867
8	(auto)bús	0,241343	46	41,07142985	0,337318531
9	iglesia	0,237779	42	37,5	0,373185312
10	plaza	0,225906	37	33,03571343	0,404782237
11	estación	0,161481	33	29,46428657	0,432963279
12	bar	0,160773	37	33,03571343	0,464560205
13	carretera	0,157223	24	21,42857164	0,485055508
14	restaurante	0,109257	26	23,21428508	0,507258753
15	tren	0,106968	20	17,85714328	0,524338173
16	ayuntamiento	0,103767	17	15,17857164	0,538855679
17	árbol	0,103427	16	14,28571492	0,552519214
18	museo	0,091731	18	16,07142836	0,567890692
19	universidad	0,09025	21	18,75	0,585824082
20	farmacia	0,080797	16	14,28571492	0,599487617
21	palacio	0,074227	12	10,71428582	0,609735269
22	bici(cleta)	0,073542	16	14,28571492	0,623398804
23	cine	0,073393	15	13,39285672	0,636208369
24	banco	0,072099	13	11,60714254	0,647309991
25	parada	0,064902	14	12,5	0,659265585
26	jardín	0,064647	12	10,71428582	0,669513237
27	centro	0,061543	10	8,928571641	0,678052946
28	hospital	0,059606	14	12,5	0,69000854
29	supermercado	0,05543	12	10,71428582	0,700256191
30	monumento	0,055358	11	9,821428359	0,709649872
31	paso-de-cebra	0,053877	8	7,142857462	0,71648164
32	oficina	0,053371	12	10,71428582	0,726729291
33	mercado	0,052902	11	9,821428359	0,736122972
34	rascacielos	0,046968	6	5,35714291	0,741246798
35	avenida	0,043776	6	5,35714291	0,746370623
36	gente	0,043614	8	7,142857462	0,753202391
37	hotel	0,042336	9	8,035714179	0,76088813

38	semáforo	0,0367	7	6,25	0,766865927
39	punte	0,035838	8	7,142857462	0,773697694
40	tranvía	0,035831	8	7,142857462	0,780529462
41	tráfico	0,035151	8	7,142857462	0,78736123
42	persona	0,031774	5	4,464285821	0,791631085
43	negocio	0,030525	5	4,464285821	0,795900939
44	metro	0,029728	6	5,35714291	0,801024765
45	municipio	0,02858	5	4,464285821	0,80529462
46	casco-antiguo	0,026451	4	3,571428731	0,808710504
47	disco(teca)	0,026117	7	6,25	0,814688301
48	barrio	0,025663	5	4,464285821	0,818958155
49	transporte-público	0,025534	4	3,571428731	0,822374039
50	peluquería	0,024897	5	4,464285821	0,826643894
51	cruce	0,023765	5	4,464285821	0,830913749
52	policía	0,023169	4	3,571428731	0,834329633
53	teatro	0,022814	6	5,35714291	0,839453459
54	piso	0,022679	4	3,571428731	0,842869342
55	instituto	0,020626	4	3,571428731	0,846285226
56	panadería	0,020371	4	3,571428731	0,84970111
57	esquina	0,019134	3	2,678571455	0,852263023
58	peatón	0,019009	3	2,678571455	0,854824936
59	aparcamiento	0,017839	4	3,571428731	0,85824082
60	histórico	0,017562	3	2,678571455	0,860802733
61	biblioteca	0,017311	5	4,464285821	0,865072588
62	fuelle	0,017284	3	2,678571455	0,8676345
63	río	0,017197	5	4,464285821	0,871904355
64	estrada	0,016288	2	1,785714366	0,873612297
65	moto(cicleta)	0,016193	3	2,678571455	0,87617421
66	centro-comercial	0,015474	3	2,678571455	0,878736123
67	alcalde	0,015269	3	2,678571455	0,881298036
68	pasear	0,015169	4	3,571428731	0,88471392
69	aeropuerto	0,014971	3	2,678571455	0,887275833
70	apartamento	0,014262	4	3,571428731	0,890691716
71	turista	0,013695	3	2,678571455	0,893253629
72	cruzar	0,012887	4	3,571428731	0,896669513
73	paseo	0,01283	2	1,785714366	0,898377455
74	carnicería	0,012698	3	2,678571455	0,900939368
75	izquierda	0,01265	4	3,571428731	0,904355252
76	heladería	0,012127	3	2,678571455	0,906917165

77	cafetería	0,012117	2	1,785714366	0,908625107
78	derecha	0,011784	3	2,678571455	0,91118702
79	taxi	0,011582	3	2,678571455	0,913748933
80	oficio	0,011389	2	1,785714366	0,915456874
81	piscina	0,011104	2	1,785714366	0,917164816
82	habitación	0,011052	2	1,785714366	0,918872758
83	gimnasio	0,010757	4	3,571428731	0,922288642
84	catedral	0,010181	3	2,678571455	0,924850555
85	estatua	0,010128	2	1,785714366	0,926558497
86	carro	0,010081	2	1,785714366	0,928266439
87	movida	0,009913	2	1,785714366	0,929974381
88	almacén	0,009913	2	1,785714366	0,931682323
89	trabajo	0,009538	2	1,785714366	0,933390265
90	estanco	0,009282	3	2,678571455	0,935952178
91	torre	0,008929	1	0,892857183	0,936806149
92	aparcar	0,008608	3	2,678571455	0,939368061
93	cartel	0,00859	2	1,785714366	0,941076003
94	colegio	0,008178	2	1,785714366	0,942783945
95	muralla	0,008144	1	0,892857183	0,943637916
96	autovía	0,008144	1	0,892857183	0,944491887
97	lampión	0,008144	1	0,892857183	0,945345858
98	fábrica	0,007649	2	1,785714366	0,9470538
99	moderno	0,007428	1	0,892857183	0,947907771
100	cementerio	0,007147	2	1,785714366	0,949615713
101	ciudadano	0,006804	2	1,785714366	0,951323655
102	esmog	0,006775	1	0,892857183	0,952177626
103	auto(móvil)	0,006775	1	0,892857183	0,953031597
104	establecimiento	0,006775	1	0,892857183	0,953885568
105	joyería	0,006775	1	0,892857183	0,954739539
106	agencia	0,00618	1	0,892857183	0,95559351
107	correos	0,00618	1	0,892857183	0,956447481
108	playa	0,00618	1	0,892857183	0,957301452
109	bloque-de-piso	0,00618	1	0,892857183	0,958155423
110	rincón	0,00618	1	0,892857183	0,959009394
111	sitio	0,00618	1	0,892857183	0,959863365
112	despacho	0,005636	1	0,892857183	0,960717336
113	luz	0,005636	1	0,892857183	0,961571307
114	arte	0,005636	1	0,892857183	0,962425278
115	municipalidad	0,005636	1	0,892857183	0,963279249
116	acera	0,005636	1	0,892857183	0,964133219

117	tabaquería	0,005141	1	0,892857183	0,96498719
118	antiguo	0,005141	1	0,892857183	0,965841161
119	señal-de-tráfico	0,005141	1	0,892857183	0,966695132
120	pueblo	0,005141	1	0,892857183	0,967549103
121	verde	0,005141	1	0,892857183	0,968403074
122	cosmopolita	0,005141	1	0,892857183	0,969257045
123	girar	0,004749	2	1,785714366	0,970964987
124	hostal	0,004689	1	0,892857183	0,971818958
125	glorieta	0,004689	1	0,892857183	0,972672929
126	frutería	0,004277	1	0,892857183	0,9735269
127	ruido	0,004277	1	0,892857183	0,974380871
128	centro-histórico	0,004277	1	0,892857183	0,975234842
129	vehículo	0,004277	1	0,892857183	0,976088813
130	tapas	0,003901	1	0,892857183	0,976942784
131	ir-de-compras	0,003901	1	0,892857183	0,977796755
132	historia	0,003901	1	0,892857183	0,978650726
133	zona-peatonal	0,003901	1	0,892857183	0,979504697
134	comer	0,003558	1	0,892857183	0,980358668
135	oficina-de-correos	0,003558	1	0,892857183	0,981212639
136	vivir	0,003558	1	0,892857183	0,98206661
137	viajar	0,003558	1	0,892857183	0,982920581
138	beber	0,003245	1	0,892857183	0,983774552
139	salir	0,003245	1	0,892857183	0,984628523
140	billete	0,003245	1	0,892857183	0,985482494
141	florería	0,003245	1	0,892857183	0,986336465
142	galería	0,00296	1	0,892857183	0,987190436
143	comprar	0,00296	1	0,892857183	0,988044406
144	flor	0,00296	1	0,892857183	0,988898377
145	puerto	0,00296	1	0,892857183	0,989752348
146	bajar	0,00296	1	0,892857183	0,990606319
147	ciudadanía	0,00296	1	0,892857183	0,99146029
148	pasillo	0,0027	1	0,892857183	0,992314261
149	café	0,0027	1	0,892857183	0,993168232
150	vender	0,0027	1	0,892857183	0,994022203
151	estudio	0,002463	1	0,892857183	0,994876174
152	terraza	0,002463	1	0,892857183	0,995730145
153	caminar	0,002463	1	0,892857183	0,996584116
154	metrópoli(s)	0,002463	1	0,892857183	0,997438087
155	floristería	0,002463	1	0,892857183	0,998292058

156	guardería	0,002246	1	0,892857183	0,999146029
157	pizzería	0,002049	1	0,892857183	1

CI11, “el campo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	árbol	0,439436	63	56,25	0,091970803
2	animal	0,315958	46	41,07142985	0,159124088
3	flor	0,213539	34	30,35714328	0,208759124
4	tierra	0,168267	28	25	0,249635036
5	hierba	0,167694	23	20,53571492	0,283211679
6	vaca	0,149145	26	23,21428508	0,321167883
7	campesino	0,1341	24	21,42857164	0,35620438
8	planta	0,123476	21	18,75	0,386861314
9	caballo	0,119684	23	20,53571492	0,420437956
10	casa	0,09598	14	12,5	0,440875912
11	perro	0,087281	15	13,39285672	0,462773723
12	naturaleza	0,086197	15	13,39285672	0,484671533
13	fruta	0,084111	14	12,5	0,505109489
14	río	0,073646	16	14,28571492	0,528467153
15	gallina	0,072989	15	13,39285672	0,550364964
16	rural	0,062334	8	7,142857462	0,562043796
17	verdura	0,061294	10	8,928571641	0,576642336
18	tractor	0,054285	8	7,142857462	0,588321168
19	cerdo	0,051808	12	10,71428582	0,605839416
20	maíz	0,049381	9	8,035714179	0,618978102
21	molino	0,046934	8	7,142857462	0,630656934
22	agricultura	0,045797	9	8,035714179	0,64379562
23	oveja	0,044852	7	6,25	0,654014599
24	cultivación	0,043185	8	7,142857462	0,665693431
25	campo	0,04219	6	5,35714291	0,674452555
26	aire	0,040176	7	6,25	0,684671533
27	parque	0,036433	5	4,464285821	0,691970803
28	agua	0,032783	7	6,25	0,702189781
29	verde	0,031061	6	5,35714291	0,710948905
30	cereales	0,029859	5	4,464285821	0,718248175
31	gato	0,028939	7	6,25	0,728467153
32	pato	0,023539	3	2,678571455	0,732846715
33	bosque	0,022677	3	2,678571455	0,737226277
34	sol	0,021968	5	4,464285821	0,744525547
35	pasear	0,020614	4	3,571428731	0,750364964
36	pájaro	0,020209	4	3,571428731	0,75620438
37	insecto	0,020053	5	4,464285821	0,76350365

38	tranquilidad	0,01944	5	4,464285821	0,77080292
39	conejo	0,01925	5	4,464285821	0,77810219
40	jardín	0,019228	4	3,571428731	0,783941606
41	pollo	0,019204	3	2,678571455	0,788321168
42	cabra	0,018616	4	3,571428731	0,794160584
43	burro	0,016741	4	3,571428731	0,8
44	granja	0,015597	2	1,785714366	0,802919708
45	silencio	0,015507	4	3,571428731	0,808759124
46	toro	0,015184	4	3,571428731	0,81459854
47	huerto	0,015034	3	2,678571455	0,818978102
48	paisaje	0,014611	2	1,785714366	0,82189781
49	harina	0,014126	2	1,785714366	0,824817518
50	cultivar	0,013993	3	2,678571455	0,82919708
51	césped	0,013624	2	1,785714366	0,832116788
52	bici(cleta)	0,013506	4	3,571428731	0,837956204
53	trabajo	0,012376	3	2,678571455	0,842335766
54	natura	0,012009	2	1,785714366	0,845255474
55	grano	0,011764	2	1,785714366	0,848175182
56	niebla	0,011497	3	2,678571455	0,852554745
57	rosa	0,011351	2	1,785714366	0,855474453
58	viento	0,011147	2	1,785714366	0,858394161
59	aldea	0,010824	2	1,785714366	0,861313869
60	girasol	0,010489	2	1,785714366	0,864233577
61	arroz	0,009915	2	1,785714366	0,867153285
62	lago	0,009837	2	1,785714366	0,870072993
63	iglesia	0,009815	2	1,785714366	0,872992701
64	paz	0,009454	2	1,785714366	0,875912409
65	pueblo	0,00912	2	1,785714366	0,878832117
66	cine	0,008929	1	0,892857183	0,880291971
67	gaucho	0,008929	1	0,892857183	0,881751825
68	afuera	0,008929	1	0,892857183	0,883211679
69	red	0,008929	1	0,892857183	0,884671533
70	gancho	0,008929	1	0,892857183	0,886131387
71	cultivador	0,008258	2	1,785714366	0,889051095
72	gallo	0,008258	2	1,785714366	0,891970803
73	al-aire-libre	0,008258	2	1,785714366	0,894890511
74	puerto	0,007799	1	0,892857183	0,896350365
75	olivo	0,007799	1	0,892857183	0,897810219
76	trigo	0,007799	1	0,892857183	0,899270073
77	colina	0,007799	1	0,892857183	0,900729927

78	hormiga	0,007799	1	0,892857183	0,902189781
79	paja	0,007799	1	0,892857183	0,903649635
80	manzana	0,007182	2	1,785714366	0,906569343
81	libertad	0,007182	2	1,785714366	0,909489051
82	vino	0,006926	2	1,785714366	0,912408759
83	bar	0,006812	1	0,892857183	0,913868613
84	ganado	0,006812	1	0,892857183	0,915328467
85	villa	0,006812	1	0,892857183	0,916788321
86	cultivo	0,006812	1	0,892857183	0,918248175
87	habitación	0,00595	1	0,892857183	0,919708029
88	coche	0,00595	1	0,892857183	0,921167883
89	hortaliza	0,00595	1	0,892857183	0,922627737
90	ternera	0,00595	1	0,892857183	0,924087591
91	aceituno	0,00595	1	0,892857183	0,925547445
92	abuelo	0,00595	1	0,892857183	0,927007299
93	pasto	0,00595	1	0,892857183	0,928467153
94	semilla	0,005224	2	1,785714366	0,931386861
95	leche	0,005197	1	0,892857183	0,932846715
96	gente	0,005197	1	0,892857183	0,934306569
97	ovino	0,005197	1	0,892857183	0,935766423
98	mata	0,005197	1	0,892857183	0,937226277
99	vegetal	0,005197	1	0,892857183	0,938686131
100	manzano	0,005197	1	0,892857183	0,940145985
101	calle	0,004539	1	0,892857183	0,941605839
102	trabajador	0,004539	1	0,892857183	0,943065693
103	heno	0,004539	1	0,892857183	0,944525547
104	puerco	0,004539	1	0,892857183	0,945985401
105	semen	0,004539	1	0,892857183	0,947445255
106	naranja	0,004539	1	0,892857183	0,948905109
107	chalé	0,004539	1	0,892857183	0,950364964
108	tomate	0,003965	1	0,892857183	0,951824818
109	producto	0,003965	1	0,892857183	0,953284672
110	ave	0,003965	1	0,892857183	0,954744526
111	fértil	0,003965	1	0,892857183	0,95620438
112	abeja	0,003965	1	0,892857183	0,957664234
113	hoja	0,003463	1	0,892857183	0,959124088
114	vacaciones	0,003463	1	0,892857183	0,960583942
115	espacio	0,003463	1	0,892857183	0,962043796
116	aceite	0,003025	1	0,892857183	0,96350365
117	estación	0,003025	1	0,892857183	0,964963504

118	cordero	0,003025	1	0,892857183	0,966423358
119	encierro	0,003025	1	0,892857183	0,967883212
120	piedra	0,003025	1	0,892857183	0,969343066
121	fuelle	0,002642	1	0,892857183	0,97080292
122	bestia	0,002642	1	0,892857183	0,972262774
123	miel	0,002642	1	0,892857183	0,973722628
124	raíz	0,002642	1	0,892857183	0,975182482
125	chivo	0,002642	1	0,892857183	0,976642336
126	terreno	0,002308	1	0,892857183	0,97810219
127	asno	0,002308	1	0,892857183	0,979562044
128	arboreto	0,002308	1	0,892857183	0,981021898
129	tienda	0,002016	1	0,892857183	0,982481752
130	aceña	0,002016	1	0,892857183	0,983941606
131	agricultor	0,002016	1	0,892857183	0,98540146
132	irrigación	0,002016	1	0,892857183	0,986861314
133	granjero	0,002016	1	0,892857183	0,988321168
134	pavo	0,002016	1	0,892857183	0,989781022
135	andar	0,001761	1	0,892857183	0,991240876
136	finca	0,001538	1	0,892857183	0,99270073
137	correr	0,001343	1	0,892857183	0,994160584
138	chancho	0,001343	1	0,892857183	0,995620438
139	jugar	0,001173	1	0,892857183	0,997080292
140	criadero	0,001173	1	0,892857183	0,998540146
141	trabajar	0,001025	1	0,892857183	1

CI12, “medios de transporte”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	coche	0,687179	92	82,1428597	0,113861386
2	(auto)bús	0,627778	93	83,03571343	0,228960396
3	tren	0,565213	94	83,92857313	0,34529703
4	bici(cleta)	0,50198	96	85,71428657	0,464108911
5	avión	0,45325	84	75	0,568069307
6	taxi	0,202902	48	42,85714328	0,627475248
7	metro	0,19024	47	41,96428657	0,685643564
8	moto(cicleta)	0,188392	55	49,10714328	0,753712871
9	barco	0,181332	50	44,64285672	0,815594059
10	tranvía	0,123226	26	23,21428508	0,847772277
11	a-pie	0,06111	18	16,07142836	0,870049505
12	carro	0,047691	10	8,928571641	0,882425743
13	caballo	0,033869	8	7,142857462	0,892326733
14	camión	0,027908	10	8,928571641	0,90470297
15	nave	0,025833	8	7,142857462	0,91460396
16	helicóptero	0,022956	8	7,142857462	0,92450495
17	patinetes	0,018114	7	6,25	0,933168317
18	auto(móvil)	0,008929	1	0,892857183	0,934405941
19	abono	0,008929	1	0,892857183	0,935643564
20	carne-de-conducir	0,008765	3	2,678571455	0,939356436
21	conducir	0,008267	4	3,571428731	0,944306931
22	tomar	0,006177	2	1,785714366	0,946782178
23	camioneta	0,005855	2	1,785714366	0,949257426
24	crucero	0,005676	2	1,785714366	0,951732673
25	aéreo	0,005251	1	0,892857183	0,952970297
26	vehículo	0,005251	1	0,892857183	0,954207921
27	tándem	0,004905	2	1,785714366	0,956683168
28	billete	0,004759	3	2,678571455	0,96039604
29	ciclomotor	0,0044	1	0,892857183	0,961633663
30	máquina	0,0044	1	0,892857183	0,962871287
31	coger	0,003339	2	1,785714366	0,965346535
32	correr	0,003092	2	1,785714366	0,967821782
33	guagua	0,003089	1	0,892857183	0,969059406
34	AVE	0,003089	1	0,892857183	0,97029703
35	subir	0,003089	1	0,892857183	0,971534653

36	a-nado	0,003089	1	0,892857183	0,972772277
37	góndola	0,003089	1	0,892857183	0,974009901
38	velero	0,003089	1	0,892857183	0,975247525
39	carabela	0,003089	1	0,892857183	0,976485149
40	esperar	0,002588	1	0,892857183	0,977722772
41	viajar	0,002588	1	0,892857183	0,978960396
42	ambulancia	0,002588	1	0,892857183	0,98019802
43	colectivo	0,002588	1	0,892857183	0,981435644
44	horario	0,002168	1	0,892857183	0,982673267
45	esmog	0,002168	1	0,892857183	0,983910891
46	funicular	0,002168	1	0,892857183	0,985148515
47	autocaravana	0,002168	1	0,892857183	0,986386139
48	globo-aerostático	0,002168	1	0,892857183	0,987623762
49	caminar	0,001817	1	0,892857183	0,988861386
50	submarino	0,001817	1	0,892857183	0,99009901
51	ir	0,001522	1	0,892857183	0,991336634
52	parada	0,001522	1	0,892857183	0,992574257
53	transporte-público	0,001522	1	0,892857183	0,993811881
54	andar	0,001522	1	0,892857183	0,995049505
55	gasolina	0,001275	1	0,892857183	0,996287129
56	público	0,001275	1	0,892857183	0,997524752
57	ómnibus	0,001275	1	0,892857183	0,998762376
58	hacer-dedo	0,000895	1	0,892857183	1

CI13, “trabajos del campo y del jardín”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	jardinero	0,3355	42	37,5	0,166667
2	campesino	0,25699	32	28,57143	0,293651
3	plantar	0,162166	21	18,75	0,376984
4	cortar	0,146328	25	22,32143	0,47619
5	flor	0,04982	8	7,142857	0,507937
6	cultivar	0,048955	7	6,25	0,535714
7	agricultor	0,047603	7	6,25	0,563492
8	regar	0,036224	5	4,464286	0,583333
9	granjero	0,035847	5	4,464286	0,603175
10	florista	0,029939	5	4,464286	0,623016
11	dar-agua	0,028753	4	3,571429	0,638889
12	pastor	0,028238	5	4,464286	0,65873
13	fruta	0,027617	5	4,464286	0,678571
14	coger	0,02468	4	3,571429	0,694444
15	florero	0,022845	4	3,571429	0,710317
16	escavar	0,021659	3	2,678571	0,722222
17	hierba	0,020058	5	4,464286	0,742063
18	cultivador	0,019581	3	2,678571	0,753968
19	arar	0,017885	3	2,678571	0,765873
20	jardín	0,016023	2	1,785714	0,77381
21	recoger	0,012731	2	1,785714	0,781746
22	planta	0,012515	3	2,678571	0,793651
23	árbol	0,011784	3	2,678571	0,805556
24	crecer	0,011595	3	2,678571	0,81746
25	verdura	0,009921	2	1,785714	0,825397
26	tierra	0,009195	2	1,785714	0,833333
27	gato	0,008929	1	0,892857	0,837302
28	cultivación	0,008929	1	0,892857	0,84127
29	sembrar	0,008929	1	0,892857	0,845238
30	irrigar	0,008929	1	0,892857	0,849206
31	tijeras	0,007094	1	0,892857	0,853175
32	perro	0,007094	1	0,892857	0,857143
33	molinero	0,007094	1	0,892857	0,861111
34	verdulero	0,007094	1	0,892857	0,865079
35	obrero	0,007094	1	0,892857	0,869048
36	criador	0,007094	1	0,892857	0,873016

37	extirpar	0,007094	1	0,892857	0,876984
38	mantequilla	0,007094	1	0,892857	0,880952
39	criar	0,007094	1	0,892857	0,884921
40	cosecha	0,007094	1	0,892857	0,888889
41	madurar	0,005654	2	1,785714	0,896825
42	agua	0,005636	1	0,892857	0,900794
43	leche	0,005636	1	0,892857	0,904762
44	vaca	0,005636	1	0,892857	0,90873
45	gaucho	0,005636	1	0,892857	0,912698
46	irrigación	0,005636	1	0,892857	0,916667
47	rama	0,005636	1	0,892857	0,920635
48	viticultor	0,005636	1	0,892857	0,924603
49	enterrar	0,005636	1	0,892857	0,928571
50	oveja	0,004478	1	0,892857	0,93254
51	vaquero	0,004478	1	0,892857	0,936508
52	aceite	0,004478	1	0,892857	0,940476
53	arado	0,004478	1	0,892857	0,944444
54	vender	0,003558	1	0,892857	0,948413
55	caballo	0,003558	1	0,892857	0,952381
56	granja	0,003558	1	0,892857	0,956349
57	vaso	0,002827	1	0,892857	0,960317
58	trabajador	0,002827	1	0,892857	0,964286
59	burro	0,002827	1	0,892857	0,968254
60	campo	0,002827	1	0,892857	0,972222
61	vino	0,002246	1	0,892857	0,97619
62	mariposa	0,002246	1	0,892857	0,980159
63	cosechar	0,002246	1	0,892857	0,984127
64	pájaro	0,001785	1	0,892857	0,988095
65	pez	0,001418	1	0,892857	0,992063
66	conejo	0,001127	1	0,892857	0,996032
67	cerdo	0,000895	1	0,892857	1

CI14, “los animales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	perro	0,816215	97	86,6071403	0,089814815
2	gato	0,750967	96	85,71428657	0,178703704
3	caballo	0,490249	86	76,78571343	0,258333333
4	vaca	0,298745	56	50	0,310185185
5	pez	0,229914	46	41,07142985	0,352777778
6	pájaro	0,209367	43	38,39285672	0,392592593
7	león	0,205385	47	41,96428657	0,436111111
8	cerdo	0,187271	35	31,25	0,468518519
9	gallina	0,144076	29	25,89285672	0,49537037
10	tigre	0,131441	31	27,67857015	0,524074074
11	conejo	0,13028	26	23,21428508	0,548148148
12	toro	0,120198	27	24,10714328	0,573148148
13	elefante	0,098352	24	21,42857164	0,59537037
14	mariposa	0,093563	20	17,85714328	0,613888889
15	burro	0,08815	21	18,75	0,633333333
16	tortuga	0,08156	17	15,17857164	0,649074074
17	jirafa	0,07895	21	18,75	0,668518519
18	ave	0,077444	14	12,5	0,681481481
19	ratón	0,076414	16	14,28571492	0,696296296
20	cebra	0,075974	18	16,07142836	0,712962963
21	pollo	0,074909	14	12,5	0,725925926
22	serpiente	0,058688	18	16,07142836	0,742592593
23	hormiga	0,057993	15	13,39285672	0,756481481
24	gallo	0,055714	14	12,5	0,769444444
25	oso	0,047829	12	10,71428582	0,780555556
26	lobo	0,04473	10	8,928571641	0,789814815
27	oveja	0,04124	12	10,71428582	0,800925926
28	cabra	0,04123	14	12,5	0,813888889
29	mono	0,034346	8	7,142857462	0,821296296
30	zorro	0,033572	6	5,35714291	0,826851852
31	pato	0,033493	7	6,25	0,833333333
32	delfín	0,032163	7	6,25	0,839814815
33	papagayo	0,031386	7	6,25	0,846296296
34	tiburón	0,029542	7	6,25	0,852777778
35	puerco	0,029306	7	6,25	0,859259259
36	mosquito	0,029155	9	8,035714179	0,867592593

37	hámster	0,028938	6	5,35714291	0,873148148
38	cocodrilo	0,027103	7	6,25	0,87962963
39	cucaracha	0,026991	7	6,25	0,886111111
40	topo	0,024902	6	5,35714291	0,891666667
41	caracol	0,024341	6	5,35714291	0,897222222
42	buey	0,021006	4	3,571428731	0,900925926
43	pingüino	0,019765	6	5,35714291	0,906481481
44	águila	0,017086	6	5,35714291	0,912037037
45	gaviota	0,01631	4	3,571428731	0,915740741
46	lagarto	0,01592	3	2,678571455	0,918518519
47	paloma	0,015871	5	4,464285821	0,923148148
48	salmón	0,01544	4	3,571428731	0,926851852
49	mariquita	0,015357	2	1,785714366	0,928703704
50	yegua	0,012189	2	1,785714366	0,930555556
51	avestruz	0,011175	3	2,678571455	0,933333333
52	jabalí	0,011056	2	1,785714366	0,935185185
53	canario	0,010925	2	1,785714366	0,937037037
54	araña	0,010521	4	3,571428731	0,940740741
55	panda	0,010158	2	1,785714366	0,942592593
56	abeja	0,009571	2	1,785714366	0,944444444
57	ciervo	0,009571	2	1,785714366	0,946296296
58	pitón	0,008929	1	0,892857183	0,947222222
59	pantera	0,00796	2	1,785714366	0,949074074
60	tucán	0,007815	2	1,785714366	0,950925926
61	ballena	0,007535	3	2,678571455	0,953703704
62	cabrón	0,00748	2	1,785714366	0,955555556
63	flamenco	0,00748	2	1,785714366	0,957407407
64	llama	0,007314	2	1,785714366	0,959259259
65	culebra	0,007024	4	3,571428731	0,962962963
66	mosca	0,006428	1	0,892857183	0,963888889
67	cochino	0,006428	1	0,892857183	0,964814815
68	pulpo	0,006318	2	1,785714366	0,966666667
69	insecto	0,005761	1	0,892857183	0,967592593
70	estrella-marina	0,005761	1	0,892857183	0,968518519
71	pavo	0,005761	1	0,892857183	0,969444444
72	tarántula	0,005761	1	0,892857183	0,97037037
73	cordero	0,005385	2	1,785714366	0,972222222
74	tigrillo	0,005164	1	0,892857183	0,973148148

75	bicho	0,005164	1	0,892857183	0,974074074
76	jaguar	0,005164	1	0,892857183	0,975
77	ardilla	0,005164	1	0,892857183	0,975925926
78	buitre	0,005164	1	0,892857183	0,976851852
79	oca	0,004628	1	0,892857183	0,977777778
80	mulo	0,004148	1	0,892857183	0,978703704
81	chanchó	0,004148	1	0,892857183	0,97962963
82	sierpe	0,004148	1	0,892857183	0,980555556
83	guepardo	0,004148	1	0,892857183	0,981481481
84	gamba	0,003718	1	0,892857183	0,982407407
85	rana	0,003718	1	0,892857183	0,983333333
86	cangrejo	0,003718	1	0,892857183	0,984259259
87	leopardo	0,003332	1	0,892857183	0,985185185
88	camarón	0,003332	1	0,892857183	0,986111111
89	gorila	0,003332	1	0,892857183	0,987037037
90	pécora	0,002986	1	0,892857183	0,987962963
91	hipopótamo	0,002841	2	1,785714366	0,989814815
92	orangután	0,002676	1	0,892857183	0,990740741
93	chinche	0,002399	1	0,892857183	0,991666667
94	canguro	0,002399	1	0,892857183	0,992592593
95	chivo	0,00215	1	0,892857183	0,993518519
96	alpaca	0,001927	1	0,892857183	0,994444444
97	escorpión	0,001727	1	0,892857183	0,99537037
98	cachorro	0,001548	1	0,892857183	0,996296296
99	murciélago	0,001548	1	0,892857183	0,997222222
100	rata	0,001548	1	0,892857183	0,998148148
101	foca	0,001387	1	0,892857183	0,999074074
102	bagre	0,000999	1	0,892857183	1

CI15, “juegos y distracciones”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	fútbol	0,358128	52	46,42857015	0,066326531
2	baloncesto	0,317442	50	44,64285672	0,130102041
3	leer	0,205784	32	28,57142985	0,170918367
4	libro	0,1811	32	28,57142985	0,211734694
5	bailar	0,177022	33	29,46428657	0,253826531
6	cartas	0,167424	25	22,32142836	0,285714286
7	pelota	0,137515	21	18,75	0,3125
8	tele(visión)	0,135749	21	18,75	0,339285714
9	ordenador	0,130146	21	18,75	0,366071429
10	videojuegos	0,124964	19	16,96428508	0,390306122
11	cine	0,124059	19	16,96428508	0,414540816
12	escuchar-música	0,120973	24	21,42857164	0,445153061
13	natación	0,103617	23	20,53571492	0,474489796
14	correr	0,099855	20	17,85714328	0,5
15	nadar	0,095567	20	17,85714328	0,525510204
16	jugar	0,092449	17	15,17857164	0,547193878
17	pasear	0,09174	16	14,28571492	0,567602041
18	tenis	0,077068	17	15,17857164	0,589285714
19	voleibol	0,067557	13	11,60714254	0,605867347
20	pelí(cula)	0,066639	12	10,71428582	0,621173469
21	tocar	0,065828	11	9,821428359	0,635204082
22	deporte	0,063196	12	10,71428582	0,650510204
23	balón	0,060566	10	8,928571641	0,663265306
24	disco(teca)	0,058453	9	8,035714179	0,674744898
25	pintar	0,049994	10	8,928571641	0,6875
26	ajedrez	0,049854	7	6,25	0,696428571
27	cantar	0,049265	9	8,035714179	0,707908163
28	dibujar	0,046347	9	8,035714179	0,719387755
29	muñeca	0,045574	8	7,142857462	0,729591837
30	móvil	0,042357	7	6,25	0,738520408
31	salir	0,041108	8	7,142857462	0,74872449
32	<i>Play-Station</i>	0,040487	5	4,464285821	0,755102041
33	fiesta	0,035948	5	4,464285821	0,761479592
34	salir-con-amigos	0,035292	6	5,35714291	0,769132653
35	guitarra	0,034245	6	5,35714291	0,776785714
36	juego-de-mesa	0,033711	6	5,35714291	0,784438776

37	escondite	0,030285	5	4,464285821	0,790816327
38	parque	0,029474	5	4,464285821	0,797193878
39	danza	0,024772	3	2,678571455	0,801020408
40	parque-de- diversiones	0,023899	4	3,571428731	0,806122449
41	dado	0,021499	3	2,678571455	0,80994898
42	instrumento- musical	0,020806	5	4,464285821	0,816326531
43	teatro	0,020196	4	3,571428731	0,821428571
44	balonvolea	0,019863	3	2,678571455	0,825255102
45	viajar	0,019811	4	3,571428731	0,830357143
46	bar	0,018203	3	2,678571455	0,834183673
47	playa	0,017329	4	3,571428731	0,839285714
48	piano	0,017225	4	3,571428731	0,844387755
49	caminar	0,017156	3	2,678571455	0,848214286
50	dibujos- animados	0,01677	3	2,678571455	0,852040816
51	peluche	0,016153	3	2,678571455	0,855867347
52	radio	0,015844	2	1,785714366	0,858418367
53	reír	0,015656	3	2,678571455	0,862244898
54	cocinar	0,014999	5	4,464285821	0,868622449
55	ir-de-compras	0,014788	4	3,571428731	0,87372449
56	gimnasio	0,01452	4	3,571428731	0,878826531
57	golf	0,013626	3	2,678571455	0,882653061
58	amigo	0,013538	3	2,678571455	0,886479592
59	entrenarse	0,012579	2	1,785714366	0,889030612
60	bici(cleta)	0,011824	4	3,571428731	0,894132653
61	quedar-con- amigos	0,011063	2	1,785714366	0,896683673
62	grupo	0,009298	2	1,785714366	0,899234694
63	museo	0,009124	3	2,678571455	0,903061224
64	computador	0,008929	1	0,892857183	0,904336735
65	pasárselo-bien	0,007858	1	0,892857183	0,905612245
66	tobogán	0,007858	1	0,892857183	0,906887755
67	sexo	0,007858	1	0,892857183	0,908163265
68	ocio	0,007858	1	0,892857183	0,909438776
69	juego	0,007798	2	1,785714366	0,911989796
70	comer	0,006981	2	1,785714366	0,914540816
71	andar	0,006915	1	0,892857183	0,915816327
72	patinetes	0,006915	1	0,892857183	0,917091837

73	cómic	0,006915	1	0,892857183	0,918367347
74	descansar	0,006915	1	0,892857183	0,919642857
75	karaoke	0,006915	1	0,892857183	0,920918367
76	arena	0,006915	1	0,892857183	0,922193878
77	partido	0,00664	2	1,785714366	0,924744898
78	ir-de-vacaciones	0,006086	1	0,892857183	0,926020408
79	tapear	0,006086	1	0,892857183	0,927295918
80	actuar	0,006086	1	0,892857183	0,928571429
81	ir-de-tapas	0,006086	1	0,892857183	0,929846939
82	pillapilla	0,006086	1	0,892857183	0,931122449
83	construcciones	0,006086	1	0,892857183	0,932397959
84	emborracharse	0,006086	1	0,892857183	0,933673469
85	balonmano	0,006086	1	0,892857183	0,93494898
86	esquiar	0,006086	1	0,892857183	0,93622449
87	niño	0,006086	1	0,892857183	0,9375
88	<i>Monopoly</i>	0,006086	1	0,892857183	0,93877551
89	rompecabezas	0,006086	1	0,892857183	0,94005102
90	perro	0,005844	2	1,785714366	0,942602041
91	chatear	0,005577	2	1,785714366	0,945153061
92	beber	0,005356	1	0,892857183	0,946428571
93	ir-de-bares	0,005356	1	0,892857183	0,947704082
94	tejo	0,005356	1	0,892857183	0,948979592
95	saltar	0,005356	1	0,892857183	0,950255102
96	atracción	0,005356	1	0,892857183	0,951530612
97	beisbol	0,005356	1	0,892857183	0,952806122
98	fotografía	0,005017	2	1,785714366	0,955357143
99	maquillaje	0,004908	2	1,785714366	0,957908163
100	piscina	0,004713	1	0,892857183	0,959183673
101	restaurante	0,004713	1	0,892857183	0,960459184
102	energía	0,004713	1	0,892857183	0,961734694
103	broma	0,004713	1	0,892857183	0,963010204
104	atletismo	0,004713	1	0,892857183	0,964285714
105	mar	0,004319	2	1,785714366	0,966836735
106	internet	0,004148	1	0,892857183	0,968112245
107	canción	0,004148	1	0,892857183	0,969387755
108	parqués	0,004148	1	0,892857183	0,970663265
109	piñata	0,004148	1	0,892857183	0,971938776
110	lucha	0,004148	1	0,892857183	0,973214286

111	hablar	0,004148	1	0,892857183	0,974489796
112	boxeo	0,004148	1	0,892857183	0,975765306
113	competición	0,00365	1	0,892857183	0,977040816
114	sonreír	0,00365	1	0,892857183	0,978316327
115	juguete	0,00365	1	0,892857183	0,979591837
116	excursión	0,003212	1	0,892857183	0,980867347
117	yoga	0,003212	1	0,892857183	0,982142857
118	club	0,003212	1	0,892857183	0,983418367
119	soldadito-de-plomo	0,003212	1	0,892857183	0,984693878
120	telenovela	0,002827	1	0,892857183	0,985969388
121	escribir	0,002827	1	0,892857183	0,987244898
122	patinar	0,002827	1	0,892857183	0,988520408
123	movida	0,00219	1	0,892857183	0,989795918
124	colorar	0,00219	1	0,892857183	0,991071429
125	redes-sociales	0,001927	1	0,892857183	0,992346939
126	diversión	0,001927	1	0,892857183	0,993622449
127	escalar	0,001927	1	0,892857183	0,994897959
128	caballo	0,001696	1	0,892857183	0,996173469
129	fotografiar	0,001492	1	0,892857183	0,99744898
130	colina	0,001313	1	0,892857183	0,99872449
131	montaña	0,001156	1	0,892857183	1

CI16, “profesiones y oficios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	profesor	0,714364	95	84,82142687	0,106263982
2	médico	0,213237	36	32,14285672	0,146532438
3	cocinero	0,197094	37	33,03571343	0,187919463
4	abogado	0,175484	29	25,89285672	0,220357942
5	doctor	0,162642	30	26,78571343	0,253914989
6	camarero	0,14971	31	27,67857015	0,288590604
7	fontanero	0,130173	18	16,07142836	0,308724832
8	actor	0,11486	28	25	0,340044743
9	estudiante	0,1071	17	15,17857164	0,359060403
10	empleado	0,101295	19	16,96428508	0,380313199
11	secretario	0,098878	18	16,07142836	0,400447427
12	cantante	0,098417	26	23,21428508	0,429530201
13	entrenador	0,084874	16	14,28571492	0,447427293
14	policía	0,08101	17	15,17857164	0,466442953
15	enfermero	0,075385	13	11,60714254	0,48098434
16	bombero	0,074168	12	10,71428582	0,494407159
17	peluquero	0,072462	15	13,39285672	0,511185682
18	director	0,07222	15	13,39285672	0,527964206
19	traductor	0,067372	15	13,39285672	0,544742729
20	canguro	0,062967	13	11,60714254	0,559284116
21	obrero	0,057789	9	8,035714179	0,56935123
22	pintor	0,051494	12	10,71428582	0,582774049
23	cuidador	0,051317	9	8,035714179	0,592841163
24	escritor	0,050478	14	12,5	0,608501119
25	guía-de-turismo	0,044641	9	8,035714179	0,618568233
26	maestro	0,043053	6	5,35714291	0,625279642
27	ama-de-casa	0,042341	10	8,928571641	0,636465324
28	piloto	0,041793	8	7,142857462	0,64541387
29	vendedor	0,037778	8	7,142857462	0,654362416
30	jefe	0,035977	8	7,142857462	0,663310962
31	periodista	0,03585	10	8,928571641	0,674496644
32	jardinero	0,035303	6	5,35714291	0,681208054
33	albañil	0,03469	7	6,25	0,689038031
34	futbolista	0,033659	7	6,25	0,696868009
35	oficina	0,033568	8	7,142857462	0,705816555
36	bailarín	0,031666	10	8,928571641	0,717002237
37	dentista	0,029778	6	5,35714291	0,723713647

38	farmacéutico	0,029035	7	6,25	0,731543624
39	veterinario	0,02871	7	6,25	0,739373602
40	carnicero	0,028515	6	5,35714291	0,746085011
41	banco	0,027348	4	3,571428731	0,750559284
42	juez	0,026093	4	3,571428731	0,755033557
43	jugador	0,025719	7	6,25	0,762863535
44	actriz	0,025149	8	7,142857462	0,771812081
45	chef	0,025022	7	6,25	0,779642058
46	dependiente	0,024802	5	4,464285821	0,785234899
47	político	0,022188	7	6,25	0,793064877
48	barista	0,021708	5	4,464285821	0,798657718
49	modista	0,021464	4	3,571428731	0,803131991
50	asistente	0,020136	4	3,571428731	0,807606264
51	modelo	0,019113	4	3,571428731	0,812080537
52	campesino	0,018558	3	2,678571455	0,815436242
53	trabajo	0,018304	3	2,678571455	0,818791946
54	autor	0,017768	4	3,571428731	0,823266219
55	ingeniero	0,017722	3	2,678571455	0,826621924
56	sastre	0,017217	3	2,678571455	0,829977629
57	señora-de-la-limpieza	0,01655	4	3,571428731	0,834451902
58	trabajador	0,015953	3	2,678571455	0,837807606
59	enseñante	0,015626	2	1,785714366	0,840044743
60	informático	0,015371	3	2,678571455	0,843400447
61	intérprete	0,014393	4	3,571428731	0,84787472
62	panadero	0,014348	3	2,678571455	0,851230425
63	arquitecto	0,014348	3	2,678571455	0,85458613
64	tienda	0,013144	3	2,678571455	0,857941834
65	mecánico	0,013144	3	2,678571455	0,861297539
66	alcalde	0,012915	3	2,678571455	0,864653244
67	cirujano	0,012697	2	1,785714366	0,86689038
68	técnico	0,012601	3	2,678571455	0,870246085
69	químico	0,011722	2	1,785714366	0,872483221
70	chofer	0,010997	2	1,785714366	0,874720358
71	editor	0,010152	2	1,785714366	0,876957494
72	socorrista	0,00957	2	1,785714366	0,879194631
73	cajero	0,009065	2	1,785714366	0,881431767
74	estudio	0,008929	1	0,892857183	0,882550336
75	agricultor	0,008929	1	0,892857183	0,883668904
76	atleta	0,008717	3	2,678571455	0,887024609

77	presentador	0,008637	3	2,678571455	0,890380313
78	artesano	0,008628	2	1,785714366	0,89261745
79	empresario	0,008288	2	1,785714366	0,894854586
80	fotógrafo	0,007775	3	2,678571455	0,898210291
81	notario	0,007733	1	0,892857183	0,899328859
82	enseñar	0,007638	2	1,785714366	0,901565996
83	redactor	0,007538	2	1,785714366	0,903803132
84	músico	0,006983	3	2,678571455	0,907158837
85	bailar	0,006698	1	0,892857183	0,908277405
86	charcutero	0,006698	1	0,892857183	0,909395973
87	alguacil	0,006698	1	0,892857183	0,910514541
88	repcionista	0,006698	1	0,892857183	0,91163311
89	lector	0,006698	1	0,892857183	0,912751678
90	cómico	0,006188	2	1,785714366	0,914988814
91	camionero	0,005801	1	0,892857183	0,916107383
92	cura	0,005801	1	0,892857183	0,917225951
93	modelista	0,005801	1	0,892857183	0,918344519
94	mucamo	0,005801	1	0,892857183	0,919463087
95	instructor	0,005801	1	0,892857183	0,920581655
96	estilista	0,005801	1	0,892857183	0,921700224
97	traficante	0,005801	1	0,892857183	0,922818792
98	preste	0,005801	1	0,892857183	0,92393736
99	comerciante	0,005801	1	0,892857183	0,925055928
100	florista	0,005713	2	1,785714366	0,927293065
101	escuela	0,00536	2	1,785714366	0,929530201
102	fábrica	0,005024	1	0,892857183	0,93064877
103	monja	0,005024	1	0,892857183	0,931767338
104	educador	0,005024	1	0,892857183	0,932885906
105	carpintero	0,005024	1	0,892857183	0,934004474
106	cardiólogo	0,005024	1	0,892857183	0,935123043
107	azafata	0,004642	2	1,785714366	0,937360179
108	payaso	0,004351	1	0,892857183	0,938478747
109	tabaquero	0,004351	1	0,892857183	0,939597315
110	taxista	0,004351	1	0,892857183	0,940715884
111	logopeda	0,004351	1	0,892857183	0,941834452
112	plomero	0,004351	1	0,892857183	0,94295302
113	programador	0,004351	1	0,892857183	0,944071588
114	dueño	0,004351	1	0,892857183	0,945190157
115	rey	0,004351	1	0,892857183	0,946308725
116	bibliotecario	0,004351	1	0,892857183	0,947427293

117	botón	0,004351	1	0,892857183	0,948545861
118	ginecólogo	0,004351	1	0,892857183	0,94966443
119	biólogo	0,004298	2	1,785714366	0,951901566
120	presidente	0,003769	1	0,892857183	0,953020134
121	seguro	0,003769	1	0,892857183	0,954138702
122	gestor	0,003769	1	0,892857183	0,955257271
123	esteticista	0,003769	1	0,892857183	0,956375839
124	aeronauta	0,003769	1	0,892857183	0,957494407
125	emprendedor	0,003769	1	0,892857183	0,958612975
126	nadador	0,003769	1	0,892857183	0,959731544
127	escultor	0,003769	1	0,892857183	0,960850112
128	neurólogo	0,003769	1	0,892857183	0,96196868
129	cocinar	0,003264	1	0,892857183	0,963087248
130	florero	0,003264	1	0,892857183	0,964205817
131	limpiar	0,003264	1	0,892857183	0,965324385
132	cuidador-de-perro	0,003264	1	0,892857183	0,966442953
133	economista	0,003264	1	0,892857183	0,967561521
134	dirigente	0,003264	1	0,892857183	0,968680089
135	tenista	0,003264	1	0,892857183	0,969798658
136	maquillador	0,003264	1	0,892857183	0,970917226
137	representante	0,003264	1	0,892857183	0,972035794
138	animador	0,003264	1	0,892857183	0,973154362
139	frutero	0,003264	1	0,892857183	0,974272931
140	andrólogo	0,003264	1	0,892857183	0,975391499
141	negocios	0,002827	1	0,892857183	0,976510067
142	gobernador	0,002827	1	0,892857183	0,977628635
143	ordenar	0,002827	1	0,892857183	0,978747204
144	zapatero	0,002827	1	0,892857183	0,979865772
145	portero	0,002827	1	0,892857183	0,98098434
146	heladero	0,002827	1	0,892857183	0,982102908
147	pastor	0,002449	1	0,892857183	0,983221477
148	hidráulico	0,002449	1	0,892857183	0,984340045
149	cartonero	0,002449	1	0,892857183	0,985458613
150	psicólogo	0,002449	1	0,892857183	0,986577181
151	restaurante	0,002121	1	0,892857183	0,987695749
152	militar	0,002121	1	0,892857183	0,988814318
153	tallador	0,002121	1	0,892857183	0,989932886
154	fisioterapeuta	0,001837	1	0,892857183	0,991051454
155	guionista	0,001591	1	0,892857183	0,992170022

156	osteópata	0,001591	1	0,892857183	0,993288591
157	guardia	0,001591	1	0,892857183	0,994407159
158	sirviente	0,001378	1	0,892857183	0,995525727
159	ayudante	0,001378	1	0,892857183	0,996644295
160	pescador	0,001378	1	0,892857183	0,997762864
161	hospital	0,001193	1	0,892857183	0,998881432
162	dibujador	0,001193	1	0,892857183	1

ANEXO IV – LISTADO ALFABÉTICO DE LÉXICO DISPONIBLE

(CORPUS GENERAL)

CI01, “las partes del cuerpo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	alto	0,0011816	1	0,999999978	0,999261448
2	aparato	0,00343744	1	0,999999978	0,989660266
3	azul	0,0051833	1	0,999999978	0,973412112
4	bajo	0,00128276	1	0,999999978	0,997784343
5	barriga	0,14163156	24	23,99999946	0,783604136
6	bigote	0,0128414	3	2,999999933	0,947562777
7	boca	0,41181397	65	64,99999762	0,390694239
8	brazo	0,4177554	63	62,99999952	0,342688331
9	cabellos	0,06044529	10	10,00000015	0,891432792
10	cabeza	0,53324872	65	64,99999762	0,185376662
11	cadera	0,01580146	5	5,000000075	0,940177253
12	cara	0,33971396	45	44,99999881	0,56351551
13	carne	0,00343744	1	0,999999978	0,986706056
14	cejas	0,06678414	16	15,99999964	0,875184638
15	cerebro	0,0128209	4	3,999999911	0,950516987
16	cintura	0,03574285	9	9,000000358	0,922451994
17	codos	0,05182045	10	10,00000015	0,898818316
18	corazón	0,09316909	20	20,0000003	0,812407681
19	coronilla	0,00781587	1	0,999999978	0,964549483
20	cuello	0,21489586	34	34,00000036	0,658788774
21	cuerpo	0,00562705	1	0,999999978	0,971196455
22	cuerpo-humano	0,00848499	1	0,999999978	0,962333826
23	culo	0,01081932	4	3,999999911	0,957902511
24	dedo	0,39431173	67	67,00000167	0,440177253
25	dientes	0,17093405	35	34,9999994	0,744460857
26	espalda	0,26327646	50	50	0,600443131
27	estómago	0,08756647	20	20,0000003	0,840472674
28	feo	0,0040512	1	0,999999978	0,983013294
29	frente	0,02417241	8	7,999999821	0,928360414
30	garganta	0,07312658	14	14,00000006	0,863367799
31	genitales	0,0040512	1	0,999999978	0,983751846
32	gordo	0,00373173	1	0,999999978	0,984490399
33	guapo	0,00128276	1	0,999999978	0,99704579
34	hígado	0,00719952	1	0,999999978	0,96676514

35	hombro	0,09187716	18	18,00000072	0,825701625
36	hueso	0,00247479	1	0,999999978	0,991137371
37	intestino	0,00343744	1	0,999999978	0,988921713
38	labios	0,09925174	19	18,99999976	0,797636632
39	lengua	0,16308302	29	28,99999917	0,765878877
40	mano	0,68998682	91	91,00000262	0,137370753
41	marrón	0,00562705	1	0,999999978	0,970457903
42	mejilla	0,0381039	7	7,00000003	0,915805022
43	mentón	0,00193427	1	0,999999978	0,994091581
44	muelas	0,00343744	1	0,999999978	0,988183161
45	muñeca	0,01443267	4	3,999999911	0,94534712
46	músculo	0,01701428	4	3,999999911	0,933530281
47	muslo	0,00537332	2	1,999999955	0,97267356
48	nalgas	0,01082335	3	2,999999933	0,954948301
49	nariz	0,37995452	61	61,00000143	0,485228951
50	naso	0,01971822	3	2,999999933	0,930576071
51	neurona	0,00247479	1	0,999999978	0,991875923
52	oído	0,00781587	1	0,999999978	0,963810931
53	oír	0,00663176	1	0,999999978	0,968242245
54	ojos	0,71077543	95	94,99999881	0,070162482
55	ombligo	0,00193427	1	0,999999978	0,994830133
56	orejas	0,22095841	45	44,99999881	0,633677991
57	panza	0,01036588	2	1,999999955	0,960856721
58	pecho	0,06640178	12	11,99999973	0,884047267
59	pelirrojo	0,00719952	1	0,999999978	0,967503693
60	pelo	0,36899704	61	61,00000143	0,53028065
61	pene	0,00437949	2	1,999999955	0,980797637
62	pestañas	0,03943514	8	7,999999821	0,910635155
63	pezón	0,00227963	1	0,999999978	0,993353028
64	pie	0,52412736	81	81,00000024	0,245199409
65	piel	0,08052495	17	17,00000018	0,853028065
66	pierna	0,47413734	69	68,99999976	0,296159527
67	pulmones	0,00343744	1	0,999999978	0,987444609
68	pulso	0,01629602	4	3,999999911	0,93648449
69	respirar	0,00343744	1	0,999999978	0,985228951
70	rodilla	0,21331339	39	38,99999857	0,687592319
71	rubio	0,01073254	2	1,999999955	0,959379616
72	salud	0,00425479	2	1,999999955	0,982274742
73	sangre	0,01127292	3	2,999999933	0,952732644
74	seno	0,00649789	2	1,999999955	0,96971935

75	sien	0,00343744	1	0,999999978	0,985967504
76	sobrecejos	0,00501653	2	1,999999955	0,977104874
77	tacto	0,00247479	1	0,999999978	0,992614476
78	talón	0,0015118	1	0,999999978	0,995568685
79	testa	0,00439803	1	0,999999978	0,979320532
80	tetas	0,00516145	2	1,999999955	0,97562777
81	tez	0,00781587	1	0,999999978	0,963072378
82	tobillo	0,04391922	8	7,999999821	0,904726736
83	tórax	0,0092114	1	0,999999978	0,961595273
84	trasero	0,00291667	1	0,999999978	0,990398818
85	tripa	0,01542782	3	2,999999933	0,94239291
86	uñas	0,20997067	42	41,99999869	0,718611521
87	vagina	0,00440906	2	1,999999955	0,978581979
88	ver	0,00100259	1	0,999999978	1
89	verde	0,00139258	1	0,999999978	0,996307238
90	vientre	0,00721757	2	1,999999955	0,966026588
91	vista	0,0011816	1	0,999999978	0,998522895

CI02, "la ropa"

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abrigo	0,126946315	23	20,53571492	0,689001264
2	algodón	0,005332824	2	1,785714366	0,96965866
3	anillo	0,01486198	5	4,464285821	0,927939317
4	anorak	0,005949898	1	0,892857183	0,967130215
5	bajero	0,003024977	1	0,892857183	0,992414665
6	bañador	0,019839035	5	4,464285821	0,912768647
7	bikini	0,012454879	4	3,571428731	0,936788875
8	blusa	0,061510492	11	9,821428359	0,756005057
9	bolsillo	0,005196992	1	0,892857183	0,970922882
10	bolso	0,058352523	11	9,821428359	0,769911504
11	botas	0,037500799	8	7,142857462	0,8482933
12	bragas	0,005196992	1	0,892857183	0,975979772
13	bufanda	0,182538986	34	30,35714328	0,585335019
14	calcetines	0,14323926	29	25,89285672	0,621997472
15	calzados	0,005196992	1	0,892857183	0,972187105
16	calzas	0,002307847	1	0,892857183	0,996207332
17	calzoncillos	0,017906409	4	3,571428731	0,917825537
18	camisa	0,28956607	44	39,28571343	0,321112516
19	camiseta	0,481934398	70	62,5	0,088495575
20	chaleco	0,008928572	1	0,892857183	0,955752212
21	chanclas	0,012055307	3	2,678571455	0,940581542
22	chancletas	0,001760727	1	0,892857183	1
23	chándal	0,005196992	1	0,892857183	0,973451327
24	chaqueta	0,255687803	47	41,96428657	0,437420986
25	cinturón	0,039425083	10	8,928571641	0,825537295
26	collar	0,012042059	3	2,678571455	0,94437421
27	corbata	0,020279812	3	2,678571455	0,906447535
28	falda	0,252731293	42	37,5	0,490518331
29	gafas	0,055289257	11	9,821428359	0,798988622
30	gafas-de-sol	0,00310404	2	1,785714366	0,988621997
31	gorro	0,033433791	10	8,928571641	0,872313527
32	guantes	0,03516937	9	8,035714179	0,859671302
33	jersey	0,259401023	45	40,17857015	0,378002528
34	lente-de-contacto	0,005196992	1	0,892857183	0,97471555
35	maleta	0,008928572	1	0,892857183	0,95448799
36	mantilla	0,003964944	1	0,892857183	0,98482933
37	medias	0,030320408	6	5,35714291	0,879898862
38	minifalda	0,062010843	11	9,821428359	0,742098609
39	mochila	0,009413114	2	1,785714366	0,953223767

40	pantalones	0,409810811	69	61,6071403	0,265486726
41	pañuelo	0,007798741	1	0,892857183	0,960809102
42	pendientes	0,003024977	1	0,892857183	0,98988622
43	pijama	0,021087589	4	3,571428731	0,902654867
44	playera	0,003964944	1	0,892857183	0,983565107
45	polera	0,00681188	1	0,892857183	0,96460177
46	ponerse	0,012008871	2	1,785714366	0,946902655
47	pulsera	0,015028615	3	2,678571455	0,921618205
48	quitarse	0,004539359	1	0,892857183	0,97977244
49	rebeca	0,007965708	2	1,785714366	0,95954488
50	reloj	0,026498398	6	5,35714291	0,887484197
51	remera	0,011469782	3	2,678571455	0,950695322
52	ropa	0,003024977	1	0,892857183	0,991150442
53	ropa-interior	0,007564336	2	1,785714366	0,963337547
54	saco	0,008928572	1	0,892857183	0,957016435
55	sandalias	0,004180116	2	1,785714366	0,982300885
56	seda	0,004658003	2	1,785714366	0,978508217
57	sombrero	0,038218297	10	8,928571641	0,83817952
58	sudadera	0,057073303	12	10,71428582	0,785082174
59	sujetador	0,022082914	8	7,142857462	0,897597977
60	tacones	0,068812624	14	12,5	0,728192162
61	tenis	0,00201581	1	0,892857183	0,997471555
62	top	0,001760727	1	0,892857183	0,998735777
63	traje	0,097627945	17	15,17857164	0,710493047
64	traje-de-baño	0,012797104	3	2,678571455	0,931731985
65	vaqueros	0,240881488	41	36,60714328	0,542351454
66	vestido	0,136809915	30	26,78571343	0,659924147
67	vestirse	0,005949898	1	0,892857183	0,965865992
68	zapata	0,003964944	1	0,892857183	0,986093552
69	zapatero	0,002307847	1	0,892857183	0,99494311
70	zapatilla-de-deporte	0,003024977	1	0,892857183	0,993678887
71	zapatillas	0,041901674	11	9,821428359	0,81289507
72	zapatos	0,439514309	71	63,3928597	0,178255373

CI03, “partes de la casa (sin muebles)”

N°	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	altillo	0,007480752	1	0,892857183	0,967213115
2	aseo	0,010068518	2	1,785714366	0,954098361
3	atelier	0,00525136	1	0,892857183	0,981967213
4	balcón	0,031048205	9	8,035714179	0,927868852
5	cámara	0,041195538	7	6,25	0,9
6	cantina	0,007488422	2	1,785714366	0,96557377
7	cocina	0,680141151	94	83,92857313	0,154098361
8	comedor	0,058273014	12	10,71428582	0,798360656
9	corredor	0,008086188	2	1,785714366	0,962295082
10	cuarto	0,174641341	27	24,10714328	0,580327869
11	cuarto-de-baño	0,614644945	91	81,25	0,303278689
12	cuarto-de-estar	0,040128678	8	7,142857462	0,913114754
13	cuarto-matrimonial	0,00525136	1	0,892857183	0,983606557
14	cuarto-trasero	0,006274131	2	1,785714366	0,980327869
15	cueva	0,001275199	1	0,892857183	1
16	despacho	0,016280396	3	2,678571455	0,945901639
17	dormitorio	0,06136163	10	8,928571641	0,778688525
18	elevador	0,006567967	2	1,785714366	0,97704918
19	entrada	0,04129323	11	9,821428359	0,88852459
20	escalera	0,094394602	19	16,96428508	0,663934426
21	escritorio	0,0030886	1	0,892857183	0,991803279
22	estudio	0,054905318	13	11,60714254	0,819672131
23	garaje	0,14945288	32	28,57142985	0,632786885
24	habitación	0,335915476	49	43,75	0,383606557
25	ingreso	0,010459727	3	2,678571455	0,950819672
26	jardín	0,226448298	45	40,17857015	0,536065574
27	lavandería	0,002168145	1	0,892857183	0,998360656
28	librería	0,003686365	1	0,892857183	0,98852459
29	mansarda	0,006987588	2	1,785714366	0,970491803
30	muro	0,006987588	2	1,785714366	0,973770492
31	oficina	0,021971243	4	3,571428731	0,940983607
32	pared	0,028426496	4	3,571428731	0,93442623
33	pasillo	0,043551773	9	8,035714179	0,850819672
34	patio	0,062821046	13	11,60714254	0,762295082
35	piscina	0,0030886	1	0,892857183	0,993442623
36	piso	0,046888702	10	8,928571641	0,836065574
37	portal	0,004399822	1	0,892857183	0,985245902
38	pórtico	0,0030886	1	0,892857183	0,990163934
39	puerta	0,076693319	15	13,39285672	0,714754098

40	sala	0,004399822	1	0,892857183	0,986885246
41	salón	0,265763253	48	42,85714328	0,462295082
42	salón-de-baile	0,002587766	1	0,892857183	0,996721311
43	sobresuelo	0,0030886	1	0,892857183	0,995081967
44	sótano	0,00958765	3	2,678571455	0,959016393
45	techo	0,079044521	16	14,28571492	0,690163934
46	terraza	0,042107698	12	10,71428582	0,870491803
47	ventana	0,073338203	16	14,28571492	0,740983607

CI04, “los muebles de la casa”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	alfombra	0,026155	8	7,142857	0,930303
2	almohada	0,007481	1	0,892857	0,980303
3	armario	0,313539	54	48,21429	0,512121
4	banco	0,002168	1	0,892857	0,99697
5	bañera	0,003089	1	0,892857	0,995455
6	baño	0,003686	1	0,892857	0,989394
7	caja	0,007481	1	0,892857	0,977273
8	cajón	0,060212	13	11,60714	0,80303
9	cama	0,510855	76	67,85714	0,115152
10	cenicero	0,003089	1	0,892857	0,993939
11	cocina	0,031606	4	3,571429	0,909091
12	cortina	0,007481	1	0,892857	0,978788
13	cuadro	0,010759	4	3,571429	0,968182
14	despensa	0,0044	1	0,892857	0,987879
15	despertador	0,002168	1	0,892857	0,998485
16	diván	0,01794	4	3,571429	0,95
17	ducha	0,121397	25	22,32143	0,595455
18	escalera	0,003089	1	0,892857	0,990909
19	escritorio	0,059886	14	12,5	0,824242
20	espejo	0,062708	13	11,60714	0,783333
21	estantería	0,049966	9	8,035714	0,890909
22	flexo	0,006268	1	0,892857	0,984848
23	frigorífico	0,067602	12	10,71429	0,763636
24	grifo	0,011519	2	1,785714	0,962121
25	guardarropa	0,030218	6	5,357143	0,918182
26	hielera	0,003089	1	0,892857	0,992424
27	horno	0,075149	18	16,07143	0,745455
28	lámpara	0,107842	24	21,42857	0,631818
29	lavabo	0,009954	2	1,785714	0,971212
30	lavadora	0,075863	15	13,39286	0,718182
31	lavavajillas	0,037307	8	7,142857	0,90303
32	librería	0,009651	2	1,785714	0,974242
33	luz	0,001275	1	0,892857	1
34	mesa	0,440463	73	65,17857	0,339394
35	mesilla-de-noche	0,101045	22	19,64286	0,695455
36	mueble	0,026061	4	3,571429	0,936364
37	nevera	0,053642	13	11,60714	0,843939
38	papelera	0,008929	1	0,892857	0,975758
39	puerta	0,050938	11	9,821428	0,877273

40	radio	0,00632	2	1,785714	0,983333
41	refrigerador	0,019333	5	4,464286	0,943939
42	reloj	0,016985	4	3,571429	0,959091
43	retrato	0,0044	1	0,892857	0,986364
44	ropero	0,017857	2	1,785714	0,95303
45	silla	0,458743	75	66,96429	0,228788
46	sillón	0,154345	30	26,78571	0,557576
47	sofá	0,345038	60	53,57143	0,430303
48	tele(visión)	0,106236	20	17,85714	0,662121
49	ventana	0,053459	11	9,821428	0,860606

CI05, “alimentos y bebidas”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aceite	0,092777	24	21,42857	0,631579
2	aceituna	0,020091	5	4,464286	0,877649
3	agua	0,658388	91	81,25	0,062201
4	agua-con-gas	0,020574	3	2,678571	0,874231
5	agua-sin-gas	0,011556	2	1,785714	0,921394
6	aguacate	0,024287	5	4,464286	0,855092
7	aguardiente	0,010045	2	1,785714	0,928913
8	ajillo	0,004628	1	0,892857	0,976077
9	ajo	0,035873	9	8,035714	0,806562
10	albaricoque	0,005921	1	0,892857	0,964457
11	albóndiga	0,005454	1	0,892857	0,966507
12	alcachofa	0,009377	2	1,785714	0,937116
13	alcohol	0,005024	1	0,892857	0,971292
14	alfajor	0,009377	2	1,785714	0,935748
15	almeja	0,002399	1	0,892857	0,994532
16	amargo	0,002827	1	0,892857	0,990431
17	ananás	0,014559	3	2,678571	0,902256
18	animal	0,006978	1	0,892857	0,954204
19	apio	0,003069	1	0,892857	0,989064
20	arroz	0,191119	35	31,25	0,386193
21	arroz-con-leche	0,004263	1	0,892857	0,97676
22	asado	0,003617	1	0,892857	0,984279
23	atún	0,011862	3	2,678571	0,91661
24	azafrán	0,008224	1	0,892857	0,9419
25	azúcar	0,1319	30	26,78571	0,503759
26	banana	0,017701	4	3,571429	0,888585
27	berenjena	0,024784	5	4,464286	0,848257
28	biscocho	0,002604	1	0,892857	0,991798
29	bocadillo	0,134283	21	18,75	0,483254
30	bocata	0,005024	1	0,892857	0,972659
31	bombón	0,001465	1	0,892857	0,999316
32	burrito	0,007576	1	0,892857	0,945318
33	caballo	0,001243	1	0,892857	1
34	café	0,139737	26	23,21429	0,4689
35	calabacín	0,034239	7	6,25	0,811347
36	calabaza	0,014096	3	2,678571	0,911141
37	caldo	0,009581	3	2,678571	0,932331
38	calimocho	0,006978	1	0,892857	0,958305

39	camarón	0,004899	2	1,785714	0,974026
40	caña	0,006978	1	0,892857	0,954887
41	cangrejo	0,003617	1	0,892857	0,984962
42	carne	0,182782	34	30,35714	0,409433
43	castaña	0,003332	1	0,892857	0,987013
44	cebolla	0,073468	17	15,17857	0,691046
45	cerdo	0,018939	5	4,464286	0,885851
46	cereales	0,010048	2	1,785714	0,927546
47	cereza	0,028818	6	5,357143	0,829802
48	cerveza	0,282548	47	41,96429	0,094327
49	chocolate	0,10957	24	21,42857	0,570062
50	chontaduro	0,005454	1	0,892857	0,967874
51	chorizo	0,02471	5	4,464286	0,851675
52	chupito	0,007576	1	0,892857	0,944634
53	churros	0,037384	7	6,25	0,80041
54	cidra	0,003617	1	0,892857	0,983595
55	ciruela	0,006978	1	0,892857	0,95352
56	<i>Coca-Cola</i>	0,043163	7	6,25	0,759398
57	cóctel	0,004628	1	0,892857	0,97471
58	coliflor	0,001875	1	0,892857	0,997949
59	conejo	0,00788	2	1,785714	0,943951
60	cordero	0,006978	1	0,892857	0,957621
61	crema	0,004263	1	0,892857	0,977444
62	crema-catalana	0,001875	1	0,892857	0,998633
63	dulce	0,03912	8	7,142857	0,795625
64	dulce	0,008891	2	1,785714	0,938483
65	dulce-de-leche	0,012754	3	2,678571	0,914559
66	embutido	0,00221	1	0,892857	0,995215
67	empanada	0,022933	3	2,678571	0,868079
68	ensalada	0,055134	12	10,71429	0,74231
69	espagueti	0,008224	1	0,892857	0,942584
70	espárrago	0,002604	1	0,892857	0,993165
71	espinaca	0,003927	1	0,892857	0,980178
72	flambeado	0,003927	1	0,892857	0,980861
73	fresa	0,06035	13	11,60714	0,718387
74	frijoles	0,026118	5	4,464286	0,842105
75	frito	0,016011	3	2,678571	0,894737
76	fruta	0,041837	9	8,035714	0,777854
77	galleta	0,060103	12	10,71429	0,726589
78	gamba	0,01171	3	2,678571	0,920027
79	garbanzos	0,01424	3	2,678571	0,909091

80	gaseoso	0,005454	1	0,892857	0,969925
81	gaspacho	0,076027	13	11,60714	0,679426
82	hamburguesa	0,013953	2	1,785714	0,912509
83	harina	0,044054	10	8,928572	0,754614
84	helado	0,009408	3	2,678571	0,934381
85	hielo	0,007576	1	0,892857	0,946685
86	hierbabuena	0,011839	2	1,785714	0,917977
87	huevo	0,125256	26	23,21429	0,521531
88	jamón	0,227918	39	34,82143	0,304853
89	judías	0,010524	3	2,678571	0,924812
90	jugo	0,006428	1	0,892857	0,958988
91	langosta	0,003332	1	0,892857	0,987697
92	leche	0,266612	49	43,75	0,161996
93	lechuga	0,028614	6	5,357143	0,833903
94	legumbres	0,0196	3	2,678571	0,879699
95	lentejas	0,006016	2	1,785714	0,96309
96	levadura	0,002035	1	0,892857	0,997266
97	licor	0,005454	1	0,892857	0,969241
98	limón	0,084624	20	17,85714	0,67054
99	macarrones	0,005454	1	0,892857	0,967191
100	mango	0,004628	1	0,892857	0,975393
101	manteca	0,004263	1	0,892857	0,978127
102	mantequilla	0,059106	11	9,821428	0,734108
103	manzana	0,259139	49	43,75	0,195489
104	mariscos	0,032217	9	8,035714	0,820916
105	mejillones	0,015789	4	3,571429	0,897471
106	melocotón	0,042864	9	8,035714	0,76555
107	melón	0,011503	2	1,785714	0,922761
108	mermelada	0,004071	2	1,785714	0,979494
109	miel	0,006138	2	1,785714	0,961722
110	mojito	0,009652	2	1,785714	0,93028
111	montadito	0,003332	1	0,892857	0,986329
112	naranja	0,241581	47	41,96429	0,227614
113	naranjada	0,005454	1	0,892857	0,968558
114	nata	0,007576	1	0,892857	0,947368
115	orégano	0,002827	1	0,892857	0,991114
116	paella	0,23787	37	33,03571	0,278195
117	palta	0,005921	1	0,892857	0,963773
118	pan	0,223488	40	35,71429	0,362269
119	papa	0,023037	4	3,571429	0,866029
120	parrilla	0,002604	1	0,892857	0,992481

121	pasta	0,170736	31	27,67857	0,451128
122	patata	0,109742	22	19,64286	0,553657
123	pavo	0,002399	1	0,892857	0,993848
124	pechuga-de-pollo	0,005454	1	0,892857	0,970608
125	pepino	0,019513	4	3,571429	0,882433
126	pera	0,007234	2	1,785714	0,952153
127	perejil	0,014445	4	3,571429	0,90499
128	pescado	0,225759	44	39,28571	0,334928
129	pimentón	0,007276	3	2,678571	0,949419
130	pimienta	0,023108	6	5,357143	0,863295
131	pimiento	0,042638	9	8,035714	0,771702
132	piña	0,022331	6	5,357143	0,87218
133	piña-colada	0,003332	1	0,892857	0,985646
134	plátano	0,104056	20	17,85714	0,596719
135	pollo	0,066688	13	11,60714	0,709501
136	postre	0,006978	1	0,892857	0,952837
137	puerco	0,005024	1	0,892857	0,971975
138	pulpo	0,008245	2	1,785714	0,941217
139	quesadilla	0,006978	1	0,892857	0,955571
140	queso	0,121596	25	22,32143	0,538619
141	refresco	0,014277	3	2,678571	0,90704
142	ron	0,01031	2	1,785714	0,926179
143	sal	0,102587	27	24,10714	0,615174
144	salado	0,003927	1	0,892857	0,982228
145	salchicha	0,006159	2	1,785714	0,960355
146	salmón	0,008525	2	1,785714	0,93985
147	sandía	0,006978	1	0,892857	0,956254
148	sangría	0,105035	19	16,96429	0,583049
149	seco	0,003927	1	0,892857	0,981545
150	solomillo	0,003069	1	0,892857	0,98838
151	sopa	0,015508	4	3,571429	0,900205
152	taco	0,006978	1	0,892857	0,956938
153	tapas	0,09256	17	15,17857	0,643199
154	tarta	0,039138	10	8,928572	0,790157
155	té	0,041666	8	7,142857	0,783322
156	tequila	0,02576	4	3,571429	0,844839
157	ternera	0,031778	7	6,25	0,825701
158	tinto-de-verano	0,007576	1	0,892857	0,946001
159	tomate	0,274992	50	44,64286	0,128503
160	torrezno	0,00221	1	0,892857	0,996582

161	tortilla	0,071329	14	12,5	0,700615
162	tostada	0,002827	1	0,892857	0,989747
163	uva	0,003927	1	0,892857	0,982912
164	vaca	0,007271	2	1,785714	0,950786
165	vegetal	0,00221	1	0,892857	0,995899
166	verdura	0,023299	6	5,357143	0,859193
167	vinagre	0,028282	7	6,25	0,838688
168	vino	0,24004	37	33,03571	0,252905
169	vino-blanco	0,033684	5	4,464286	0,814764
170	vino-tinto	0,049694	8	7,142857	0,747779
171	yogur	0,005896	2	1,785714	0,965824
172	zanahoria	0,090186	20	17,85714	0,656869
173	zumو	0,176995	30	26,78571	0,429938
174	zumو-de-fruta	0,016647	3	2,678571	0,890636
175	zumو-de-naranja	0,016414	3	2,678571	0,892686

CI06, “objetos colocados en la mesa para la comida”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aceite	0,038536	10	8,928572	0,816981
2	agua	0,01847	4	3,571429	0,928302
3	alimento	0,004148	1	0,892857	0,984906
4	bandeja	0,002827	1	0,892857	0,992453
5	bebida	0,013658	3	2,678571	0,939623
6	botella	0,280678	55	49,10714	0,669811
7	cacerola	0,005024	1	0,892857	0,977358
8	cambiar	0,004148	1	0,892857	0,983019
9	cazo	0,002334	1	0,892857	0,996226
10	comedor	0,006086	1	0,892857	0,964151
11	comer	0,002827	1	0,892857	0,990566
12	comida	0,003424	1	0,892857	0,988679
13	copa	0,02677	4	3,571429	0,858491
14	cubierto	0,005024	1	0,892857	0,975472
15	cuchara	0,349408	59	52,67857	0,566038
16	cucharada	0,020038	3	2,678571	0,920755
17	cucharilla	0,098805	18	16,07143	0,769811
18	cucharón	0,003424	1	0,892857	0,986792
19	cuchillo	0,453779	68	60,71429	0,454717
20	cuenca	0,007685	3	2,678571	0,958491
21	desayuno	0,002334	1	0,892857	0,99434
22	ensaladera	0,025116	7	6,25	0,884906
23	jarra	0,022015	6	5,357143	0,896226
24	mantel	0,034716	9	8,035714	0,850943
25	mesa	0,006086	1	0,892857	0,966038
26	olla	0,008902	3	2,678571	0,95283
27	pan	0,006981	2	1,785714	0,962264
28	papel	0,004754	2	1,785714	0,981132
29	pimienta	0,005232	2	1,785714	0,971698
30	plato	0,692113	94	83,92857	0,177358
31	plato-hondo	0,020105	5	4,464286	0,90566
32	plato-llano	0,020052	5	4,464286	0,915094
33	pulir	0,005024	1	0,892857	0,973585
34	sal	0,036828	9	8,035714	0,833962
35	sartén	0,015544	3	2,678571	0,933962
36	servilleta	0,060212	15	13,39286	0,798113
37	taza	0,025734	7	6,25	0,871698
38	tenedor	0,221298	35	31,25	0,735849
39	tomar	0,006086	1	0,892857	0,967925
40	vaso	0,493172	79	70,53571	0,326415

41	vinagre	0,010019	4	3,571429	0,94717
42	vino-blanco	0,001591	1	0,892857	1
43	vino-tinto	0,001927	1	0,892857	0,998113

CI07, “la cocina y sus utensilios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abrir	0,004478	1	0,892857183	0,945205479
2	agua	0,002246	1	0,892857183	0,983561644
3	apagar	0,005636	1	0,892857183	0,926027397
4	balanza	0,005636	1	0,892857183	0,939726027
5	balda	0,004478	1	0,892857183	0,953424658
6	bandeja	0,029133	6	5,35714291	0,82739726
7	batidor	0,041419	8	7,142857462	0,717808219
8	bol	0,008929	1	0,892857183	0,893150685
9	botella	0,030165	7	6,25	0,810958904
10	cacerola	0,011273	2	1,785714366	0,879452055
11	cafetera	0,002246	1	0,892857183	0,989041096
12	carbón	0,001127	1	0,892857183	1
13	cazo	0,012022	3	2,678571455	0,873972603
14	cazuela	0,018123	3	2,678571455	0,857534247
15	cocina	0,007094	1	0,892857183	0,906849315
16	cocinar	0,009921	2	1,785714366	0,890410959
17	comedor	0,007094	1	0,892857183	0,909589041
18	comida	0,003558	1	0,892857183	0,961643836
19	congelador	0,007094	1	0,892857183	0,912328767
20	cubierto	0,003558	1	0,892857183	0,97260274
21	cuchara	0,135013	24	21,42857164	0,350684932
22	cucharilla	0,0144	3	2,678571455	0,865753425
23	cuchillo	0,155303	27	24,10714328	0,284931507
24	cuenca	0,040143	6	5,35714291	0,734246575
25	despensa	0,002827	1	0,892857183	0,975342466
26	enharinar	0,007094	1	0,892857183	0,917808219
27	ensaladera	0,033472	7	6,25	0,775342466
28	envase	0,008929	1	0,892857183	0,895890411
29	escudilla	0,001785	1	0,892857183	0,991780822
30	escurridor	0,004478	1	0,892857183	0,950684932
31	espátula	0,005636	1	0,892857183	0,942465753
32	esponja	0,003558	1	0,892857183	0,964383562
33	espumadera	0,005636	1	0,892857183	0,928767123
34	estufa	0,005636	1	0,892857183	0,931506849
35	frigorífico	0,108946	19	16,96428508	0,493150685
36	fuego	0,03092	6	5,35714291	0,791780822

37	grasera	0,005636	1	0,892857183	0,936986301
38	grifo	0,024003	4	3,571428731	0,838356164
39	guantes	0,007094	1	0,892857183	0,904109589
40	hervidor	0,002827	1	0,892857183	0,978082192
41	horno	0,389606	53	47,32142985	0,145205479
42	(horno)microondas	0,056066	9	8,035714179	0,679452055
43	jabón	0,003558	1	0,892857183	0,969863014
44	jarra	0,002246	1	0,892857183	0,98630137
45	lavaplatos	0,004478	1	0,892857183	0,947945205
46	lavavajillas	0,090867	15	13,39285672	0,583561644
47	libro-de-cocina	0,003558	1	0,892857183	0,967123288
48	licuadora	0,001418	1	0,892857183	0,997260274
49	mesa	0,063249	11	9,821428359	0,654794521
50	mezclar	0,010652	2	1,785714366	0,884931507
51	molde	0,002827	1	0,892857183	0,980821918
52	nevera	0,10836	18	16,07142836	0,542465753
53	olla	0,084341	15	13,39285672	0,624657534
54	paella	0,020767	4	3,571428731	0,849315068
55	papel	0,005636	1	0,892857183	0,923287671
56	placa	0,005636	1	0,892857183	0,934246575
57	plato	0,172674	24	21,42857164	0,210958904
58	recipiente	0,007094	1	0,892857183	0,915068493
59	refrigerador	0,04216	6	5,35714291	0,695890411
60	rodillo	0,007094	1	0,892857183	0,920547945
61	sartén	0,110797	17	15,17857164	0,397260274
62	taza	0,007421	2	1,785714366	0,901369863
63	tenedor	0,036962	8	7,142857462	0,756164384
64	tostador	0,001785	1	0,892857183	0,994520548
65	trapo	0,003722	2	1,785714366	0,95890411
66	vaso	0,109405	16	14,28571492	0,44109589

CI08, “la escuela: muebles y materiales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	agenda	0,005908215	3	2,678571455	0,96762142
2	alumno	0,007887587	2	1,785714366	0,95890411
3	apuntes	0,009584014	3	2,678571455	0,948941469
4	armario	0,025492486	7	6,25	0,904109589
5	aula	0,011832476	3	2,678571455	0,942714819
6	boleta	0,004147849	1	0,892857183	0,97758406
7	bolí(grafo)	0,36574918	64	57,1428597	0,325031133
8	bolso	0,045729965	9	8,035714179	0,782067248
9	borrador	0,116734877	24	21,42857164	0,712328767
10	calculador	0,01681616	3	2,678571455	0,927770859
11	cartel	0,00104351	1	0,892857183	1
12	cátedra	0,029776724	5	4,464285821	0,879202989
13	celo	0,006570499	1	0,892857183	0,96388543
14	clase	0,02455434	5	4,464285821	0,910336239
15	color	0,036044519	8	7,142857462	0,866749689
16	compañero	0,002618469	1	0,892857183	0,98505604
17	computador	0,021984624	4	3,571428731	0,915317559
18	copiadora	0,003052384	1	0,892857183	0,98132005
19	cuaderno	0,344558567	67	59,82142687	0,480697385
20	deberes	0,00192692	1	0,892857183	0,99128269
21	diccionario	0,029421568	8	7,142857462	0,889165629
22	edificio	0,006570499	1	0,892857183	0,9626401
23	ejercicio	0,003558206	1	0,892857183	0,97882939
24	escribir	0,001418013	1	0,892857183	0,99626401
25	escritorio	0,040309761	6	5,35714291	0,809464508
26	estuche	0,057638772	16	14,28571492	0,770859278
27	estudiante	0,002618469	1	0,892857183	0,98381071
28	fotocopia	0,005236938	2	1,785714366	0,97260274
29	goma	0,044203345	10	8,928571641	0,794520548
30	goma-de-borrar	0,005236938	2	1,785714366	0,97011208
31	hoja	0,039687119	10	8,928571641	0,821917808
32	horario	0,001216433	1	0,892857183	0,99875467
33	impresora	0,002246237	1	0,892857183	0,98879203
34	ir	0,008928572	1	0,892857183	0,95392279
35	lámpara	0,001418013	1	0,892857183	0,99501868
36	lápiz	0,348755896	58	51,78571343	0,397260274
37	libro	0,388528198	76	67,8571403	0,094645081

38	luz	0,001652996	1	0,892857183	0,99252802
39	mapa	0,014010834	4	3,571428731	0,932752179
40	mapamundi	0,001216433	1	0,892857183	0,99750934
41	marcador	0,027966425	5	4,464285821	0,895392279
42	mesa	0,384773791	59	52,67857313	0,168119552
43	mochila	0,195003152	38	33,92857015	0,597758406
44	ordenador	0,178551197	38	33,92857015	0,645080946
45	papel	0,140372336	30	26,78571343	0,682440847
46	papelera	0,031376302	5	4,464285821	0,872976339
47	pastel	0,01851481	3	2,678571455	0,924034869
48	pegamento	0,002246237	1	0,892857183	0,98630137
49	perchero	0,008254929	2	1,785714366	0,95641345
50	pincel	0,003052384	1	0,892857183	0,98256538
51	pinchar	0,00192692	1	0,892857183	0,99003736
52	pizarra	0,307754308	56	50	0,550435866
53	pluma	0,01307642	2	1,785714366	0,938978829
54	profesor	0,013524048	3	2,678571455	0,936488169
55	proyector	0,001652996	1	0,892857183	0,99377335
56	puerta	0,009905555	2	1,785714366	0,945205479
57	pupitre	0,040602017	6	5,35714291	0,801992528
58	registro	0,002246237	1	0,892857183	0,9875467
59	regla	0,038989238	11	9,821428359	0,835616438
60	reloj	0,00900836	3	2,678571455	0,95267746
61	rotulador	0,069617681	16	14,28571492	0,732254047
62	sacapuntas	0,067909956	15	13,39285672	0,750933998
63	silla	0,379523367	62	55,3571403	0,245330012
64	sillón	0,003052384	1	0,892857183	0,98007472
65	subrayar	0,004147849	1	0,892857183	0,9750934
66	tabla	0,007659319	1	0,892857183	0,96014944
67	tablero	0,004147849	1	0,892857183	0,97633873
68	tableta	0,004835203	1	0,892857183	0,97384807
69	tijeras	0,037150998	8	7,142857462	0,845579078
70	tiza	0,036431495	9	8,035714179	0,856787049
71	ventana	0,021129509	4	3,571428731	0,920298879
72	volver	0,007659319	1	0,892857183	0,96139477

CI09, “iluminación y calefacción”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	acondicionador	0,006085586	1	0,892857183	0,973684211
2	aire	0,001313361	1	0,892857183	0,997076023
3	aire-acondicionado	0,058199514	9	8,035714179	0,640350877
4	antorcha	0,007371268	1	0,892857183	0,947368421
5	apagar	0,018256757	3	2,678571455	0,859649123
6	bombilla	0,06588643	10	8,928571641	0,584795322
7	calefacción	0,031161118	6	5,35714291	0,751461988
8	calefactor	0,007371268	1	0,892857183	0,932748538
9	calentador	0,037709363	6	5,35714291	0,704678363
10	calentamiento	0,006848778	2	1,785714366	0,953216374
11	calentar	0,028940191	6	5,35714291	0,769005848
12	caliente	0,020619711	4	3,571428731	0,81871345
13	calor	0,063645363	10	8,928571641	0,614035088
14	calorífero	0,007371268	1	0,892857183	0,941520468
15	calorífico	0,042270858	8	7,142857462	0,687134503
16	candela	0,006085586	1	0,892857183	0,961988304
17	central-nuclear	0,00502415	1	0,892857183	0,976608187
18	chimenea	0,024566591	4	3,571428731	0,780701754
19	circuito	0,004147849	1	0,892857183	0,98245614
20	corriente-eléctrica	0,008448539	2	1,785714366	0,926900585
21	electricidad	0,018422676	4	3,571428731	0,850877193
22	encender	0,008928572	1	0,892857183	0,921052632
23	enchufe	0,014742536	2	1,785714366	0,888888889
24	energía-eléctrica	0,011755685	2	1,785714366	0,906432749
25	energía-renovable	0,006085586	1	0,892857183	0,956140351
26	estrella	0,004147849	1	0,892857183	0,979532164
27	estufa	0,024186159	4	3,571428731	0,792397661
28	farola	0,01629984	2	1,785714366	0,865497076
29	flexo	0,014742536	2	1,785714366	0,894736842
30	frío	0,019967657	4	3,571428731	0,830409357
31	fuego	0,263790518	39	34,82142985	0,520467836
32	gas	0,007371268	1	0,892857183	0,938596491
33	hacer-calor	0,006085586	1	0,892857183	0,964912281
34	hacer-frío	0,007371268	1	0,892857183	0,935672515

35	hielo	0,001590831	1	0,892857183	0,994152047
36	hogar	0,011519117	2	1,785714366	0,912280702
37	horno	0,013952722	2	1,785714366	0,900584795
38	iluminación	0,002334015	1	0,892857183	0,99122807
39	interruptor	0,006085586	1	0,892857183	0,970760234
40	lámpara	0,627458036	80	71,42857313	0,233918129
41	lamparilla	0,000895168	1	0,892857183	1
42	led	0,007371268	1	0,892857183	0,944444444
43	leña	0,011109738	2	1,785714366	0,918128655
44	luz	0,422833234	59	52,67857313	0,406432749
45	madera	0,004147849	1	0,892857183	0,985380117
46	neón	0,015021139	4	3,571428731	0,877192982
47	nevera	0,007371268	1	0,892857183	0,929824561
48	noche	0,006085586	1	0,892857183	0,967836257
49	pantalla	0,006085586	1	0,892857183	0,959064327
50	pila	0,003424389	1	0,892857183	0,988304094
51	radiador	0,015014158	2	1,785714366	0,883040936
52	sol	0,071262822	12	10,71428582	0,555555556
53	temperatura	0,03762383	5	4,464285821	0,719298246
54	termo(sifón)	0,035461847	5	4,464285821	0,733918129
55	vela	0,048937883	8	7,142857462	0,66374269
56	ventana	0,023553282	5	4,464285821	0,807017544
57	ventilador	0,019542441	3	2,678571455	0,839181287

CI10, “la ciudad”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	acera	0,005636	1	0,892857183	0,964133219
2	aeropuerto	0,014971	3	2,678571455	0,887275833
3	agencia	0,00618	1	0,892857183	0,95559351
4	alcalde	0,015269	3	2,678571455	0,881298036
5	almacén	0,009913	2	1,785714366	0,931682323
6	antiguo	0,005141	1	0,892857183	0,965841161
7	aparcamiento	0,017839	4	3,571428731	0,85824082
8	aparcar	0,008608	3	2,678571455	0,939368061
9	apartamento	0,014262	4	3,571428731	0,890691716
10	árbol	0,103427	16	14,28571492	0,552519214
11	arte	0,005636	1	0,892857183	0,962425278
12	(auto)bús	0,241343	46	41,07142985	0,337318531
13	auto(móvil)	0,006775	1	0,892857183	0,953031597
14	autovía	0,008144	1	0,892857183	0,944491887
15	avenida	0,043776	6	5,35714291	0,746370623
16	ayuntamiento	0,103767	17	15,17857164	0,538855679
17	bajar	0,00296	1	0,892857183	0,990606319
18	banco	0,072099	13	11,60714254	0,647309991
19	bar	0,160773	37	33,03571343	0,464560205
20	barrio	0,025663	5	4,464285821	0,818958155
21	beber	0,003245	1	0,892857183	0,983774552
22	biblioteca	0,017311	5	4,464285821	0,865072588
23	bici(cleta)	0,073542	16	14,28571492	0,623398804
24	billete	0,003245	1	0,892857183	0,985482494
25	bloque-de-piso	0,00618	1	0,892857183	0,958155423
26	café	0,0027	1	0,892857183	0,993168232
27	cafetería	0,012117	2	1,785714366	0,908625107
28	calle	0,487558	61	54,46428657	0,052092229
29	caminar	0,002463	1	0,892857183	0,996584116
30	carnicería	0,012698	3	2,678571455	0,900939368
31	carretera	0,157223	24	21,42857164	0,485055508
32	carro	0,010081	2	1,785714366	0,928266439
33	cartel	0,00859	2	1,785714366	0,941076003
34	casa	0,268297	42	37,5	0,264730999
35	casco-antiguo	0,026451	4	3,571428731	0,808710504
36	catedral	0,010181	3	2,678571455	0,924850555
37	cementerio	0,007147	2	1,785714366	0,949615713

38	centro	0,061543	10	8,928571641	0,678052946
39	centro-comercial	0,015474	3	2,678571455	0,878736123
40	centro-histórico	0,004277	1	0,892857183	0,975234842
41	cine	0,073393	15	13,39285672	0,636208369
42	ciudadanía	0,00296	1	0,892857183	0,99146029
43	ciudadano	0,006804	2	1,785714366	0,951323655
44	coche	0,314487	52	46,42857015	0,096498719
45	colegio	0,008178	2	1,785714366	0,942783945
46	comer	0,003558	1	0,892857183	0,980358668
47	comprar	0,00296	1	0,892857183	0,988044406
48	correos	0,00618	1	0,892857183	0,956447481
49	cosmopolita	0,005141	1	0,892857183	0,969257045
50	cruce	0,023765	5	4,464285821	0,830913749
51	cruzar	0,012887	4	3,571428731	0,896669513
52	derecha	0,011784	3	2,678571455	0,91118702
53	despacho	0,005636	1	0,892857183	0,960717336
54	disco(teca)	0,026117	7	6,25	0,814688301
55	edificio	0,249073	39	34,82142985	0,298035867
56	escuela	0,272103	48	42,85714328	0,228864219
57	esmog	0,006775	1	0,892857183	0,952177626
58	esquina	0,019134	3	2,678571455	0,852263023
59	establecimiento	0,006775	1	0,892857183	0,953885568
60	estación	0,161481	33	29,46428657	0,432963279
61	estanco	0,009282	3	2,678571455	0,935952178
62	estatua	0,010128	2	1,785714366	0,926558497
63	estrada	0,016288	2	1,785714366	0,873612297
64	estudio	0,002463	1	0,892857183	0,994876174
65	fábrica	0,007649	2	1,785714366	0,9470538
66	farmacia	0,080797	16	14,28571492	0,599487617
67	flor	0,00296	1	0,892857183	0,988898377
68	florería	0,003245	1	0,892857183	0,986336465
69	floristería	0,002463	1	0,892857183	0,998292058
70	frutería	0,004277	1	0,892857183	0,9735269
71	fuelle	0,017284	3	2,678571455	0,8676345
72	galería	0,00296	1	0,892857183	0,987190436
73	gente	0,043614	8	7,142857462	0,753202391
74	gimnasio	0,010757	4	3,571428731	0,922288642
75	girar	0,004749	2	1,785714366	0,970964987
76	glorieta	0,004689	1	0,892857183	0,972672929

77	guardería	0,002246	1	0,892857183	0,999146029
78	habitación	0,011052	2	1,785714366	0,918872758
79	heladería	0,012127	3	2,678571455	0,906917165
80	historia	0,003901	1	0,892857183	0,978650726
81	histórico	0,017562	3	2,678571455	0,860802733
82	hospital	0,059606	14	12,5	0,69000854
83	hostal	0,004689	1	0,892857183	0,971818958
84	hotel	0,042336	9	8,035714179	0,76088813
85	iglesia	0,237779	42	37,5	0,373185312
86	instituto	0,020626	4	3,571428731	0,846285226
87	ir-de-compras	0,003901	1	0,892857183	0,977796755
88	izquierda	0,01265	4	3,571428731	0,904355252
89	jardín	0,064647	12	10,71428582	0,669513237
90	joyería	0,006775	1	0,892857183	0,954739539
91	lampión	0,008144	1	0,892857183	0,945345858
92	luz	0,005636	1	0,892857183	0,961571307
93	mercado	0,052902	11	9,821428359	0,736122972
94	metro	0,029728	6	5,35714291	0,801024765
95	metrópoli(s)	0,002463	1	0,892857183	0,997438087
96	moderno	0,007428	1	0,892857183	0,947907771
97	monumento	0,055358	11	9,821428359	0,709649872
98	moto(cicleta)	0,016193	3	2,678571455	0,87617421
99	movida	0,009913	2	1,785714366	0,929974381
100	municipalidad	0,005636	1	0,892857183	0,963279249
101	municipio	0,02858	5	4,464285821	0,80529462
102	muralla	0,008144	1	0,892857183	0,943637916
103	museo	0,091731	18	16,07142836	0,567890692
104	negocio	0,030525	5	4,464285821	0,795900939
105	oficina	0,053371	12	10,71428582	0,726729291
106	oficina-de-correos	0,003558	1	0,892857183	0,981212639
107	oficio	0,011389	2	1,785714366	0,915456874
108	palacio	0,074227	12	10,71428582	0,609735269
109	panadería	0,020371	4	3,571428731	0,84970111
110	parada	0,064902	14	12,5	0,659265585
111	parque	0,300974	49	43,75	0,187873612
112	pasear	0,015169	4	3,571428731	0,88471392
113	paseo	0,01283	2	1,785714366	0,898377455
114	pasillo	0,0027	1	0,892857183	0,992314261
115	paso-de-cebra	0,053877	8	7,142857462	0,71648164

116	peatón	0,019009	3	2,678571455	0,854824936
117	peluquería	0,024897	5	4,464285821	0,826643894
118	persona	0,031774	5	4,464285821	0,791631085
119	piscina	0,011104	2	1,785714366	0,917164816
120	piso	0,022679	4	3,571428731	0,842869342
121	pizzería	0,002049	1	0,892857183	1
122	playa	0,00618	1	0,892857183	0,957301452
123	plaza	0,225906	37	33,03571343	0,404782237
124	policía	0,023169	4	3,571428731	0,834329633
125	pueblo	0,005141	1	0,892857183	0,967549103
126	puente	0,035838	8	7,142857462	0,773697694
127	puerto	0,00296	1	0,892857183	0,989752348
128	rascacielos	0,046968	6	5,35714291	0,741246798
129	restaurante	0,109257	26	23,21428508	0,507258753
130	rincón	0,00618	1	0,892857183	0,959009394
131	río	0,017197	5	4,464285821	0,871904355
132	ruido	0,004277	1	0,892857183	0,974380871
133	salir	0,003245	1	0,892857183	0,984628523
134	semáforo	0,0367	7	6,25	0,766865927
135	señal-de-tráfico	0,005141	1	0,892857183	0,966695132
136	sitio	0,00618	1	0,892857183	0,959863365
137	supermercado	0,05543	12	10,71428582	0,700256191
138	tabaquería	0,005141	1	0,892857183	0,96498719
139	tapas	0,003901	1	0,892857183	0,976942784
140	taxi	0,011582	3	2,678571455	0,913748933
141	teatro	0,022814	6	5,35714291	0,839453459
142	terrace	0,002463	1	0,892857183	0,995730145
143	tienda	0,309836	58	51,78571343	0,146029035
144	torre	0,008929	1	0,892857183	0,936806149
145	trabajo	0,009538	2	1,785714366	0,933390265
146	tráfico	0,035151	8	7,142857462	0,78736123
147	transporte-público	0,025534	4	3,571428731	0,822374039
148	tranvía	0,035831	8	7,142857462	0,780529462
149	tren	0,106968	20	17,85714328	0,524338173
150	turista	0,013695	3	2,678571455	0,893253629
151	universidad	0,09025	21	18,75	0,585824082
152	vehículo	0,004277	1	0,892857183	0,976088813
153	vender	0,0027	1	0,892857183	0,994022203
154	verde	0,005141	1	0,892857183	0,968403074

155	viajar	0,003558	1	0,892857183	0,982920581
156	vivir	0,003558	1	0,892857183	0,98206661
157	zona-peatonal	0,003901	1	0,892857183	0,979504697

CI11, “el campo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abeja	0,003965	1	0,892857183	0,957664234
2	abuelo	0,00595	1	0,892857183	0,927007299
3	aceite	0,003025	1	0,892857183	0,96350365
4	aceituno	0,00595	1	0,892857183	0,925547445
5	aceña	0,002016	1	0,892857183	0,983941606
6	afuera	0,008929	1	0,892857183	0,883211679
7	agricultor	0,002016	1	0,892857183	0,98540146
8	agricultura	0,045797	9	8,035714179	0,64379562
9	agua	0,032783	7	6,25	0,702189781
10	aire	0,040176	7	6,25	0,684671533
11	al-aire-libre	0,008258	2	1,785714366	0,894890511
12	aldea	0,010824	2	1,785714366	0,861313869
13	andar	0,001761	1	0,892857183	0,991240876
14	animal	0,315958	46	41,07142985	0,159124088
15	árbol	0,439436	63	56,25	0,091970803
16	arboreto	0,002308	1	0,892857183	0,981021898
17	arroz	0,009915	2	1,785714366	0,867153285
18	asno	0,002308	1	0,892857183	0,979562044
19	ave	0,003965	1	0,892857183	0,954744526
20	bar	0,006812	1	0,892857183	0,913868613
21	bestia	0,002642	1	0,892857183	0,972262774
22	bici(cleta)	0,013506	4	3,571428731	0,837956204
23	bosque	0,022677	3	2,678571455	0,737226277
24	burro	0,016741	4	3,571428731	0,8
25	caballo	0,119684	23	20,53571492	0,420437956
26	cabra	0,018616	4	3,571428731	0,794160584
27	calle	0,004539	1	0,892857183	0,941605839
28	campesino	0,1341	24	21,42857164	0,35620438
29	campo	0,04219	6	5,35714291	0,674452555
30	casa	0,09598	14	12,5	0,440875912
31	cerdo	0,051808	12	10,71428582	0,605839416
32	cereales	0,029859	5	4,464285821	0,718248175
33	césped	0,013624	2	1,785714366	0,832116788
34	chalé	0,004539	1	0,892857183	0,950364964
35	chanchó	0,001343	1	0,892857183	0,995620438
36	chivo	0,002642	1	0,892857183	0,976642336
37	cine	0,008929	1	0,892857183	0,880291971

38	coche	0,00595	1	0,892857183	0,921167883
39	colina	0,007799	1	0,892857183	0,900729927
40	conejo	0,01925	5	4,464285821	0,77810219
41	cordero	0,003025	1	0,892857183	0,966423358
42	correr	0,001343	1	0,892857183	0,994160584
43	criadero	0,001173	1	0,892857183	0,998540146
44	cultivación	0,043185	8	7,142857462	0,665693431
45	cultivador	0,008258	2	1,785714366	0,889051095
46	cultivar	0,013993	3	2,678571455	0,82919708
47	cultivo	0,006812	1	0,892857183	0,918248175
48	encierro	0,003025	1	0,892857183	0,967883212
49	espacio	0,003463	1	0,892857183	0,962043796
50	estación	0,003025	1	0,892857183	0,964963504
51	fértil	0,003965	1	0,892857183	0,95620438
52	finca	0,001538	1	0,892857183	0,99270073
53	flor	0,213539	34	30,35714328	0,208759124
54	fruta	0,084111	14	12,5	0,505109489
55	fuelle	0,002642	1	0,892857183	0,97080292
56	gallina	0,072989	15	13,39285672	0,550364964
57	gallo	0,008258	2	1,785714366	0,891970803
58	ganado	0,006812	1	0,892857183	0,915328467
59	gancho	0,008929	1	0,892857183	0,886131387
60	gato	0,028939	7	6,25	0,728467153
61	gaucho	0,008929	1	0,892857183	0,881751825
62	gente	0,005197	1	0,892857183	0,934306569
63	girasol	0,010489	2	1,785714366	0,864233577
64	granja	0,015597	2	1,785714366	0,802919708
65	granjero	0,002016	1	0,892857183	0,988321168
66	grano	0,011764	2	1,785714366	0,848175182
67	habitación	0,00595	1	0,892857183	0,919708029
68	harina	0,014126	2	1,785714366	0,824817518
69	heno	0,004539	1	0,892857183	0,944525547
70	hierba	0,167694	23	20,53571492	0,283211679
71	hoja	0,003463	1	0,892857183	0,959124088
72	hormiga	0,007799	1	0,892857183	0,902189781
73	hortaliza	0,00595	1	0,892857183	0,922627737
74	huerto	0,015034	3	2,678571455	0,818978102
75	iglesia	0,009815	2	1,785714366	0,872992701
76	insecto	0,020053	5	4,464285821	0,76350365
77	irrigación	0,002016	1	0,892857183	0,986861314

78	jardín	0,019228	4	3,571428731	0,783941606
79	jugar	0,001173	1	0,892857183	0,997080292
80	lago	0,009837	2	1,785714366	0,870072993
81	leche	0,005197	1	0,892857183	0,932846715
82	libertad	0,007182	2	1,785714366	0,909489051
83	maíz	0,049381	9	8,035714179	0,618978102
84	manzana	0,007182	2	1,785714366	0,906569343
85	manzano	0,005197	1	0,892857183	0,940145985
86	mata	0,005197	1	0,892857183	0,937226277
87	miel	0,002642	1	0,892857183	0,973722628
88	molino	0,046934	8	7,142857462	0,630656934
89	naranja	0,004539	1	0,892857183	0,948905109
90	natura	0,012009	2	1,785714366	0,845255474
91	naturaleza	0,086197	15	13,39285672	0,484671533
92	niebla	0,011497	3	2,678571455	0,852554745
93	olivo	0,007799	1	0,892857183	0,897810219
94	oveja	0,044852	7	6,25	0,654014599
95	ovino	0,005197	1	0,892857183	0,935766423
96	paisaje	0,014611	2	1,785714366	0,82189781
97	paja	0,007799	1	0,892857183	0,903649635
98	pájaro	0,020209	4	3,571428731	0,75620438
99	parque	0,036433	5	4,464285821	0,691970803
100	pasear	0,020614	4	3,571428731	0,750364964
101	pasto	0,00595	1	0,892857183	0,928467153
102	pato	0,023539	3	2,678571455	0,732846715
103	pavo	0,002016	1	0,892857183	0,989781022
104	paz	0,009454	2	1,785714366	0,875912409
105	perro	0,087281	15	13,39285672	0,462773723
106	piedra	0,003025	1	0,892857183	0,969343066
107	planta	0,123476	21	18,75	0,386861314
108	pollo	0,019204	3	2,678571455	0,788321168
109	producto	0,003965	1	0,892857183	0,953284672
110	pueblo	0,00912	2	1,785714366	0,878832117
111	puerco	0,004539	1	0,892857183	0,945985401
112	puerto	0,007799	1	0,892857183	0,896350365
113	raíz	0,002642	1	0,892857183	0,975182482
114	red	0,008929	1	0,892857183	0,884671533
115	río	0,073646	16	14,28571492	0,528467153
116	rosa	0,011351	2	1,785714366	0,855474453
117	rural	0,062334	8	7,142857462	0,562043796

118	semen	0,004539	1	0,892857183	0,947445255
119	semilla	0,005224	2	1,785714366	0,931386861
120	silencio	0,015507	4	3,571428731	0,808759124
121	sol	0,021968	5	4,464285821	0,744525547
122	ternera	0,00595	1	0,892857183	0,924087591
123	terreno	0,002308	1	0,892857183	0,97810219
124	tienda	0,002016	1	0,892857183	0,982481752
125	tierra	0,168267	28	25	0,249635036
126	tomate	0,003965	1	0,892857183	0,951824818
127	toro	0,015184	4	3,571428731	0,81459854
128	trabajador	0,004539	1	0,892857183	0,943065693
129	trabajar	0,001025	1	0,892857183	1
130	trabajo	0,012376	3	2,678571455	0,842335766
131	tractor	0,054285	8	7,142857462	0,588321168
132	tranquilidad	0,01944	5	4,464285821	0,77080292
133	trigo	0,007799	1	0,892857183	0,899270073
134	vaca	0,149145	26	23,21428508	0,321167883
135	vacaciones	0,003463	1	0,892857183	0,960583942
136	vegetal	0,005197	1	0,892857183	0,938686131
137	verde	0,031061	6	5,35714291	0,710948905
138	verdura	0,061294	10	8,928571641	0,576642336
139	viento	0,011147	2	1,785714366	0,858394161
140	villa	0,006812	1	0,892857183	0,916788321
141	vino	0,006926	2	1,785714366	0,912408759

CI12, “medios de transporte”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	a-nado	0,003089	1	0,892857183	0,972772277
2	a-pie	0,06111	18	16,07142836	0,870049505
3	abono	0,008929	1	0,892857183	0,935643564
4	aéreo	0,005251	1	0,892857183	0,952970297
5	ambulancia	0,002588	1	0,892857183	0,98019802
6	andar	0,001522	1	0,892857183	0,995049505
7	(auto)bús	0,627778	93	83,03571343	0,228960396
8	auto(móvil)	0,008929	1	0,892857183	0,934405941
9	autocaravana	0,002168	1	0,892857183	0,986386139
10	AVE	0,003089	1	0,892857183	0,97029703
11	avión	0,45325	84	75	0,568069307
12	barco	0,181332	50	44,64285672	0,815594059
13	bici(cleta)	0,50198	96	85,71428657	0,464108911
14	billete	0,004759	3	2,678571455	0,96039604
15	caballo	0,033869	8	7,142857462	0,892326733
16	caminar	0,001817	1	0,892857183	0,988861386
17	camión	0,027908	10	8,928571641	0,90470297
18	camioneta	0,005855	2	1,785714366	0,949257426
19	carabela	0,003089	1	0,892857183	0,976485149
20	carne-de-conducir	0,008765	3	2,678571455	0,939356436
21	carro	0,047691	10	8,928571641	0,882425743
22	ciclomotor	0,0044	1	0,892857183	0,961633663
23	coche	0,687179	92	82,1428597	0,113861386
24	coger	0,003339	2	1,785714366	0,965346535
25	colectivo	0,002588	1	0,892857183	0,981435644
26	conducir	0,008267	4	3,571428731	0,944306931
27	correr	0,003092	2	1,785714366	0,967821782
28	crucero	0,005676	2	1,785714366	0,951732673
29	esmog	0,002168	1	0,892857183	0,983910891
30	esperar	0,002588	1	0,892857183	0,977722772
31	funicular	0,002168	1	0,892857183	0,985148515
32	gasolina	0,001275	1	0,892857183	0,996287129
33	globo-aerostático	0,002168	1	0,892857183	0,987623762
34	góndola	0,003089	1	0,892857183	0,974009901
35	guagua	0,003089	1	0,892857183	0,969059406
36	hacer-dedo	0,000895	1	0,892857183	1

37	helicóptero	0,022956	8	7,142857462	0,92450495
38	horario	0,002168	1	0,892857183	0,982673267
39	ir	0,001522	1	0,892857183	0,991336634
40	máquina	0,0044	1	0,892857183	0,962871287
41	metro	0,19024	47	41,96428657	0,685643564
42	moto(cicleta)	0,188392	55	49,10714328	0,753712871
43	nave	0,025833	8	7,142857462	0,91460396
44	ómnibus	0,001275	1	0,892857183	0,998762376
45	parada	0,001522	1	0,892857183	0,992574257
46	patinetes	0,018114	7	6,25	0,933168317
47	público	0,001275	1	0,892857183	0,997524752
48	subir	0,003089	1	0,892857183	0,971534653
49	submarino	0,001817	1	0,892857183	0,99009901
50	tándem	0,004905	2	1,785714366	0,956683168
51	taxi	0,202902	48	42,85714328	0,627475248
52	tomar	0,006177	2	1,785714366	0,946782178
53	transporte-público	0,001522	1	0,892857183	0,993811881
54	tranvía	0,123226	26	23,21428508	0,847772277
55	tren	0,565213	94	83,92857313	0,34529703
56	vehículo	0,005251	1	0,892857183	0,954207921
57	velero	0,003089	1	0,892857183	0,975247525
58	viajar	0,002588	1	0,892857183	0,978960396

CI13, “trabajos del campo y del jardín”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aceite	0,004478	1	0,892857	0,940476
2	agricultor	0,047603	7	6,25	0,563492
3	agua	0,005636	1	0,892857	0,900794
4	arado	0,004478	1	0,892857	0,944444
5	arar	0,017885	3	2,678571	0,765873
6	árbol	0,011784	3	2,678571	0,805556
7	burro	0,002827	1	0,892857	0,968254
8	caballo	0,003558	1	0,892857	0,952381
9	campesino	0,25699	32	28,57143	0,293651
10	campo	0,002827	1	0,892857	0,972222
11	cerdo	0,000895	1	0,892857	1
12	coger	0,02468	4	3,571429	0,694444
13	conejo	0,001127	1	0,892857	0,996032
14	cortar	0,146328	25	22,32143	0,47619
15	cosecha	0,007094	1	0,892857	0,888889
16	cosechar	0,002246	1	0,892857	0,984127
17	crecer	0,011595	3	2,678571	0,81746
18	criador	0,007094	1	0,892857	0,873016
19	criar	0,007094	1	0,892857	0,884921
20	cultivación	0,008929	1	0,892857	0,84127
21	cultivador	0,019581	3	2,678571	0,753968
22	cultivar	0,048955	7	6,25	0,535714
23	dar-agua	0,028753	4	3,571429	0,638889
24	enterrar	0,005636	1	0,892857	0,928571
25	escavar	0,021659	3	2,678571	0,722222
26	extirpar	0,007094	1	0,892857	0,876984
27	flor	0,04982	8	7,142857	0,507937
28	florero	0,022845	4	3,571429	0,710317
29	florista	0,029939	5	4,464286	0,623016
30	fruta	0,027617	5	4,464286	0,678571
31	gato	0,008929	1	0,892857	0,837302
32	gaucho	0,005636	1	0,892857	0,912698
33	granja	0,003558	1	0,892857	0,956349
34	granjero	0,035847	5	4,464286	0,603175
35	hierba	0,020058	5	4,464286	0,742063
36	irrigación	0,005636	1	0,892857	0,916667

37	irrigar	0,008929	1	0,892857	0,849206
38	jardín	0,016023	2	1,785714	0,77381
39	jardinero	0,3355	42	37,5	0,166667
40	leche	0,005636	1	0,892857	0,904762
41	madurar	0,005654	2	1,785714	0,896825
42	mantequilla	0,007094	1	0,892857	0,880952
43	mariposa	0,002246	1	0,892857	0,980159
44	molinero	0,007094	1	0,892857	0,861111
45	obrero	0,007094	1	0,892857	0,869048
46	oveja	0,004478	1	0,892857	0,93254
47	pájaro	0,001785	1	0,892857	0,988095
48	pastor	0,028238	5	4,464286	0,65873
49	perro	0,007094	1	0,892857	0,857143
50	pez	0,001418	1	0,892857	0,992063
51	planta	0,012515	3	2,678571	0,793651
52	plantar	0,162166	21	18,75	0,376984
53	rama	0,005636	1	0,892857	0,920635
54	recoger	0,012731	2	1,785714	0,781746
55	regar	0,036224	5	4,464286	0,583333
56	sembrar	0,008929	1	0,892857	0,845238
57	tierra	0,009195	2	1,785714	0,833333
58	tijeras	0,007094	1	0,892857	0,853175
59	trabajador	0,002827	1	0,892857	0,964286
60	vaca	0,005636	1	0,892857	0,90873
61	vaquero	0,004478	1	0,892857	0,936508
62	vaso	0,002827	1	0,892857	0,960317
63	vender	0,003558	1	0,892857	0,948413
64	verdulero	0,007094	1	0,892857	0,865079
65	verdura	0,009921	2	1,785714	0,825397
66	vino	0,002246	1	0,892857	0,97619
67	viticultor	0,005636	1	0,892857	0,924603

CI14, “los animales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abeja	0,009571	2	1,785714366	0,944444444
2	águila	0,017086	6	5,35714291	0,912037037
3	alpaca	0,001927	1	0,892857183	0,994444444
4	araña	0,010521	4	3,571428731	0,940740741
5	ardilla	0,005164	1	0,892857183	0,975925926
6	ave	0,077444	14	12,5	0,681481481
7	avestruz	0,011175	3	2,678571455	0,933333333
8	bagre	0,000999	1	0,892857183	1
9	ballena	0,007535	3	2,678571455	0,953703704
10	bicho	0,005164	1	0,892857183	0,974074074
11	buey	0,021006	4	3,571428731	0,900925926
12	buitre	0,005164	1	0,892857183	0,976851852
13	burro	0,08815	21	18,75	0,633333333
14	caballo	0,490249	86	76,78571343	0,258333333
15	cabra	0,04123	14	12,5	0,813888889
16	cabrón	0,00748	2	1,785714366	0,955555556
17	cachorro	0,001548	1	0,892857183	0,996296296
18	camarón	0,003332	1	0,892857183	0,986111111
19	canario	0,010925	2	1,785714366	0,937037037
20	cangrejo	0,003718	1	0,892857183	0,984259259
21	canguro	0,002399	1	0,892857183	0,992592593
22	caracol	0,024341	6	5,35714291	0,897222222
23	cebra	0,075974	18	16,07142836	0,712962963
24	cerdo	0,187271	35	31,25	0,468518519
25	chanchó	0,004148	1	0,892857183	0,97962963
26	chinche	0,002399	1	0,892857183	0,991666667
27	chivo	0,00215	1	0,892857183	0,993518519
28	ciervo	0,009571	2	1,785714366	0,946296296
29	cochino	0,006428	1	0,892857183	0,964814815
30	cocodrilo	0,027103	7	6,25	0,87962963
31	conejo	0,13028	26	23,21428508	0,548148148
32	cordero	0,005385	2	1,785714366	0,972222222
33	cucaracha	0,026991	7	6,25	0,886111111
34	culebra	0,007024	4	3,571428731	0,962962963
35	delfín	0,032163	7	6,25	0,839814815
36	elefante	0,098352	24	21,42857164	0,59537037

37	escorpión	0,001727	1	0,892857183	0,99537037
38	estrella-marina	0,005761	1	0,892857183	0,968518519
39	flamenco	0,00748	2	1,785714366	0,957407407
40	foca	0,001387	1	0,892857183	0,999074074
41	gallina	0,144076	29	25,89285672	0,49537037
42	gallo	0,055714	14	12,5	0,769444444
43	gamba	0,003718	1	0,892857183	0,982407407
44	gato	0,750967	96	85,71428657	0,178703704
45	gaviota	0,01631	4	3,571428731	0,915740741
46	gorila	0,003332	1	0,892857183	0,987037037
47	guepardo	0,004148	1	0,892857183	0,981481481
48	hámster	0,028938	6	5,35714291	0,873148148
49	hipopótamo	0,002841	2	1,785714366	0,989814815
50	hormiga	0,057993	15	13,39285672	0,756481481
51	insecto	0,005761	1	0,892857183	0,967592593
52	jabalí	0,011056	2	1,785714366	0,935185185
53	jaguar	0,005164	1	0,892857183	0,975
54	jirafa	0,07895	21	18,75	0,668518519
55	lagarto	0,01592	3	2,678571455	0,918518519
56	león	0,205385	47	41,96428657	0,436111111
57	leopardo	0,003332	1	0,892857183	0,985185185
58	llama	0,007314	2	1,785714366	0,959259259
59	lobo	0,04473	10	8,928571641	0,789814815
60	mariposa	0,093563	20	17,85714328	0,613888889
61	mariquita	0,015357	2	1,785714366	0,928703704
62	mono	0,034346	8	7,142857462	0,821296296
63	mosca	0,006428	1	0,892857183	0,963888889
64	mosquito	0,029155	9	8,035714179	0,867592593
65	mulo	0,004148	1	0,892857183	0,978703704
66	murciélago	0,001548	1	0,892857183	0,997222222
67	oca	0,004628	1	0,892857183	0,977777778
68	orangután	0,002676	1	0,892857183	0,990740741
69	oso	0,047829	12	10,71428582	0,780555556
70	oveja	0,04124	12	10,71428582	0,800925926
71	pájaro	0,209367	43	38,39285672	0,392592593
72	paloma	0,015871	5	4,464285821	0,923148148
73	panda	0,010158	2	1,785714366	0,942592593
74	pantera	0,00796	2	1,785714366	0,949074074

75	papagayo	0,031386	7	6,25	0,846296296
76	pato	0,033493	7	6,25	0,833333333
77	pavo	0,005761	1	0,892857183	0,969444444
78	pécora	0,002986	1	0,892857183	0,987962963
79	perro	0,816215	97	86,6071403	0,089814815
80	pez	0,229914	46	41,07142985	0,352777778
81	pingüino	0,019765	6	5,35714291	0,906481481
82	pitón	0,008929	1	0,892857183	0,947222222
83	pollo	0,074909	14	12,5	0,725925926
84	puerco	0,029306	7	6,25	0,859259259
85	pulpo	0,006318	2	1,785714366	0,966666667
86	rana	0,003718	1	0,892857183	0,983333333
87	rata	0,001548	1	0,892857183	0,998148148
88	ratón	0,076414	16	14,28571492	0,696296296
89	salmón	0,01544	4	3,571428731	0,926851852
90	serpiente	0,058688	18	16,07142836	0,742592593
91	sierpe	0,004148	1	0,892857183	0,980555556
92	tarántula	0,005761	1	0,892857183	0,97037037
93	tiburón	0,029542	7	6,25	0,852777778
94	tigre	0,131441	31	27,67857015	0,524074074
95	tigrillo	0,005164	1	0,892857183	0,973148148
96	topo	0,024902	6	5,35714291	0,891666667
97	toro	0,120198	27	24,10714328	0,573148148
98	tortuga	0,08156	17	15,17857164	0,649074074
99	tucán	0,007815	2	1,785714366	0,950925926
100	vaca	0,298745	56	50	0,310185185
101	yegua	0,012189	2	1,785714366	0,930555556
102	zorro	0,033572	6	5,35714291	0,826851852

CI15, “juegos y distracciones”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	actuar	0,006086	1	0,892857183	0,928571429
2	ajedrez	0,049854	7	6,25	0,696428571
3	amigo	0,013538	3	2,678571455	0,886479592
4	andar	0,006915	1	0,892857183	0,915816327
5	arena	0,006915	1	0,892857183	0,922193878
6	atletismo	0,004713	1	0,892857183	0,964285714
7	atracción	0,005356	1	0,892857183	0,951530612
8	bailar	0,177022	33	29,46428657	0,253826531
9	balón	0,060566	10	8,928571641	0,663265306
10	baloncesto	0,317442	50	44,64285672	0,130102041
11	balonmano	0,006086	1	0,892857183	0,93494898
12	balonvolea	0,019863	3	2,678571455	0,825255102
13	bar	0,018203	3	2,678571455	0,834183673
14	beber	0,005356	1	0,892857183	0,946428571
15	beisbol	0,005356	1	0,892857183	0,952806122
16	bici(cleta)	0,011824	4	3,571428731	0,894132653
17	boxeo	0,004148	1	0,892857183	0,975765306
18	broma	0,004713	1	0,892857183	0,963010204
19	caballo	0,001696	1	0,892857183	0,996173469
20	caminar	0,017156	3	2,678571455	0,848214286
21	canción	0,004148	1	0,892857183	0,969387755
22	cantar	0,049265	9	8,035714179	0,707908163
23	cartas	0,167424	25	22,32142836	0,285714286
24	chatear	0,005577	2	1,785714366	0,945153061
25	cine	0,124059	19	16,96428508	0,414540816
26	club	0,003212	1	0,892857183	0,983418367
27	cocinar	0,014999	5	4,464285821	0,868622449
28	colina	0,001313	1	0,892857183	0,99872449
29	colorar	0,00219	1	0,892857183	0,991071429
30	comer	0,006981	2	1,785714366	0,914540816
31	cómic	0,006915	1	0,892857183	0,918367347
32	competición	0,00365	1	0,892857183	0,977040816
33	computador	0,008929	1	0,892857183	0,904336735
34	construcciones	0,006086	1	0,892857183	0,932397959
35	correr	0,099855	20	17,85714328	0,5
36	dado	0,021499	3	2,678571455	0,80994898

37	danza	0,024772	3	2,678571455	0,801020408
38	deporte	0,063196	12	10,71428582	0,650510204
39	descansar	0,006915	1	0,892857183	0,919642857
40	dibujar	0,046347	9	8,035714179	0,719387755
41	dibujos-animados	0,01677	3	2,678571455	0,852040816
42	disco(teca)	0,058453	9	8,035714179	0,674744898
43	diversión	0,001927	1	0,892857183	0,993622449
44	emborracharse	0,006086	1	0,892857183	0,933673469
45	energía	0,004713	1	0,892857183	0,961734694
46	entrenarse	0,012579	2	1,785714366	0,889030612
47	escalar	0,001927	1	0,892857183	0,994897959
48	escondite	0,030285	5	4,464285821	0,790816327
49	escribir	0,002827	1	0,892857183	0,987244898
50	escuchar-música	0,120973	24	21,42857164	0,445153061
51	esquiar	0,006086	1	0,892857183	0,93622449
52	excursión	0,003212	1	0,892857183	0,980867347
53	fiesta	0,035948	5	4,464285821	0,761479592
54	fotografía	0,005017	2	1,785714366	0,955357143
55	fotografiar	0,001492	1	0,892857183	0,99744898
56	fútbol	0,358128	52	46,42857015	0,066326531
57	gimnasio	0,01452	4	3,571428731	0,878826531
58	golf	0,013626	3	2,678571455	0,882653061
59	grupo	0,009298	2	1,785714366	0,899234694
60	guitarra	0,034245	6	5,35714291	0,776785714
61	hablar	0,004148	1	0,892857183	0,974489796
62	instrumento-musical	0,020806	5	4,464285821	0,816326531
63	internet	0,004148	1	0,892857183	0,968112245
64	ir-de-bares	0,005356	1	0,892857183	0,947704082
65	ir-de-compras	0,014788	4	3,571428731	0,87372449
66	ir-de-tapas	0,006086	1	0,892857183	0,929846939
67	ir-de-vacaciones	0,006086	1	0,892857183	0,926020408
68	juego	0,007798	2	1,785714366	0,911989796
69	juego-de-mesa	0,033711	6	5,35714291	0,784438776
70	jugar	0,092449	17	15,17857164	0,547193878
71	juguete	0,00365	1	0,892857183	0,979591837
72	karaoke	0,006915	1	0,892857183	0,920918367
73	leer	0,205784	32	28,57142985	0,170918367

74	libro	0,1811	32	28,57142985	0,211734694
75	lucha	0,004148	1	0,892857183	0,973214286
76	maquillaje	0,004908	2	1,785714366	0,957908163
77	mar	0,004319	2	1,785714366	0,966836735
78	<i>Monopoly</i>	0,006086	1	0,892857183	0,93877551
79	montaña	0,001156	1	0,892857183	1
80	movida	0,00219	1	0,892857183	0,989795918
81	móvil	0,042357	7	6,25	0,738520408
82	muñeca	0,045574	8	7,142857462	0,729591837
83	museo	0,009124	3	2,678571455	0,903061224
84	nadar	0,095567	20	17,85714328	0,525510204
85	natación	0,103617	23	20,53571492	0,474489796
86	niño	0,006086	1	0,892857183	0,9375
87	ocio	0,007858	1	0,892857183	0,909438776
88	ordenador	0,130146	21	18,75	0,366071429
89	parque	0,029474	5	4,464285821	0,797193878
90	parque-de- diversiones	0,023899	4	3,571428731	0,806122449
91	parqués	0,004148	1	0,892857183	0,970663265
92	partido	0,00664	2	1,785714366	0,924744898
93	pasárselo-bien	0,007858	1	0,892857183	0,905612245
94	pasear	0,09174	16	14,28571492	0,567602041
95	patinar	0,002827	1	0,892857183	0,988520408
96	patinetes	0,006915	1	0,892857183	0,917091837
97	pelí(cula)	0,066639	12	10,71428582	0,621173469
98	pelota	0,137515	21	18,75	0,3125
99	peluche	0,016153	3	2,678571455	0,855867347
100	perro	0,005844	2	1,785714366	0,942602041
101	piano	0,017225	4	3,571428731	0,844387755
102	pillapilla	0,006086	1	0,892857183	0,931122449
103	piñata	0,004148	1	0,892857183	0,971938776
104	pintar	0,049994	10	8,928571641	0,6875
105	piscina	0,004713	1	0,892857183	0,959183673
106	<i>Play-Station</i>	0,040487	5	4,464285821	0,755102041
107	playa	0,017329	4	3,571428731	0,839285714
108	quedar-con- amigos	0,011063	2	1,785714366	0,896683673
109	radio	0,015844	2	1,785714366	0,858418367
110	redes-sociales	0,001927	1	0,892857183	0,992346939

111	reír	0,015656	3	2,678571455	0,862244898
112	restaurante	0,004713	1	0,892857183	0,960459184
113	rompecabezas	0,006086	1	0,892857183	0,94005102
114	salir	0,041108	8	7,142857462	0,74872449
115	salir-con-amigos	0,035292	6	5,35714291	0,769132653
116	saltar	0,005356	1	0,892857183	0,950255102
117	sexo	0,007858	1	0,892857183	0,908163265
118	soldadito-de-plomo	0,003212	1	0,892857183	0,984693878
119	sonreír	0,00365	1	0,892857183	0,978316327
120	tapear	0,006086	1	0,892857183	0,927295918
121	teatro	0,020196	4	3,571428731	0,821428571
122	tejo	0,005356	1	0,892857183	0,948979592
123	tele(visión)	0,135749	21	18,75	0,339285714
124	telenovela	0,002827	1	0,892857183	0,985969388
125	tenis	0,077068	17	15,17857164	0,589285714
126	tobogán	0,007858	1	0,892857183	0,906887755
127	tocar	0,065828	11	9,821428359	0,635204082
128	viajar	0,019811	4	3,571428731	0,830357143
129	videojuegos	0,124964	19	16,96428508	0,390306122
130	voleibol	0,067557	13	11,60714254	0,605867347
131	yoga	0,003212	1	0,892857183	0,982142857

CI16, “profesiones y oficios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abogado	0,175484	29	25,89285672	0,220357942
2	actor	0,11486	28	25	0,340044743
3	actriz	0,025149	8	7,142857462	0,771812081
4	aeronauta	0,003769	1	0,892857183	0,957494407
5	agricultor	0,008929	1	0,892857183	0,883668904
6	albañil	0,03469	7	6,25	0,689038031
7	alcalde	0,012915	3	2,678571455	0,864653244
8	alguacil	0,006698	1	0,892857183	0,910514541
9	ama-de-casa	0,042341	10	8,928571641	0,636465324
10	andrólogo	0,003264	1	0,892857183	0,975391499
11	animador	0,003264	1	0,892857183	0,973154362
12	arquitecto	0,014348	3	2,678571455	0,85458613
13	artesano	0,008628	2	1,785714366	0,89261745
14	asistente	0,020136	4	3,571428731	0,807606264
15	atleta	0,008717	3	2,678571455	0,887024609
16	autor	0,017768	4	3,571428731	0,823266219
17	ayudante	0,001378	1	0,892857183	0,996644295
18	azafata	0,004642	2	1,785714366	0,937360179
19	bailar	0,006698	1	0,892857183	0,908277405
20	bailarín	0,031666	10	8,928571641	0,717002237
21	banco	0,027348	4	3,571428731	0,750559284
22	barista	0,021708	5	4,464285821	0,798657718
23	bibliotecario	0,004351	1	0,892857183	0,947427293
24	biólogo	0,004298	2	1,785714366	0,951901566
25	bombero	0,074168	12	10,71428582	0,494407159
26	botón	0,004351	1	0,892857183	0,948545861
27	cajero	0,009065	2	1,785714366	0,881431767
28	camarero	0,14971	31	27,67857015	0,288590604
29	camionero	0,005801	1	0,892857183	0,916107383
30	campesino	0,018558	3	2,678571455	0,815436242
31	canguro	0,062967	13	11,60714254	0,559284116
32	cantante	0,098417	26	23,21428508	0,429530201
33	cardiólogo	0,005024	1	0,892857183	0,935123043
34	carnicero	0,028515	6	5,35714291	0,746085011
35	carpintero	0,005024	1	0,892857183	0,934004474
36	cartonero	0,002449	1	0,892857183	0,985458613
37	charcutero	0,006698	1	0,892857183	0,909395973

38	chef	0,025022	7	6,25	0,779642058
39	chofer	0,010997	2	1,785714366	0,874720358
40	cirujano	0,012697	2	1,785714366	0,86689038
41	cocinar	0,003264	1	0,892857183	0,963087248
42	cocinero	0,197094	37	33,03571343	0,187919463
43	comerciante	0,005801	1	0,892857183	0,925055928
44	cómico	0,006188	2	1,785714366	0,914988814
45	cuidador	0,051317	9	8,035714179	0,592841163
46	cuidador-de-perro	0,003264	1	0,892857183	0,966442953
47	cura	0,005801	1	0,892857183	0,917225951
48	dentista	0,029778	6	5,35714291	0,723713647
49	dependiente	0,024802	5	4,464285821	0,785234899
50	dibujador	0,001193	1	0,892857183	1
51	director	0,07222	15	13,39285672	0,527964206
52	dirigente	0,003264	1	0,892857183	0,968680089
53	doctor	0,162642	30	26,78571343	0,253914989
54	dueño	0,004351	1	0,892857183	0,945190157
55	economista	0,003264	1	0,892857183	0,967561521
56	editor	0,010152	2	1,785714366	0,876957494
57	educador	0,005024	1	0,892857183	0,932885906
58	empleado	0,101295	19	16,96428508	0,380313199
59	emprendedor	0,003769	1	0,892857183	0,958612975
60	empresario	0,008288	2	1,785714366	0,894854586
61	enfermero	0,075385	13	11,60714254	0,48098434
62	enseñante	0,015626	2	1,785714366	0,840044743
63	enseñar	0,007638	2	1,785714366	0,901565996
64	entrenador	0,084874	16	14,28571492	0,447427293
65	escritor	0,050478	14	12,5	0,608501119
66	escuela	0,00536	2	1,785714366	0,929530201
67	escultor	0,003769	1	0,892857183	0,960850112
68	esteticista	0,003769	1	0,892857183	0,956375839
69	estilista	0,005801	1	0,892857183	0,921700224
70	estudiante	0,1071	17	15,17857164	0,359060403
71	estudio	0,008929	1	0,892857183	0,882550336
72	fábrica	0,005024	1	0,892857183	0,93064877
73	farmacéutico	0,029035	7	6,25	0,731543624
74	fisioterapeuta	0,001837	1	0,892857183	0,991051454
75	florero	0,003264	1	0,892857183	0,964205817
76	florista	0,005713	2	1,785714366	0,927293065

77	fontanero	0,130173	18	16,07142836	0,308724832
78	fotógrafo	0,007775	3	2,678571455	0,898210291
79	frutero	0,003264	1	0,892857183	0,974272931
80	futbolista	0,033659	7	6,25	0,696868009
81	gestor	0,003769	1	0,892857183	0,955257271
82	ginecólogo	0,004351	1	0,892857183	0,94966443
83	gobernador	0,002827	1	0,892857183	0,977628635
84	guardia	0,001591	1	0,892857183	0,994407159
85	guía-de-turismo	0,044641	9	8,035714179	0,618568233
86	guionista	0,001591	1	0,892857183	0,992170022
87	heladero	0,002827	1	0,892857183	0,982102908
88	hidráulico	0,002449	1	0,892857183	0,984340045
89	hospital	0,001193	1	0,892857183	0,998881432
90	informático	0,015371	3	2,678571455	0,843400447
91	ingeniero	0,017722	3	2,678571455	0,826621924
92	instructor	0,005801	1	0,892857183	0,920581655
93	intérprete	0,014393	4	3,571428731	0,84787472
94	jardinero	0,035303	6	5,35714291	0,681208054
95	jefe	0,035977	8	7,142857462	0,663310962
96	juez	0,026093	4	3,571428731	0,755033557
97	jugador	0,025719	7	6,25	0,762863535
98	lector	0,006698	1	0,892857183	0,912751678
99	limpiar	0,003264	1	0,892857183	0,965324385
100	logopeda	0,004351	1	0,892857183	0,941834452
101	maestro	0,043053	6	5,35714291	0,625279642
102	maquillador	0,003264	1	0,892857183	0,970917226
103	mecánico	0,013144	3	2,678571455	0,861297539
104	médico	0,213237	36	32,14285672	0,146532438
105	militar	0,002121	1	0,892857183	0,988814318
106	modelista	0,005801	1	0,892857183	0,918344519
107	modelo	0,019113	4	3,571428731	0,812080537
108	modista	0,021464	4	3,571428731	0,803131991
109	monja	0,005024	1	0,892857183	0,931767338
110	mucamo	0,005801	1	0,892857183	0,919463087
111	músico	0,006983	3	2,678571455	0,907158837
112	nadador	0,003769	1	0,892857183	0,959731544
113	negocios	0,002827	1	0,892857183	0,976510067
114	neurólogo	0,003769	1	0,892857183	0,96196868
115	notario	0,007733	1	0,892857183	0,899328859
116	obrero	0,057789	9	8,035714179	0,56935123

117	oficina	0,033568	8	7,142857462	0,705816555
118	ordenar	0,002827	1	0,892857183	0,978747204
119	osteópata	0,001591	1	0,892857183	0,993288591
120	panadero	0,014348	3	2,678571455	0,851230425
121	pastor	0,002449	1	0,892857183	0,983221477
122	payaso	0,004351	1	0,892857183	0,938478747
123	peluquero	0,072462	15	13,39285672	0,511185682
124	periodista	0,03585	10	8,928571641	0,674496644
125	pescador	0,001378	1	0,892857183	0,997762864
126	piloto	0,041793	8	7,142857462	0,64541387
127	pintor	0,051494	12	10,71428582	0,582774049
128	plomero	0,004351	1	0,892857183	0,94295302
129	policía	0,08101	17	15,17857164	0,466442953
130	político	0,022188	7	6,25	0,793064877
131	portero	0,002827	1	0,892857183	0,98098434
132	presentador	0,008637	3	2,678571455	0,890380313
133	presidente	0,003769	1	0,892857183	0,953020134
134	preste	0,005801	1	0,892857183	0,92393736
135	profesor	0,714364	95	84,82142687	0,106263982
136	programador	0,004351	1	0,892857183	0,944071588
137	psicólogo	0,002449	1	0,892857183	0,986577181
138	químico	0,011722	2	1,785714366	0,872483221
139	repcionista	0,006698	1	0,892857183	0,91163311
140	redactor	0,007538	2	1,785714366	0,903803132
141	representante	0,003264	1	0,892857183	0,972035794
142	restaurante	0,002121	1	0,892857183	0,987695749
143	rey	0,004351	1	0,892857183	0,946308725
144	sastre	0,017217	3	2,678571455	0,829977629
145	secretario	0,098878	18	16,07142836	0,400447427
146	seguro	0,003769	1	0,892857183	0,954138702
147	señora-de-la-limpieza	0,01655	4	3,571428731	0,834451902
148	sirviente	0,001378	1	0,892857183	0,995525727
149	socorrista	0,00957	2	1,785714366	0,879194631
150	tabaquero	0,004351	1	0,892857183	0,939597315
151	tallador	0,002121	1	0,892857183	0,989932886
152	taxista	0,004351	1	0,892857183	0,940715884
153	técnico	0,012601	3	2,678571455	0,870246085
154	tenista	0,003264	1	0,892857183	0,969798658
155	tienda	0,013144	3	2,678571455	0,857941834

156	trabajador	0,015953	3	2,678571455	0,837807606
157	trabajo	0,018304	3	2,678571455	0,818791946
158	traductor	0,067372	15	13,39285672	0,544742729
159	traficante	0,005801	1	0,892857183	0,922818792
160	vendedor	0,037778	8	7,142857462	0,654362416
161	veterinario	0,02871	7	6,25	0,739373602
162	zapatero	0,002827	1	0,892857183	0,979865772

ANEXO V – LISTADO GENERAL DE LÉXICO DISPONIBLE

(CORPUS B2_ b)

CI01, “las partes del cuerpo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	mano	0,70837861	48	95,99999785	0,061146497
2	ojos	0,6513288	47	93,99999976	0,121019108
3	pierna	0,60147297	43	86,00000143	0,175796178
4	cabeza	0,56961423	36	72,00000286	0,221656051
5	pie	0,55654699	47	93,99999976	0,281528662
6	boca	0,52944398	39	77,99999714	0,331210191
7	dedo	0,46233949	41	81,99999928	0,38343949
8	brazo	0,45347556	38	75,99999905	0,431847134
9	nariz	0,43172881	36	72,00000286	0,477707006
10	pelo	0,41362104	35	69,99999881	0,522292994
11	cara	0,36953753	25	50	0,554140127
12	orejas	0,32147488	30	60,00000238	0,592356688
13	espalda	0,29661363	32	63,99999857	0,633121019
14	cuello	0,25987297	27	54,00000215	0,667515924
15	rodilla	0,24885885	25	50	0,699363057
16	uñas	0,21045408	23	46,00000083	0,72866242
17	lengua	0,18896568	20	40,0000006	0,754140127
18	dientes	0,1774254	21	41,99999869	0,78089172
19	hombro	0,12907742	12	23,99999946	0,796178344
20	labios	0,12809826	13	25,99999905	0,812738854
21	barriga	0,11000639	13	25,99999905	0,829299363
22	pecho	0,09088355	12	23,99999946	0,844585987
23	pestañas	0,07515361	8	15,99999964	0,85477707
24	cejas	0,0749658	7	14,00000006	0,863694268
25	cintura	0,07182431	8	15,99999964	0,87388535
26	codos	0,0695781	10	20,0000003	0,886624204
27	muñeca	0,06933831	8	15,99999964	0,896815287
28	piel	0,06525035	6	11,99999973	0,904458599
29	cabellos	0,06121696	5	10,00000015	0,910828025
30	corazón	0,05311658	4	7,999999821	0,915923567
31	mejilla	0,05300115	7	14,00000006	0,924840764
32	frente	0,04371168	5	10,00000015	0,931210191
33	trasero	0,0367991	5	10,00000015	0,937579618

34	cadera	0,03436467	5	10,00000015	0,943949045
35	rostro	0,03405455	2	3,999999911	0,946496815
36	estómago	0,03304558	4	7,999999821	0,951592357
37	puño	0,02384488	2	3,999999911	0,954140127
38	tripa	0,02250144	2	3,999999911	0,956687898
39	cuerpo	0,02101363	2	3,999999911	0,959235669
40	ombligo	0,01977235	2	3,999999911	0,961783439
41	garganta	0,01934235	3	5,999999866	0,965605096
42	nalgas	0,01567094	2	3,999999911	0,968152866
43	sobrecejos	0,01563174	1	1,999999955	0,969426752
44	tobillo	0,01335532	2	3,999999911	0,971974522
45	bíceps	0,0103666	1	1,999999955	0,973248408
46	pupila	0,0103666	1	1,999999955	0,974522293
47	olfato	0,0095491	1	1,999999955	0,975796178
48	músculo	0,0095491	1	1,999999955	0,977070064
49	riñón	0,00879606	1	1,999999955	0,978343949
50	pulmones	0,0081024	1	1,999999955	0,979617834
51	glúteo	0,00746345	1	1,999999955	0,98089172
52	muelas	0,00687489	1	1,999999955	0,982165605
53	culo	0,00665369	2	3,999999911	0,984713376
54	colita	0,00633274	1	1,999999955	0,985987261
55	panza	0,00633274	1	1,999999955	0,987261146
56	talón	0,00419972	1	1,999999955	0,988535032
57	mentón	0,00386853	1	1,999999955	0,989808917
58	barba	0,00328245	1	1,999999955	0,991082803
59	rastro	0,00328245	1	1,999999955	0,992356688
60	paladar	0,00278516	1	1,999999955	0,993630573
61	sangre	0,00278516	1	1,999999955	0,994904459
62	ver	0,00256552	1	1,999999955	0,996178344
63	médico	0,0023632	1	1,999999955	0,997452229
64	cerebro	0,00217684	1	1,999999955	0,998726115
65	pulgar	0,00200518	1	1,999999955	1

CI02, "la ropa"

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	camiseta	0,609961331	38	75,99999905	0,063973064
2	pantalones	0,55659759	43	86,00000143	0,136363636
3	falda	0,521908998	37	74,00000095	0,198653199
4	zapatos	0,514442086	40	80,00000119	0,265993266
5	camisa	0,441626191	29	57,99999833	0,314814815
6	sudadera	0,437353641	30	60,00000238	0,365319865
7	vaqueros	0,375662982	26	51,99999809	0,409090909
8	chaqueta	0,24462299	24	47,99999893	0,449494949
9	sombrero	0,214348674	22	43,99999976	0,486531987
10	bufanda	0,212874815	22	43,99999976	0,523569024
11	calcetines	0,203128695	21	41,99999869	0,558922559
12	vestido	0,188340113	16	31,99999928	0,585858586
13	jersey	0,177477971	15	30,00000119	0,611111111
14	bolso	0,16805914	16	31,99999928	0,638047138
15	cinturón	0,160558224	12	23,99999946	0,658249158
16	traje	0,149300545	15	30,00000119	0,683501684
17	minifalda	0,131340444	10	20,00000003	0,7003367
18	zapatillas	0,121659353	12	23,99999946	0,720538721
19	abrigo	0,117233962	13	25,99999905	0,742424242
20	tacones	0,116492882	14	28,00000012	0,765993266
21	blusa	0,115164272	10	20,00000003	0,782828283
22	bañador	0,096230097	11	21,99999988	0,801346801
23	pijama	0,091781095	8	15,99999964	0,814814815
24	gorro	0,086104937	10	20,00000003	0,831649832
25	botas	0,073819496	8	15,99999964	0,845117845
26	sujetador	0,062678814	7	14,00000006	0,856902357
27	mochila	0,056511249	5	10,00000015	0,865319865
28	guantes	0,05616384	8	15,99999964	0,878787879
29	gafas	0,054199327	7	14,00000006	0,890572391
30	ropa-interior	0,048420023	6	11,99999973	0,900673401
31	ropa	0,038156971	3	5,999999866	0,905723906
32	chándal	0,036094394	3	5,999999866	0,910774411
33	bragas	0,034903485	4	7,999999821	0,917508418
34	suéter	0,032630373	2	3,999999911	0,920875421
35	anillo	0,031649657	5	10,00000015	0,929292929
36	vestirse	0,022785522	2	3,999999911	0,932659933
37	rebeca	0,021866739	3	5,999999866	0,937710438

38	reloj	0,021400414	3	5,999999866	0,942760943
39	chaleco	0,02	1	1,999999955	0,944444444
40	calzoncillos	0,019981118	2	3,999999911	0,947811448
41	chanclas	0,018491071	3	5,999999866	0,952861953
42	capucha	0,016226429	1	1,999999955	0,954545455
43	gafas-de-sol	0,014998401	2	3,999999911	0,957912458
44	sandalias	0,013164851	1	1,999999955	0,95959596
45	tenis	0,013164851	1	1,999999955	0,961279461
46	corbata	0,012433324	2	3,999999911	0,964646465
47	maleta	0,011858027	1	1,999999955	0,966329966
48	complemento	0,009620672	1	1,999999955	0,968013468
49	estampado	0,008665665	1	1,999999955	0,96969697
50	pantufla	0,008665665	1	1,999999955	0,971380471
51	traje-de-baño	0,007805458	1	1,999999955	0,973063973
52	elegante	0,007805458	1	1,999999955	0,974747475
53	deportivo	0,00703064	1	1,999999955	0,976430976
54	lavar	0,006332736	1	1,999999955	0,978114478
55	collar	0,006332736	1	1,999999955	0,97979798
56	sostén	0,005704109	1	1,999999955	0,981481481
57	estirar	0,005137885	1	1,999999955	0,983164983
58	zapatilla-de-deporte	0,005137885	1	1,999999955	0,984848485
59	bikini	0,004627867	1	1,999999955	0,986531987
60	colorado	0,004627867	1	1,999999955	0,988215488
61	elástico	0,003381974	1	1,999999955	0,98989899
62	medias	0,002743868	1	1,999999955	0,991582492
63	albornoz	0,002743868	1	1,999999955	0,993265993
64	pulsera	0,002471495	1	1,999999955	0,994949495
65	pajarita	0,002226159	1	1,999999955	0,996632997
66	colgante	0,002005177	1	1,999999955	0,998316498
67	gemelos	0,002005177	1	1,999999955	1

CI03, “partes de la casa (sin muebles)”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	cuarto-de-baño	0,684572935	47	93,99999976	0,111638955
2	cocina	0,684232116	48	95,99999785	0,225653207
3	habitación	0,508510292	34	68,00000072	0,306413302
4	jardín	0,362669349	41	81,99999928	0,403800475
5	salón	0,292330235	27	54,00000215	0,467933492
6	garaje	0,207323298	27	54,00000215	0,532066508
7	techo	0,196412519	21	41,99999869	0,581947743
8	cuarto	0,170653328	11	21,99999988	0,60807601
9	pasillo	0,170356348	14	28,00000012	0,641330166
10	comedor	0,164211139	17	34,00000036	0,681710214
11	escalera	0,11866454	18	36,00000143	0,724465558
12	piso	0,102793455	10	20,00000003	0,748218527
13	puerta	0,096483856	10	20,00000003	0,771971496
14	ventana	0,092840612	10	20,00000003	0,795724466
15	terrazza	0,088384591	12	23,99999946	0,824228029
16	entrada	0,083903477	10	20,00000003	0,847980998
17	cámara	0,065276816	4	7,999999821	0,857482185
18	cuarto-trastero	0,053882025	4	7,999999821	0,866983373
19	balcón	0,053062402	7	14,00000006	0,883610451
20	pared	0,05172161	6	11,99999973	0,897862233
21	dormitorio	0,04389526	3	5,999999866	0,904988124
22	sótano	0,043232173	4	7,999999821	0,914489311
23	recibidor	0,038392711	4	7,999999821	0,923990499
24	cuarto-de-estar	0,037559643	3	5,999999866	0,93111639
25	estudio	0,035409037	6	11,99999973	0,945368171
26	estantería	0,02	1	1,999999955	0,947743468
27	cama	0,02	1	1,999999955	0,950118765
28	ingreso	0,015652196	2	3,999999911	0,954869359
29	casa	0,01403966	1	1,999999955	0,957244656
30	alcoba	0,011763047	1	1,999999955	0,959619952
31	corredor	0,009855601	1	1,999999955	0,961995249
32	azotea	0,009855601	1	1,999999955	0,964370546
33	parque	0,009855601	1	1,999999955	0,966745843
34	planta	0,008257459	1	1,999999955	0,96912114
35	oficina	0,008257459	1	1,999999955	0,971496437
36	aseo	0,008257459	1	1,999999955	0,973871734
37	buhardilla	0,00771309	2	3,999999911	0,978622328

38	cantina	0,007249901	2	3,999999911	0,983372922
39	muro	0,005796595	1	1,999999955	0,985748219
40	despacho	0,005796595	1	1,999999955	0,988123515
41	espacio	0,004856644	1	1,999999955	0,990498812
42	suelo	0,004069111	1	1,999999955	0,992874109
43	ascensor	0,003409281	1	1,999999955	0,995249406
44	portal	0,002856446	1	1,999999955	0,997624703
45	patio	0,002005177	1	1,999999955	1

CI04, “los muebles de la casa”

N°	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	silla	0,763007104	49	98,00000191	0,11036036
2	mesa	0,631066442	42	83,99999738	0,204954955
3	cama	0,59248817	42	83,99999738	0,29954955
4	armario	0,476490438	42	83,99999738	0,394144144
5	sofá	0,42817679	40	80,00000119	0,484234234
6	sillón	0,279084623	27	54,00000215	0,545045045
7	mesilla-de-noche	0,271197557	24	47,99999893	0,599099099
8	horno	0,204054072	20	40,0000006	0,644144144
9	lámpara	0,163142428	15	30,00000119	0,677927928
10	estantería	0,14716956	18	36,00000143	0,718468468
11	lavadora	0,130393282	14	28,00000012	0,75
12	nevera	0,121058576	11	21,99999988	0,774774775
13	cajón	0,083406389	7	14,00000006	0,790540541
14	ducha	0,074727118	10	20,0000003	0,813063063
15	alfombra	0,07177645	9	18,00000072	0,833333333
16	frigorífico	0,069919817	8	15,99999964	0,851351351
17	lavavajillas	0,069059886	8	15,99999964	0,869369369
18	escritorio	0,061806098	8	15,99999964	0,887387387
19	cocina	0,055156551	4	7,999999821	0,896396396
20	espejo	0,045970064	7	14,00000006	0,912162162
21	ventana	0,038818583	3	5,999999866	0,918918919
22	diván	0,037602469	4	7,999999821	0,927927928
23	refrigerador	0,034717917	2	3,999999911	0,932432432
24	puerta	0,032069284	4	7,999999821	0,941441441
25	tele(visión)	0,030936243	4	7,999999821	0,95045045
26	fregadero	0,027987728	2	3,999999911	0,954954955
27	baño	0,021645064	3	5,999999866	0,961711712
28	secadora	0,017261561	2	3,999999911	0,966216216
29	cortina	0,014807723	2	3,999999911	0,970720721
30	(horno)microondas	0,01432275	2	3,999999911	0,975225225
31	mueble	0,012625673	1	1,999999955	0,977477477
32	piano	0,010830854	1	1,999999955	0,97972973
33	ropero	0,00929118	1	1,999999955	0,981981982
34	cuadro	0,00929118	1	1,999999955	0,984234234
35	grifo	0,007970381	1	1,999999955	0,986486486
36	cómoda	0,00586537	1	1,999999955	0,988738739

37	bañadora	0,00586537	1	1,999999955	0,990990991
38	lavabo	0,005031571	1	1,999999955	0,993243243
39	congelador	0,004316302	1	1,999999955	0,995495495
40	almohada	0,002337462	1	1,999999955	0,997747748
41	chimenea	0,002005177	1	1,999999955	1

CI05, “alimentos y bebidas”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	agua	0,635452032	41	81,99999928	0,048235294
2	manzana	0,383473247	33	66,00000262	0,087058824
3	cerveza	0,350475401	27	54,00000215	0,118823529
4	naranja	0,345326662	26	51,99999809	0,149411765
5	jamón	0,325120628	36	72,00000286	0,191764706
6	paella	0,263280928	21	41,99999869	0,216470588
7	pescado	0,257610768	26	51,99999809	0,247058824
8	tomate	0,252142251	27	54,00000215	0,278823529
9	leche	0,24110207	22	43,99999976	0,304705882
10	sangría	0,238604978	21	41,99999869	0,329411765
11	vino	0,199763492	17	34,00000036	0,349411765
12	bocadillo	0,196689829	19	37,99999952	0,371764706
13	plátano	0,189315781	20	40,0000006	0,395294118
14	pan	0,185082033	16	31,99999928	0,414117647
15	pasta	0,181169912	15	30,00000119	0,431764706
16	zumos	0,181150109	14	28,00000012	0,448235294
17	queso	0,166866481	24	47,99999893	0,476470588
18	fresa	0,15604502	17	34,00000036	0,496470588
19	tequila	0,149518877	12	23,99999946	0,510588235
20	limón	0,137682617	13	25,99999905	0,525882353
21	mojito	0,136998296	9	18,00000072	0,536470588
22	zanahoria	0,132242933	13	25,99999905	0,551764706
23	arroz	0,123259909	11	21,99999988	0,564705882
24	pollo	0,121179998	14	28,00000012	0,581176471
25	gaspacho	0,117847465	12	23,99999946	0,595294118
26	patata	0,116352327	13	25,99999905	0,610588235
27	hamburguesa	0,114162713	8	15,99999964	0,62
28	tortilla	0,113463052	12	23,99999946	0,634117647
29	carne	0,095831797	9	18,00000072	0,644705882
30	galleta	0,087273262	8	15,99999964	0,654117647
31	<i>Coca-Cola</i>	0,086081639	6	11,99999973	0,661176471
32	café	0,085591353	13	25,99999905	0,676470588
33	té	0,07781332	7	14,00000006	0,684705882
34	ron	0,076130331	7	14,00000006	0,692941176
35	cebolla	0,075653031	12	23,99999946	0,707058824
36	aceite	0,072743438	8	15,99999964	0,716470588
37	chorizo	0,06541273	8	15,99999964	0,725882353

38	sal	0,060935002	8	15,99999964	0,735294118
39	pimiento	0,060442969	7	14,00000006	0,743529412
40	huevo	0,057096533	7	14,00000006	0,751764706
41	atún	0,056594711	4	7,999999821	0,756470588
42	churros	0,056321803	5	10,00000015	0,762352941
43	azúcar	0,056029309	8	15,99999964	0,771764706
44	aceituna	0,055749707	5	10,00000015	0,777647059
45	ensalada	0,048778988	6	11,99999973	0,784705882
46	mariscos	0,047801897	6	11,99999973	0,791764706
47	taco	0,047506496	3	5,999999866	0,795294118
48	fruta	0,047499754	4	7,999999821	0,8
49	mantequilla	0,045268916	9	18,00000072	0,810588235
50	cereza	0,039653819	5	10,00000015	0,816470588
51	vino-blanco	0,038132474	3	5,999999866	0,82
52	harina	0,036867961	6	11,99999973	0,827058824
53	verdura	0,035862036	3	5,999999866	0,830588235
54	banana	0,03483187	3	5,999999866	0,834117647
55	comida	0,033842344	2	3,999999911	0,836470588
56	pera	0,033388533	4	7,999999821	0,841176471
57	zumode-naranja	0,033277433	2	3,999999911	0,843529412
58	chocolate	0,031912126	4	7,999999821	0,848235294
59	melocotón	0,031867865	3	5,999999866	0,851764706
60	melón	0,031536765	4	7,999999821	0,856470588
61	vino-tinto	0,031387627	3	5,999999866	0,86
62	tapas	0,029777484	3	5,999999866	0,863529412
63	pechuga-de-pollo	0,029664798	3	5,999999866	0,867058824
64	lechuga	0,028647833	4	7,999999821	0,871764706
65	calabacín	0,027727311	3	5,999999866	0,875294118
66	ajo	0,027629152	5	10,00000015	0,881176471
67	zumode-fruta	0,025358286	2	3,999999911	0,883529412
68	hierbabuena	0,024589323	2	3,999999911	0,885882353
69	uva	0,024367057	3	5,999999866	0,889411765
70	piña	0,024075644	3	5,999999866	0,892941176
71	salmón	0,022639114	3	5,999999866	0,896470588
72	vinagre	0,02054783	4	7,999999821	0,901176471
73	berenjena	0,020254388	2	3,999999911	0,903529412
74	burrito	0,02	1	1,999999955	0,904705882
75	palomitas	0,02	1	1,999999955	0,905882353
76	salchicha	0,018834291	3	5,999999866	0,909411765

77	judías	0,018242102	1	1,999999955	0,910588235
78	jugo	0,018242102	1	1,999999955	0,911764706
79	sandía	0,017584488	3	5,999999866	0,915294118
80	cereales	0,016708827	2	3,999999911	0,917647059
81	salado	0,016638717	1	1,999999955	0,918823529
82	pepino	0,016638717	1	1,999999955	0,92
83	espaguete	0,016638717	1	1,999999955	0,921176471
84	cordero	0,015535321	2	3,999999911	0,923529412
85	jamón-cocido	0,01536929	2	3,999999911	0,925882353
86	butifarra	0,015176258	1	1,999999955	0,927058824
87	tarta	0,015096962	2	3,999999911	0,929411765
88	dulce-de-leche	0,014786473	2	3,999999911	0,931764706
89	ternera	0,013842343	1	1,999999955	0,932941176
90	jalapeño	0,013842343	1	1,999999955	0,934117647
91	brécol	0,013842343	1	1,999999955	0,935294118
92	res	0,013842343	1	1,999999955	0,936470588
93	frijoles	0,013486818	2	3,999999911	0,938823529
94	beber	0,012786262	2	3,999999911	0,941176471
95	legumbres	0,012625673	1	1,999999955	0,942352941
96	piña-colada	0,012625673	1	1,999999955	0,943529412
97	hortaliza	0,012625673	1	1,999999955	0,944705882
98	asado	0,011515941	1	1,999999955	0,945882353
99	agua-tónica	0,011515941	1	1,999999955	0,947058824
100	pavo	0,011515941	1	1,999999955	0,948235294
101	perrito-caliente	0,01063735	2	3,999999911	0,950588235
102	embutido	0,01050375	1	1,999999955	0,951764706
103	pastel	0,00986605	2	3,999999911	0,954117647
104	pimienta	0,00968008	2	3,999999911	0,956470588
105	vaca	0,009580524	1	1,999999955	0,957647059
106	pollastre	0,008738445	1	1,999999955	0,958823529
107	albaricoque	0,008738445	1	1,999999955	0,96
108	cerdo	0,008738445	1	1,999999955	0,961176471
109	empanada	0,007970381	1	1,999999955	0,962352941
110	bombón	0,007970381	1	1,999999955	0,963529412
111	ciruela	0,007970381	1	1,999999955	0,964705882
112	calabaza	0,007269825	1	1,999999955	0,965882353
113	coco	0,007269825	1	1,999999955	0,967058824
114	postre	0,006999576	2	3,999999911	0,969411765
115	croqueta	0,006630845	1	1,999999955	0,970588235
116	comer	0,006630845	1	1,999999955	0,971764706

117	bollo	0,006630845	1	1,999999955	0,972941176
118	cóctel	0,006630845	1	1,999999955	0,974117647
119	frito	0,006048028	1	1,999999955	0,975294118
120	pincho	0,006048028	1	1,999999955	0,976470588
121	kiwi	0,006048028	1	1,999999955	0,977647059
122	albahaca	0,005516437	1	1,999999955	0,978823529
123	pepito	0,005516437	1	1,999999955	0,98
124	cococha	0,005516437	1	1,999999955	0,981176471
125	chirimoya	0,005516437	1	1,999999955	0,982352941
126	salchichón	0,005031571	1	1,999999955	0,983529412
127	perejil	0,005031571	1	1,999999955	0,984705882
128	hambre	0,005031571	1	1,999999955	0,985882353
129	pimentón	0,004589322	1	1,999999955	0,987058824
130	sopa	0,004185944	1	1,999999955	0,988235294
131	café-con-leche	0,004185944	1	1,999999955	0,989411765
132	champiñón	0,003818021	1	1,999999955	0,990588235
133	café-solo	0,003818021	1	1,999999955	0,991764706
134	manteca	0,003818021	1	1,999999955	0,992941176
135	apio	0,003482437	1	1,999999955	0,994117647
136	tinto-de-verano	0,003482437	1	1,999999955	0,995294118
137	espárrago	0,002897164	1	1,999999955	0,996470588
138	camarón	0,002897164	1	1,999999955	0,997647059
139	cangrejo	0,002642518	1	1,999999955	0,998823529
140	langosta	0,002410254	1	1,999999955	1

CI06, “objetos colocados en la mesa para la comida”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	cuchillo	0,76044482	46	92,00000167	0,114427861
2	plato	0,696026623	44	87,99999952	0,223880597
3	vaso	0,620776772	41	81,99999928	0,325870647
4	cuchara	0,600198209	43	86,00000143	0,432835821
5	tenedor	0,565229416	39	77,99999714	0,529850746
6	mantel	0,477595717	39	77,99999714	0,626865672
7	botella	0,34520179	30	60,00000238	0,701492537
8	servilleta	0,247670174	24	47,99999893	0,76119403
9	sal	0,112899065	11	21,99999988	0,788557214
10	cucharilla	0,110618323	8	15,99999964	0,808457711
11	aceite	0,093990147	11	21,99999988	0,835820896
12	pimienta	0,092433512	10	20,0000003	0,860696517
13	vinagre	0,061547745	8	15,99999964	0,880597015
14	agua	0,057554368	5	10,00000015	0,893034826
15	copa	0,057141647	4	7,999999821	0,902985075
16	jarra	0,042891916	4	7,999999821	0,912935323
17	ensaladera	0,033419218	2	3,999999911	0,917910448
18	bebida	0,03242179	3	5,999999866	0,925373134
19	pan	0,029163193	3	5,999999866	0,932835821
20	plato-hondo	0,026290569	2	3,999999911	0,937810945
21	comer	0,020468483	3	5,999999866	0,945273632
22	papel	0,018244579	2	3,999999911	0,950248756
23	navaja	0,017719747	1	1,999999955	0,952736318
24	cubierto	0,017367754	3	5,999999866	0,960199005
25	beber	0,015298693	2	3,999999911	0,965174129
26	pasta	0,009673755	1	1,999999955	0,967661692
27	utensilio	0,008570824	1	1,999999955	0,970149254
28	vino	0,008570824	1	1,999999955	0,972636816
29	centro-de-mesa	0,008164369	2	3,999999911	0,97761194
30	vajilla	0,007593641	1	1,999999955	0,980099502
31	comida	0,005960808	1	1,999999955	0,982587065
32	vino-tinto	0,005960808	1	1,999999955	0,985074627
33	vino-seco	0,0052812	1	1,999999955	0,987562189
34	sentarse	0,004679076	1	1,999999955	0,990049751
35	fruta	0,003672951	1	1,999999955	0,997512438

CI07, “la cocina y sus utensilios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	horno	0,445441127	27	54,00000215	0,09375
2	sartén	0,400939256	29	57,99999833	0,194444444
3	olla	0,270124078	18	36,00000143	0,256944444
4	plato	0,23930271	18	36,00000143	0,319444444
5	cuchara	0,224970534	18	36,00000143	0,381944444
6	cuchillo	0,205783516	16	31,99999928	0,4375
7	nevera	0,142611116	11	21,99999988	0,475694444
8	tenedor	0,126008689	9	18,00000072	0,506944444
9	(horno)microondas	0,110641956	10	20,0000003	0,541666667
10	colador	0,099093094	9	18,00000072	0,572916667
11	cacerola	0,095629901	8	15,99999964	0,600694444
12	vaso	0,074722193	6	11,99999973	0,621527778
13	cuenca	0,068861157	5	10,00000015	0,638888889
14	batidor	0,068684436	7	14,00000006	0,663194444
15	paella	0,065068699	6	11,99999973	0,684027778
16	lavavajillas	0,065022886	6	11,99999973	0,704861111
17	cubierto	0,054526702	4	7,999999821	0,71875
18	cucharón	0,048147786	5	10,00000015	0,736111111
19	cazo	0,038141645	4	7,999999821	0,75
20	frigorífico	0,037080728	3	5,999999866	0,760416667
21	cazuela	0,037048217	4	7,999999821	0,774305556
22	rallador	0,033937994	3	5,999999866	0,784722222
23	refrigerador	0,033164848	2	3,999999911	0,791666667
24	cocinar	0,029757585	4	7,999999821	0,805555556
25	fuego	0,028861156	3	5,999999866	0,815972222
26	esponja	0,028012257	3	5,999999866	0,826388889
27	mesa	0,027170362	4	7,999999821	0,840277778
28	sacacorchos	0,021400414	3	5,999999866	0,850694444
29	olla-a-presión	0,020854296	2	3,999999911	0,857638889
30	cafetera	0,02	1	1,999999955	0,861111111
31	licuadora	0,019346589	2	3,999999911	0,868055556
32	afilador	0,018868959	2	3,999999911	0,875
33	balanza	0,01733133	2	3,999999911	0,881944444
34	embudo	0,016226429	1	1,999999955	0,885416667
35	navaja	0,016226429	1	1,999999955	0,888888889
36	mantel	0,016226429	1	1,999999955	0,892361111
37	cucharilla	0,015308793	2	3,999999911	0,899305556

38	rodillo	0,014435614	2	3,999999911	0,90625
39	batidora	0,013164851	1	1,999999955	0,909722222
40	mezclador	0,013164851	1	1,999999955	0,913194444
41	trapo	0,013164851	1	1,999999955	0,916666667
42	abrebotellas	0,011930292	3	5,999999866	0,927083333
43	hervidor	0,010680926	1	1,999999955	0,930555556
44	pelador	0,010680926	1	1,999999955	0,934027778
45	grifo	0,010680926	1	1,999999955	0,9375
46	cortar	0,010331975	2	3,999999911	0,944444444
47	molde	0,008665665	1	1,999999955	0,947916667
48	botella	0,008665665	1	1,999999955	0,951388889
49	pala-de-freír	0,008665665	1	1,999999955	0,954861111
50	freír	0,008665665	1	1,999999955	0,958333333
51	taza	0,00703064	1	1,999999955	0,961805556
52	espumadera	0,00703064	1	1,999999955	0,965277778
53	espátula	0,00703064	1	1,999999955	0,96875
54	servilleta	0,005704109	1	1,999999955	0,972222222
55	calentar	0,005704109	1	1,999999955	0,975694444
56	mezclar	0,005704109	1	1,999999955	0,979166667
57	silla	0,005704109	1	1,999999955	0,982638889
58	abrelatas	0,005051436	2	3,999999911	0,989583333
59	molinillo	0,004627867	1	1,999999955	0,993055556
60	escurridor	0,003754687	1	1,999999955	0,996527778
61	agua	0,003754687	1	1,999999955	1

CI08, “la escuela: muebles y materiales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	bolí(grafo)	0,624877334	42	83,99999738	0,06095791
2	silla	0,560440719	36	72,00000286	0,113207547
3	pizarra	0,560122848	39	77,99999714	0,169811321
4	lápiz	0,499719381	37	74,00000095	0,223512337
5	libro	0,481130153	36	72,00000286	0,275761974
6	cuaderno	0,40448308	33	66,00000262	0,323657475
7	mesa	0,388972521	25	50	0,359941945
8	estuche	0,379247427	28	56,00000024	0,400580552
9	mochila	0,323707402	26	51,99999809	0,438316401
10	profe(sor)	0,278022587	36	72,00000286	0,490566038
11	estudiar	0,226635858	28	56,00000024	0,531204644
12	ordenador	0,209623486	21	41,99999869	0,561683599
13	borrador	0,188381001	17	34,00000036	0,586357039
14	estudiante	0,133570641	18	36,00000143	0,612481858
15	escribir	0,12941815	17	34,00000036	0,637155298
16	leer	0,123478919	17	34,00000036	0,661828737
17	papel	0,118069962	9	18,00000072	0,674891147
18	sacapuntas	0,109751455	12	23,99999946	0,692307692
19	alumno	0,083195001	14	28,00000012	0,712626996
20	rotulador	0,078340262	9	18,00000072	0,725689405
21	clase	0,074912056	9	18,00000072	0,738751814
22	apuntes	0,068062894	7	14,00000006	0,748911466
23	enseñar	0,066228643	7	14,00000006	0,759071118
24	examen	0,062445763	10	20,00000003	0,773584906
25	armario	0,061703347	4	7,999999821	0,779390421
26	pupitre	0,05820176	5	10,00000015	0,786647315
27	goma-de-borrar	0,052397802	4	7,999999821	0,79245283
28	marcador	0,0495185	7	14,00000006	0,802612482
29	color	0,04371915	5	10,00000015	0,809869376
30	tiza	0,043310452	6	11,99999973	0,818577649
31	tableta	0,03977184	6	11,99999973	0,827285922
32	regla	0,037804585	5	10,00000015	0,834542816
33	aprender	0,037623931	6	11,99999973	0,843251089
34	tijeras	0,031283285	6	11,99999973	0,851959361
35	ejercicio	0,029524386	3	5,999999866	0,856313498
36	papelera	0,029491721	2	3,999999911	0,859216255
37	escuchar	0,028645176	4	7,999999821	0,865021771

38	registro	0,027304316	2	3,999999911	0,867924528
39	ventana	0,027042652	3	5,999999866	0,872278665
40	diccionario	0,026486574	4	7,999999821	0,87808418
41	goma	0,023949379	4	7,999999821	0,883889695
42	despacho	0,023637723	5	10,00000015	0,891146589
43	aula	0,021933107	2	3,999999911	0,894049347
44	asignatura	0,018875049	4	7,999999821	0,899854862
45	manual	0,017925216	1	1,999999955	0,901306241
46	cátedra	0,017925216	1	1,999999955	0,90275762
47	historia	0,01501652	2	3,999999911	0,905660377
48	oficina	0,01439903	1	1,999999955	0,907111756
49	curso	0,013700644	2	3,999999911	0,910014514
50	bolso	0,013474131	2	3,999999911	0,912917271
51	matemática	0,013458719	2	3,999999911	0,915820029
52	nota	0,013458719	2	3,999999911	0,918722787
53	puerta	0,013159713	2	3,999999911	0,921625544
54	tomar-apuntes	0,013151759	2	3,999999911	0,924528302
55	perchero	0,013099012	3	5,999999866	0,928882438
56	pintura	0,012905287	1	1,999999955	0,930333817
57	aula-magna	0,012905287	1	1,999999955	0,931785196
58	secretaría	0,012307751	3	5,999999866	0,936139332
59	pastel	0,011566503	1	1,999999955	0,937590711
60	calculador	0,0110055	2	3,999999911	0,940493469
61	subrayar	0,010811167	2	3,999999911	0,943396226
62	pensar	0,00929118	1	1,999999955	0,944847605
63	enseñanza	0,00929118	1	1,999999955	0,946298984
64	escuela	0,00929118	1	1,999999955	0,947750363
65	computador	0,00929118	1	1,999999955	0,949201742
66	aprobar	0,008327321	1	1,999999955	0,95065312
67	opinar	0,008327321	1	1,999999955	0,952104499
68	página	0,008327321	1	1,999999955	0,953555878
69	colegio	0,007463451	1	1,999999955	0,955007257
70	ordenar	0,007463451	1	1,999999955	0,956458636
71	pluma	0,007463451	1	1,999999955	0,957910015
72	actividad	0,007463451	1	1,999999955	0,959361393
73	maestro	0,007463451	1	1,999999955	0,960812772
74	deberes	0,007335746	2	3,999999911	0,96371553
75	bachillerato	0,006689199	1	1,999999955	0,965166909
76	ciencia	0,006689199	1	1,999999955	0,966618287
77	aprobar-examen	0,006689199	1	1,999999955	0,968069666

78	repetir	0,006689199	1	1,999999955	0,969521045
79	prestar-atención	0,005995267	1	1,999999955	0,970972424
80	hoja	0,005995267	1	1,999999955	0,972423803
81	repasar	0,005373323	1	1,999999955	0,973875181
82	mapa	0,005373323	1	1,999999955	0,97532656
83	revisar	0,004815899	1	1,999999955	0,976777939
84	director	0,004815899	1	1,999999955	0,978229318
85	pintar	0,004815899	1	1,999999955	0,979680697
86	tabla-periódica	0,004815899	1	1,999999955	0,981132075
87	borrar	0,004815899	1	1,999999955	0,982583454
88	tutoría	0,004316302	1	1,999999955	0,984034833
89	pantalla	0,004316302	1	1,999999955	0,985486212
90	carpeta	0,004316302	1	1,999999955	0,986937591
91	entregar	0,003868532	1	1,999999955	0,98838897
92	luz	0,003868532	1	1,999999955	0,989840348
93	reseña	0,003467214	1	1,999999955	0,991291727
94	literatura	0,003107528	1	1,999999955	0,992743106
95	informe	0,002785156	1	1,999999955	0,994194485
96	resumen	0,002496226	1	1,999999955	0,995645864
97	opinión	0,002237269	1	1,999999955	0,997097242
98	química	0,002237269	1	1,999999955	0,998548621
99	valoración	0,002005177	1	1,999999955	1

CI09, “iluminación y calefacción”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	lámpara	0,804322839	47	93,99999976	0,139053254
2	luz	0,515237451	32	63,99999857	0,233727811
3	fuego	0,330727935	24	47,99999893	0,304733728
4	bombilla	0,254878402	20	40,0000006	0,363905325
5	calor	0,243989661	24	47,99999893	0,434911243
6	sol	0,208063886	18	36,00000143	0,48816568
7	radiador	0,167096436	15	30,00000119	0,532544379
8	iluminar	0,134498641	12	23,99999946	0,568047337
9	calefacción	0,124975942	10	20,0000003	0,597633136
10	encender	0,114825211	10	20,0000003	0,627218935
11	apagar	0,091260664	8	15,99999964	0,650887574
12	aire- acondicionado	0,089713357	7	14,00000006	0,671597633
13	calentar	0,085136369	7	14,00000006	0,692307692
14	mechero	0,082267769	10	20,0000003	0,721893491
15	farola	0,074488804	6	11,99999973	0,73964497
16	horno	0,071512699	6	11,99999973	0,75739645
17	vela	0,061925936	8	15,99999964	0,781065089
18	frío	0,058550335	7	14,00000006	0,801775148
19	chimenea	0,057646122	5	10,00000015	0,816568047
20	electricidad	0,053551454	5	10,00000015	0,831360947
21	araña	0,02807932	2	3,999999911	0,837278107
22	madera	0,025914831	4	7,999999821	0,849112426
23	neón	0,023895262	2	3,999999911	0,855029586
24	led	0,023526095	2	3,999999911	0,860946746
25	linterna	0,020166166	2	3,999999911	0,866863905
26	calefactor	0,02	1	1,999999955	0,869822485
27	calentador	0,02	1	1,999999955	0,872781065
28	calorífico	0,02	1	1,999999955	0,875739645
29	iluminado	0,018910699	3	5,999999866	0,884615385
30	electricista	0,018681511	2	3,999999911	0,890532544
31	estufa	0,01811306	2	3,999999911	0,896449704
32	estrella	0,017559642	2	3,999999911	0,902366864
33	candela	0,017559642	2	3,999999911	0,908284024
34	hogar	0,016756885	1	1,999999955	0,911242604
35	abanico	0,016756885	1	1,999999955	0,914201183
36	caliente	0,015832158	2	3,999999911	0,920118343
37	ventilador	0,015652196	2	3,999999911	0,926035503

38	energía-eléctrica	0,014062519	3	5,999999866	0,934911243
39	refrescador	0,01403966	1	1,999999955	0,937869822
40	hilo-eléctrico	0,01403966	1	1,999999955	0,940828402
41	refriar	0,011763047	1	1,999999955	0,943786982
42	candelabro	0,011763047	1	1,999999955	0,946745562
43	claro	0,01166674	2	3,999999911	0,952662722
44	sombra	0,010987575	2	3,999999911	0,958579882
45	ultravioleta	0,009855601	1	1,999999955	0,961538462
46	lámina	0,009855601	1	1,999999955	0,964497041
47	insolación	0,008257459	1	1,999999955	0,967455621
48	panel-solar	0,008257459	1	1,999999955	0,970414201
49	ventana	0,006918464	1	1,999999955	0,973372781
50	(horno)microondas	0,006918464	1	1,999999955	0,976331361
51	oscuro	0,006918464	1	1,999999955	0,979289941
52	acondicionar	0,006918464	1	1,999999955	0,982248521
53	termo(sifón)	0,006918464	1	1,999999955	0,985207101
54	nevera	0,005796595	1	1,999999955	0,98816568
55	eléctrico	0,005796595	1	1,999999955	0,99112426
56	verano	0,004856644	1	1,999999955	0,99408284
57	aire	0,004856644	1	1,999999955	0,99704142
58	luminoso	0,004069111	1	1,999999955	1

CI10, “la ciudad”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	calle	0,88570106	50	100	0,070422535
2	plaza	0,461875796	29	57,99999833	0,111267606
3	iglesia	0,382334292	27	54,00000215	0,149295775
4	coche	0,301854372	25	50	0,184507042
5	tienda	0,299393058	29	57,99999833	0,225352113
6	parque	0,290621251	28	56,00000024	0,264788732
7	escuela	0,267535418	22	43,99999976	0,295774648
8	(auto)bús	0,253651202	27	54,00000215	0,333802817
9	estación	0,245425075	28	56,00000024	0,373239437
10	casa	0,205742136	19	37,99999952	0,4
11	carretera	0,195171684	12	23,99999946	0,416901408
12	mercado	0,19052574	14	28,00000012	0,436619718
13	semáforo	0,160329968	16	31,99999928	0,45915493
14	hospital	0,159847409	18	36,00000143	0,484507042
15	museo	0,149430767	19	37,99999952	0,511267606
16	restaurante	0,149250537	19	37,99999952	0,538028169
17	monumento	0,142489821	11	21,99999988	0,553521127
18	ayuntamiento	0,128914341	13	25,99999905	0,571830986
19	avenida	0,115746886	8	15,99999964	0,583098592
20	cine	0,115698859	11	21,99999988	0,598591549
21	bici(cleta)	0,1130739	9	18,00000072	0,611267606
22	barrio	0,112609319	7	14,00000006	0,621126761
23	edificio	0,098870285	9	18,00000072	0,633802817
24	jardín	0,085557312	8	15,99999964	0,645070423
25	bar	0,081786118	12	23,99999946	0,661971831
26	árbol	0,070514552	6	11,99999973	0,670422535
27	universidad	0,070020616	9	18,00000072	0,683098592
28	rascacielos	0,066918164	8	15,99999964	0,694366197
29	paso-de-cebra	0,066771328	5	10,00000015	0,701408451
30	parada	0,062388927	10	20,0000003	0,715492958
31	palacio	0,060984746	6	11,99999973	0,723943662
32	cruce	0,057482496	7	14,00000006	0,733802817
33	farmacia	0,056668267	8	15,99999964	0,745070423
34	aeropuerto	0,054216936	7	14,00000006	0,754929577
35	metro	0,052684359	6	11,99999973	0,763380282
36	supermercado	0,052303996	5	10,00000015	0,770422535
37	cruzar	0,050927822	5	10,00000015	0,777464789

38	pasear	0,049926862	6	11,99999973	0,785915493
39	tren	0,047735348	6	11,99999973	0,794366197
40	biblioteca	0,04648364	5	10,00000015	0,801408451
41	disco(teca)	0,042612225	5	10,00000015	0,808450704
42	banco	0,041409049	5	10,00000015	0,815492958
43	negocio	0,041311264	4	7,999999821	0,821126761
44	policía	0,040155392	7	14,00000006	0,830985915
45	caminar	0,038327079	3	5,999999866	0,835211268
46	vía	0,037584949	3	5,999999866	0,83943662
47	taxi	0,035862818	3	5,999999866	0,843661972
48	centro	0,029632729	2	3,999999911	0,846478873
49	trabajar	0,029310055	3	5,999999866	0,850704225
50	rincón	0,029072981	2	3,999999911	0,853521127
51	autovía	0,027350849	2	3,999999911	0,856338028
52	vivir	0,027239017	3	5,999999866	0,86056338
53	tráfico	0,026946979	2	3,999999911	0,863380282
54	piso	0,026593886	5	10,00000015	0,870422535
55	visitar	0,02643091	3	5,999999866	0,874647887
56	señal-de-tráfico	0,02538435	3	5,999999866	0,878873239
57	peatón	0,024754172	2	3,999999911	0,881690141
58	apartamento	0,023338107	2	3,999999911	0,884507042
59	habitación	0,023106847	2	3,999999911	0,887323944
60	esquina	0,023106847	2	3,999999911	0,890140845
61	andar	0,02084025	2	3,999999911	0,892957746
62	panadería	0,020262009	2	3,999999911	0,895774648
63	país	0,02	1	1,999999955	0,897183099
64	urbano	0,02	1	1,999999955	0,898591549
65	derecha	0,018918285	2	3,999999911	0,901408451
66	izquierda	0,018918285	2	3,999999911	0,904225352
67	despacho	0,018470034	2	3,999999911	0,907042254
68	alcalde	0,018096749	1	1,999999955	0,908450704
69	oficio	0,018096749	1	1,999999955	0,909859155
70	pueblo	0,018096749	1	1,999999955	0,911267606
71	estatua	0,017060071	2	3,999999911	0,914084507
72	establecimiento	0,015488982	2	3,999999911	0,916901408
73	teatro	0,014845178	3	5,999999866	0,921126761
74	manzana	0,014816364	1	1,999999955	0,922535211
75	estanco	0,013406401	1	1,999999955	0,923943662
76	pastelería	0,013406401	1	1,999999955	0,925352113
77	paseo	0,012681305	2	3,999999911	0,928169014

78	quiosco	0,012681305	2	3,999999911	0,930985915
79	redonda	0,012130613	1	1,999999955	0,932394366
80	paseante	0,012130613	1	1,999999955	0,933802817
81	comprar	0,012130613	1	1,999999955	0,935211268
82	castro	0,012130613	1	1,999999955	0,936619718
83	turista	0,012130613	1	1,999999955	0,938028169
84	alcaldía	0,012130613	1	1,999999955	0,93943662
85	moto(cicleta)	0,012130613	1	1,999999955	0,94084507
86	café	0,011820192	2	3,999999911	0,943661972
87	centro-comercial	0,010976233	1	1,999999955	0,945070423
88	colegio	0,010976233	1	1,999999955	0,946478873
89	dar-una-vuelta	0,010976233	1	1,999999955	0,947887324
90	estadio	0,010976233	1	1,999999955	0,949295775
91	correr	0,010901271	2	3,999999911	0,952112676
92	trabajo	0,009931706	1	1,999999955	0,953521127
93	ir-de-compras	0,009931706	1	1,999999955	0,954929577
94	casco-antiguo	0,009931706	1	1,999999955	0,956338028
95	catedral	0,008986579	1	1,999999955	0,957746479
96	habitante	0,008986579	1	1,999999955	0,95915493
97	zona-peatonal	0,008986579	1	1,999999955	0,96056338
98	albergue	0,008131393	1	1,999999955	0,961971831
99	viajar	0,008131393	1	1,999999955	0,963380282
100	acera	0,008131393	1	1,999999955	0,964788732
101	letrero	0,007357589	1	1,999999955	0,966197183
102	peluquería	0,007357589	1	1,999999955	0,967605634
103	librería	0,006657422	1	1,999999955	0,969014085
104	estudiar	0,006657422	1	1,999999955	0,970422535
105	centro-histórico	0,006657422	1	1,999999955	0,971830986
106	planta	0,006253994	2	3,999999911	0,974647887
107	vida-nocturna	0,006023884	1	1,999999955	0,976056338
108	playa	0,006023884	1	1,999999955	0,977464789
109	gimnasio	0,006023884	1	1,999999955	0,978873239
110	parquímetro	0,005450636	1	1,999999955	0,98028169
111	villa	0,005450636	1	1,999999955	0,981690141
112	capital	0,005450636	1	1,999999955	0,983098592
113	puerto	0,005450636	1	1,999999955	0,984507042
114	barco	0,004931939	1	1,999999955	0,985915493
115	conocer	0,004931939	1	1,999999955	0,987323944
116	antiguo	0,004931939	1	1,999999955	0,988732394

117	río	0,004462603	1	1,999999955	0,990140845
118	explorar	0,004462603	1	1,999999955	0,991549296
119	carro	0,00403793	1	1,999999955	0,992957746
120	comisaría	0,00403793	1	1,999999955	0,994366197
121	comer	0,00403793	1	1,999999955	0,995774648
122	cementerio	0,00365367	1	1,999999955	0,997183099
123	aparcar	0,00365367	1	1,999999955	0,998591549
124	tribunal	0,002449129	1	1,999999955	1

CI11, “el campo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	árbol	0,560603023	34	69,38775778	0,084577114
2	tierra	0,406464636	25	51,02040768	0,146766169
3	animal	0,384502292	27	55,10203838	0,213930348
4	flor	0,293008149	17	34,69387889	0,256218905
5	hierba	0,231443837	18	36,73469424	0,300995025
6	campo	0,208890676	12	24,48979616	0,330845771
7	cortijo	0,189752907	17	34,69387889	0,373134328
8	vaca	0,182809025	17	34,69387889	0,415422886
9	campesino	0,179148316	11	22,44897932	0,44278607
10	tractor	0,15208064	15	30,6122452	0,480099502
11	cultivar	0,125214204	9	18,36734712	0,502487562
12	río	0,11026711	9	18,36734712	0,524875622
13	caballo	0,102199025	12	24,48979616	0,554726368
14	perro	0,095713869	9	18,36734712	0,577114428
15	agricultor	0,087898403	7	14,28571492	0,594527363
16	gallina	0,077377662	7	14,28571492	0,611940299
17	fruta	0,074415453	7	14,28571492	0,629353234
18	casa	0,067686468	4	8,163265139	0,639303483
19	maíz	0,066700958	7	14,28571492	0,656716418
20	verdura	0,064711459	7	14,28571492	0,674129353
21	jardín	0,055759355	3	6,12244904	0,68159204
22	pollo	0,047423877	4	8,163265139	0,691542289
23	naturaleza	0,046243619	3	6,12244904	0,699004975
24	planta	0,044841055	4	8,163265139	0,708955224
25	insecto	0,044471521	5	10,20408198	0,721393035
26	burro	0,04298909	4	8,163265139	0,731343284
27	granjero	0,040300451	3	6,12244904	0,73880597
28	agua	0,039910052	4	8,163265139	0,748756219
29	cerdo	0,039483272	5	10,20408198	0,76119403
30	trigo	0,034487631	4	8,163265139	0,771144279
31	cortar	0,033667274	2	4,081632569	0,776119403
32	pasear	0,032913689	3	6,12244904	0,78358209
33	oveja	0,032762613	4	8,163265139	0,793532338
34	sendero	0,030617818	2	4,081632569	0,798507463
35	cultivación	0,026289994	2	4,081632569	0,803482587
36	verde	0,025154179	4	8,163265139	0,813432836
37	trabajar	0,023205256	2	4,081632569	0,81840796

38	cultivador	0,022967545	2	4,081632569	0,823383085
39	al-aire-libre	0,022769878	2	4,081632569	0,828358209
40	gallo	0,022769878	2	4,081632569	0,833333333
41	tranquilidad	0,022769878	2	4,081632569	0,838308458
42	hacienda	0,020408163	1	2,040816285	0,84079602
43	bosque	0,020408163	1	2,040816285	0,843283582
44	grano	0,020156274	2	4,081632569	0,848258706
45	rastro	0,02009817	2	4,081632569	0,853233831
46	mariposa	0,017847592	3	6,12244904	0,860696517
47	parque	0,017675597	1	2,040816285	0,86318408
48	vino	0,016895296	2	4,081632569	0,868159204
49	conejo	0,016408119	2	4,081632569	0,873134328
50	ratón	0,015542889	2	4,081632569	0,878109453
51	vegetación	0,015308909	1	2,040816285	0,880597015
52	cabra	0,015308909	1	2,040816285	0,883084577
53	espantapájaros	0,013922942	2	4,081632569	0,888059701
54	puerco	0,013922942	2	4,081632569	0,893034826
55	gato	0,013259112	1	2,040816285	0,895522388
56	pradería	0,013259112	1	2,040816285	0,89800995
57	aire	0,013259112	1	2,040816285	0,900497512
58	agricultura	0,013259112	1	2,040816285	0,902985075
59	pozo	0,013259112	1	2,040816285	0,905472637
60	villa	0,013259112	1	2,040816285	0,907960199
61	granja	0,012812721	2	4,081632569	0,912935323
62	mosquito	0,011483773	1	2,040816285	0,915422886
63	verme	0,011483773	1	2,040816285	0,917910448
64	finca	0,011483773	1	2,040816285	0,92039801
65	res	0,009946144	1	2,040816285	0,922885572
66	pato	0,009946144	1	2,040816285	0,925373134
67	cereales	0,009946144	1	2,040816285	0,927860697
68	irrigación	0,009946144	1	2,040816285	0,930348259
69	galera	0,009946144	1	2,040816285	0,932835821
70	red	0,008614398	1	2,040816285	0,935323383
71	huerto	0,008614398	1	2,040816285	0,937810945
72	ganso	0,008614398	1	2,040816285	0,940298507
73	fruto	0,008614398	1	2,040816285	0,94278607
74	cultivo	0,007460967	1	2,040816285	0,945273632
75	plantar	0,007460967	1	2,040816285	0,947761194
76	cereza	0,007460967	1	2,040816285	0,950248756
77	abeja	0,007460967	1	2,040816285	0,952736318

78	instrumento	0,007460967	1	2,040816285	0,955223881
79	arar	0,006461975	1	2,040816285	0,957711443
80	hoja	0,006461975	1	2,040816285	0,960199005
81	paisaje	0,006461975	1	2,040816285	0,962686567
82	toro	0,005596744	1	2,040816285	0,965174129
83	azada	0,005596744	1	2,040816285	0,967661692
84	bici(cleta)	0,005596744	1	2,040816285	0,970149254
85	pájaro	0,005596744	1	2,040816285	0,972636816
86	cantina	0,004847364	1	2,040816285	0,975124378
87	a-pie	0,004847364	1	2,040816285	0,97761194
88	canal	0,004847364	1	2,040816285	0,980099502
89	chivo	0,004847364	1	2,040816285	0,982587065
90	uva	0,004198323	1	2,040816285	0,985074627
91	soja	0,004198323	1	2,040816285	0,987562189
92	rural	0,004198323	1	2,040816285	0,990049751
93	pavo	0,004198323	1	2,040816285	0,992537313
94	ave	0,003636185	1	2,040816285	0,995024876
95	hormiga	0,002362416	1	2,040816285	0,997512438
96	mosca	0,002046099	1	2,040816285	1

CI12, “medios de transporte”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	coche	0,843886614	48	95,99999785	0,091954023
2	(auto)bús	0,666665137	46	92,00000167	0,180076628
3	tren	0,652358174	47	93,99999976	0,270114943
4	avión	0,633473039	49	98,00000191	0,363984674
5	bici(cleta)	0,599127591	48	95,99999785	0,455938697
6	moto(cicleta)	0,298715442	34	68,00000072	0,521072797
7	metro	0,272853822	31	62,00000048	0,58045977
8	taxi	0,26460433	26	51,99999809	0,630268199
9	barco	0,231958166	27	54,00000215	0,681992337
10	tranvía	0,145628154	17	34,00000036	0,714559387
11	a-pie	0,134974703	14	28,00000012	0,74137931
12	carro	0,104649633	6	11,99999973	0,752873563
13	conducir	0,083559535	14	28,00000012	0,779693487
14	caballo	0,053501382	8	15,99999964	0,795019157
15	coger	0,046200946	5	10,00000015	0,804597701
16	camión	0,041634168	4	7,999999821	0,812260536
17	viajar	0,03878564	6	11,99999973	0,823754789
18	billete	0,036617614	7	14,00000006	0,837164751
19	helicóptero	0,032096397	6	11,99999973	0,848659004
20	nave	0,030609636	4	7,999999821	0,856321839
21	patinetes	0,028085528	4	7,999999821	0,863984674
22	carné-de-conducir	0,026464157	5	10,00000015	0,873563218
23	ciclomotor	0,02550108	3	5,999999866	0,879310345
24	tractor	0,022508195	2	3,999999911	0,883141762
25	tomar	0,022412691	3	5,999999866	0,888888889
26	yate	0,020108305	3	5,999999866	0,894636015
27	salir	0,02	1	1,999999955	0,896551724
28	velero	0,019494442	2	3,999999911	0,900383142
29	andar	0,018677335	3	5,999999866	0,906130268
30	estación	0,017322086	1	1,999999955	0,908045977
31	subir	0,017322086	1	1,999999955	0,909961686
32	embarcación	0,016884219	2	3,999999911	0,913793103
33	parar	0,014774845	2	3,999999911	0,917624521
34	tándem	0,013926919	2	3,999999911	0,921455939
35	monopatín	0,012993929	1	1,999999955	0,923371648
36	globo-aerostático	0,012428233	3	5,999999866	0,929118774

37	correr	0,011817545	2	3,999999911	0,932950192
38	pasajero	0,01131504	3	5,999999866	0,938697318
39	navío	0,011254097	1	1,999999955	0,940613027
40	góndola	0,011254097	1	1,999999955	0,942528736
41	autovía	0,009048271	2	3,999999911	0,946360153
42	transportar	0,008442109	1	1,999999955	0,948275862
43	conductor	0,008313878	2	3,999999911	0,95210728
44	llevar	0,007311748	1	1,999999955	0,954022989
45	gasolina	0,007311748	1	1,999999955	0,955938697
46	canoa	0,006332736	1	1,999999955	0,957854406
47	caminar	0,006332736	1	1,999999955	0,959770115
48	navegar	0,006332736	1	1,999999955	0,961685824
49	submarino	0,006332736	1	1,999999955	0,963601533
50	rápido	0,006236544	2	3,999999911	0,96743295
51	nave-espacial	0,006172657	2	3,999999911	0,971264368
52	perder	0,005484809	1	1,999999955	0,973180077
53	despacio	0,005401497	2	3,999999911	0,977011494
54	atraso	0,004750417	1	1,999999955	0,978927203
55	transporte-público	0,004750417	1	1,999999955	0,980842912
56	triciclo	0,004114356	1	1,999999955	0,982758621
57	pedalear	0,004114356	1	1,999999955	0,98467433
58	parapente	0,004114356	1	1,999999955	0,986590038
59	caravana	0,003563461	1	1,999999955	0,988505747
60	asiento	0,003563461	1	1,999999955	0,990421456
61	globo-dirigible	0,003563461	1	1,999999955	0,992337165
62	confortable	0,003086329	1	1,999999955	0,994252874
63	motor	0,002673083	1	1,999999955	0,996168582
64	funicular	0,002673083	1	1,999999955	0,998084291
65	camioneta	0,002315168	1	1,999999955	1

CI13, “trabajos del campo y del jardín”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	cortar	0,405599326	29	57,99999833	0,140096618
2	plantar	0,258907467	20	40,0000006	0,236714976
3	jardinero	0,201662749	16	31,99999928	0,314009662
4	cultivar	0,193461567	13	25,99999905	0,376811594
5	regar	0,139385074	11	21,99999988	0,429951691
6	campesino	0,136261523	8	15,99999964	0,468599034
7	flor	0,106400467	8	15,99999964	0,507246377
8	arar	0,095892325	8	15,99999964	0,54589372
9	dar-agua	0,087212503	11	21,99999988	0,599033816
10	granjero	0,086628109	6	11,99999973	0,628019324
11	hierba	0,077896349	7	14,00000006	0,661835749
12	cavar	0,062595688	5	10,00000015	0,685990338
13	campo	0,059999999	3	5,999999866	0,700483092
14	sembrar	0,047851015	6	11,99999973	0,729468599
15	agricultor	0,044765636	3	5,999999866	0,743961353
16	criador	0,041862484	3	5,999999866	0,758454106
17	árbol	0,032229085	3	5,999999866	0,77294686
18	jardín	0,02879806	2	3,999999911	0,782608696
19	vaquero	0,02879806	2	3,999999911	0,792270531
20	tierra	0,025373323	2	3,999999911	0,801932367
21	recoger	0,023585768	5	10,00000015	0,826086957
22	obrero	0,020733206	2	3,999999911	0,835748792
23	vender	0,02	1	1,999999955	0,84057971
24	erradicar	0,02	1	1,999999955	0,845410628
25	cultivación	0,02	1	1,999999955	0,850241546
26	agricultura	0,02	1	1,999999955	0,855072464
27	ingeniero	0,02	1	1,999999955	0,859903382
28	coger	0,019608459	3	5,999999866	0,874396135
29	trabajar	0,017830055	2	3,999999911	0,884057971
30	crecer	0,014615178	3	5,999999866	0,898550725
31	irrigación	0,01439903	1	1,999999955	0,903381643
32	agua	0,01439903	1	1,999999955	0,90821256
33	zapar	0,012836775	2	3,999999911	0,917874396
34	huerto	0,010746646	2	3,999999911	0,927536232
35	arreglar	0,010366603	1	1,999999955	0,93236715
36	viña	0,010366603	1	1,999999955	0,937198068
37	fatiga	0,010366603	1	1,999999955	0,942028986

38	jardinería	0,010366603	1	1,999999955	0,946859903
39	florero	0,010366603	1	1,999999955	0,951690821
40	fertilizar	0,007737064	2	3,999999911	0,961352657
41	césped	0,007463451	1	1,999999955	0,966183575
42	escavar	0,007463451	1	1,999999955	0,971014493
43	caballero	0,007463451	1	1,999999955	0,975845411
44	irrigar	0,005373323	1	1,999999955	0,980676329
45	extirpar	0,005373323	1	1,999999955	0,985507246
46	floristería	0,003868532	1	1,999999955	0,990338164
47	cerdo	0,002785155	1	1,999999955	0,995169082
48	verdura	0,002785155	1	1,999999955	1
49	vino	0,002785155	1	1,999999955	1

CI14, “los animales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	perro	0,903004229	49	98,00000191	0,067215364
2	gato	0,809534907	47	93,99999976	0,131687243
3	pájaro	0,546356559	41	81,99999928	0,187928669
4	caballo	0,544799984	40	80,00000119	0,242798354
5	vaca	0,378292769	29	57,99999833	0,282578875
6	león	0,322645724	33	66,00000262	0,327846365
7	pez	0,291956514	26	51,99999809	0,36351166
8	cerdo	0,285189211	28	56,00000024	0,401920439
9	conejo	0,274814904	30	60,00000238	0,443072702
10	ratón	0,249433324	26	51,99999809	0,478737997
11	jirafa	0,222832724	20	40,0000006	0,50617284
12	tigre	0,211527824	19	37,99999952	0,53223594
13	gallina	0,206976339	22	43,99999976	0,562414266
14	burro	0,173005566	20	40,0000006	0,589849108
15	pollo	0,161385313	19	37,99999952	0,615912209
16	tortuga	0,15969108	13	25,99999905	0,633744856
17	hámster	0,134284481	12	23,99999946	0,650205761
18	mono	0,132017732	20	40,0000006	0,677640604
19	elefante	0,128512338	12	23,99999946	0,694101509
20	mariposa	0,114002235	12	23,99999946	0,710562414
21	serpiente	0,112488836	15	30,00000119	0,731138546
22	mosquito	0,088295676	10	20,0000003	0,744855967
23	hormiga	0,086216152	7	14,00000006	0,754458162
24	cebra	0,083228625	9	18,00000072	0,766803841
25	cabra	0,067868851	11	21,99999988	0,781893004
26	oveja	0,059501048	8	15,99999964	0,792866941
27	cordero	0,053967878	6	11,99999973	0,801097394
28	paloma	0,052645512	5	10,00000015	0,807956104
29	tiburón	0,051833957	6	11,99999973	0,816186557
30	águila	0,050771013	5	10,00000015	0,823045267
31	zorro	0,050499231	5	10,00000015	0,829903978
32	toro	0,049334254	5	10,00000015	0,836762689
33	lobo	0,048646823	4	7,999999821	0,842249657
34	delfín	0,046756063	7	14,00000006	0,851851852
35	gallo	0,045859437	4	7,999999821	0,85733882
36	caracol	0,042023767	5	10,00000015	0,864197531
37	oso	0,03973313	7	14,00000006	0,873799726

38	rana	0,035344303	5	10,00000015	0,880658436
39	abeja	0,03317504	2	3,999999911	0,88340192
40	ballena	0,03200243	5	10,00000015	0,890260631
41	alce	0,028897626	2	3,999999911	0,893004115
42	araña	0,028449399	3	5,999999866	0,897119342
43	pato	0,026737288	2	3,999999911	0,899862826
44	nutria	0,024632445	3	5,999999866	0,903978052
45	armadillo	0,022673363	2	3,999999911	0,906721536
46	yegua	0,021677164	2	3,999999911	0,909465021
47	pavo	0,02	1	1,999999955	0,910836763
48	atún	0,019091146	3	5,999999866	0,914951989
49	sierpe	0,016511641	1	1,999999955	0,916323731
50	cocodrilo	0,015411764	2	3,999999911	0,919067215
51	pulpo	0,015045201	2	3,999999911	0,9218107
52	culebra	0,014814789	2	3,999999911	0,924554184
53	almeja	0,014491913	2	3,999999911	0,927297668
54	bisonte	0,013631714	1	1,999999955	0,92866941
55	avestruz	0,013631714	1	1,999999955	0,930041152
56	buey	0,01349241	3	5,999999866	0,934156379
57	muflón	0,012385986	1	1,999999955	0,935528121
58	ternero	0,012385986	1	1,999999955	0,936899863
59	sapo	0,011986933	2	3,999999911	0,939643347
60	camello	0,011296357	2	3,999999911	0,942386831
61	gaviota	0,011254097	1	1,999999955	0,943758573
62	canguro	0,011254097	1	1,999999955	0,945130316
63	chivo	0,011254097	1	1,999999955	0,946502058
64	puerco	0,011254097	1	1,999999955	0,9478738
65	mosca	0,011083152	2	3,999999911	0,950617284
66	cachorro	0,010225646	1	1,999999955	0,951989026
67	trucha	0,010225646	1	1,999999955	0,953360768
68	alondra	0,010225646	1	1,999999955	0,95473251
69	avispa	0,009675879	2	3,999999911	0,957475995
70	murciélago	0,009500833	2	3,999999911	0,960219479
71	anchoa	0,00929118	1	1,999999955	0,961591221
72	cabrón	0,008974832	2	3,999999911	0,964334705
73	insecto	0,008442109	1	1,999999955	0,965706447
74	gorila	0,008442109	1	1,999999955	0,967078189
75	ciervo	0,008442109	1	1,999999955	0,968449931
76	anguila	0,008442109	1	1,999999955	0,969821674
77	papagayo	0,007670632	1	1,999999955	0,971193416

78	hipopótamo	0,006969654	1	1,999999955	0,972565158
79	zoo	0,006969654	1	1,999999955	0,9739369
80	búfalo	0,006505391	2	3,999999911	0,976680384
81	salmón	0,006332736	1	1,999999955	0,978052126
82	letargo	0,006332736	1	1,999999955	0,979423868
83	verme	0,006332736	1	1,999999955	0,98079561
84	escarabajo	0,006332736	1	1,999999955	0,982167353
85	hicotea	0,005754021	1	1,999999955	0,983539095
86	reptil	0,005228193	1	1,999999955	0,984910837
87	babosa	0,005228193	1	1,999999955	0,986282579
88	bacalao	0,004750417	1	1,999999955	0,987654321
89	mirlo	0,004316302	1	1,999999955	0,989026063
90	rato	0,003563461	1	1,999999955	0,990397805
91	mejillón	0,003563461	1	1,999999955	0,991769547
92	buitre	0,003237816	1	1,999999955	0,993141289
93	búho	0,00294193	1	1,999999955	0,994513032
94	koala	0,002673083	1	1,999999955	0,995884774
95	tarántula	0,002673083	1	1,999999955	0,997256516
96	perezoso	0,002673083	1	1,999999955	0,998628258
97	caballito-de-mar	0,002428804	1	1,999999955	1

CI15, “juegos y distracciones”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	jugar	0,473836452	29	57,99999833	0,059917355
2	leer	0,326024234	21	41,99999869	0,103305785
3	baloncesto	0,311291516	21	41,99999869	0,146694215
4	fútbol	0,311100811	21	41,99999869	0,190082645
5	libro	0,296230674	21	41,99999869	0,233471074
6	tele(visión)	0,272362143	22	43,99999976	0,27892562
7	cartas	0,266988039	21	41,99999869	0,32231405
8	bailar	0,212835938	15	30,00000119	0,353305785
9	ordenador	0,195679307	15	30,00000119	0,384297521
10	móvil	0,194394991	13	25,99999905	0,411157025
11	correr	0,177313745	16	31,99999928	0,444214876
12	cine	0,157209069	12	23,99999946	0,469008264
13	tenis	0,151045844	11	21,99999988	0,491735537
14	nadar	0,145325914	13	25,99999905	0,518595041
15	videojuegos	0,133351162	10	20,0000003	0,539256198
16	escuchar-música	0,133220613	9	18,00000072	0,55785124
17	pelota	0,131815135	9	18,00000072	0,576446281
18	deporte	0,128030851	10	20,0000003	0,597107438
19	pasear	0,120822459	9	18,00000072	0,615702479
20	balón	0,113802955	10	20,0000003	0,636363636
21	redes-sociales	0,112856261	6	11,99999973	0,648760331
22	voleibol	0,102000244	8	15,99999964	0,665289256
23	natación	0,100342184	9	18,00000072	0,683884298
24	muñeca	0,087431885	7	14,00000006	0,698347107
25	<i>Play-Station</i>	0,084660284	5	10,00000015	0,708677686
26	disco(teca)	0,08409442	6	11,99999973	0,72107438
27	comer	0,075790167	7	14,00000006	0,73553719
28	juego-de-mesa	0,073251463	6	11,99999973	0,747933884
29	pelí(cula)	0,066198327	4	7,999999821	0,756198347
30	dibujar	0,062068444	5	10,00000015	0,766528926
31	ajedrez	0,052730907	5	10,00000015	0,776859504
32	ir-de-compras	0,049707759	6	11,99999973	0,789256198
33	cantar	0,04703629	5	10,00000015	0,799586777
34	pintar	0,046348095	4	7,999999821	0,80785124
35	tocar	0,039999999	2	3,999999911	0,811983471
36	damas	0,036139909	3	5,999999866	0,818181818
37	teatro	0,035757661	4	7,999999821	0,826446281

38	beber	0,03489418	3	5,999999866	0,832644628
39	piscina	0,034748875	4	7,999999821	0,840909091
40	distraer	0,03317504	2	3,999999911	0,845041322
41	periódico	0,031929422	4	7,999999821	0,853305785
42	amigo	0,031115472	3	5,999999866	0,859504132
43	divertir	0,030143354	2	3,999999911	0,863636364
44	serie	0,030143354	2	3,999999911	0,867768595
45	bar	0,029291181	2	3,999999911	0,871900826
46	atracción	0,028987389	3	5,999999866	0,878099174
47	salir-con- amigos	0,028215909	3	5,999999866	0,884297521
48	alcohol	0,026614418	2	3,999999911	0,888429752
49	escribir	0,025637412	3	5,999999866	0,894628099
50	instrumento- musical	0,025228376	2	3,999999911	0,898760331
51	esquiar	0,025228376	2	3,999999911	0,902892562
52	escondite	0,024678258	3	5,999999866	0,909090909
53	fotografía	0,022922892	2	3,999999911	0,91322314
54	paseo	0,02	1	1,999999955	0,915289256
55	<i>Monopoly</i>	0,02	1	1,999999955	0,917355372
56	surf	0,02	1	1,999999955	0,919421488
57	cocinar	0,018924728	2	3,999999911	0,923553719
58	gimnasio	0,018223751	2	3,999999911	0,92768595
59	internet	0,018172309	1	1,999999955	0,929752066
60	<i>Facebook</i>	0,018172309	1	1,999999955	0,931818182
61	aburrir	0,018172309	1	1,999999955	0,933884298
62	colorar	0,018172309	1	1,999999955	0,935950413
63	parque	0,016961811	2	3,999999911	0,940082645
64	chico	0,016511641	1	1,999999955	0,94214876
65	marioneta	0,016511641	1	1,999999955	0,944214876
66	beisbol	0,016511641	1	1,999999955	0,946280992
67	entrenarse	0,016112741	2	3,999999911	0,950413223
68	practicar	0,015002731	1	1,999999955	0,952479339
69	descansar	0,015002731	1	1,999999955	0,954545455
70	viajar	0,013927179	2	3,999999911	0,958677686
71	ciclismo	0,013631714	1	1,999999955	0,960743802
72	juguete	0,013631714	1	1,999999955	0,962809917
73	playa	0,012385986	1	1,999999955	0,964876033
74	competición	0,011254097	1	1,999999955	0,966942149
75	dado	0,011254097	1	1,999999955	0,969008264
76	consola	0,011254097	1	1,999999955	0,97107438

77	golf	0,010225646	1	1,999999955	0,973140496
78	aprender	0,00929118	1	1,999999955	0,975206612
79	navegar	0,00929118	1	1,999999955	0,977272727
80	guitarra	0,00929118	1	1,999999955	0,979338843
81	pádel	0,008442109	1	1,999999955	0,981404959
82	peluche	0,008442109	1	1,999999955	0,983471074
83	fiesta	0,007670632	1	1,999999955	0,98553719
84	bici(cleta)	0,006332736	1	1,999999955	0,987603306
85	chatear	0,006332736	1	1,999999955	0,989669421
86	visitar	0,006179745	2	3,999999911	0,993801653
87	tapear	0,005754021	1	1,999999955	0,995867769
88	volar	0,003563461	1	1,999999955	0,997933884
89	dormir	0,002428804	1	1,999999955	1

CI16, “profesiones y oficios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	profesor	0,709641576	42	83,99999738	0,066878981
2	abogado	0,491881609	30	60,00000238	0,114649682
3	camarero	0,449185342	38	75,99999905	0,175159236
4	doctor	0,416348994	29	57,99999833	0,22133758
5	cocinero	0,293964416	29	57,99999833	0,267515924
6	actor	0,281964391	24	47,99999893	0,305732484
7	peluquero	0,222062781	22	43,99999976	0,340764331
8	secretario	0,198214173	15	30,00000119	0,364649682
9	enfermero	0,195621982	13	25,99999905	0,385350318
10	médico	0,179588169	16	31,99999928	0,410828025
11	estudiante	0,148478016	13	25,99999905	0,431528662
12	policía	0,148127049	14	28,00000012	0,453821656
13	jardinero	0,145388886	12	23,99999946	0,472929936
14	juez	0,130046636	12	23,99999946	0,492038217
15	escritor	0,127062872	8	15,99999964	0,50477707
16	panadero	0,122084573	13	25,99999905	0,525477707
17	cantante	0,121199481	13	25,99999905	0,546178344
18	trabajar	0,120336145	11	21,99999988	0,563694268
19	bombero	0,115266412	8	15,99999964	0,576433121
20	bailarín	0,113824271	13	25,99999905	0,597133758
21	obrero	0,099162839	8	15,99999964	0,609872611
22	entrenador	0,098051094	8	15,99999964	0,622611465
23	empleado	0,097376049	8	15,99999964	0,635350318
24	director	0,09380886	8	15,99999964	0,648089172
25	fontanero	0,082016304	6	11,99999973	0,657643312
26	campesino	0,080714539	8	15,99999964	0,670382166
27	músico	0,076955706	12	23,99999946	0,689490446
28	pescador	0,076532155	8	15,99999964	0,702229299
29	pintor	0,076465681	8	15,99999964	0,714968153
30	jefe	0,068372279	5	10,00000015	0,722929936
31	ingeniero	0,061955404	5	10,00000015	0,73089172
32	albañil	0,057614364	4	7,999999821	0,737261146
33	actriz	0,056981489	6	11,99999973	0,746815287
34	arquitecto	0,056401234	4	7,999999821	0,753184713
35	barista	0,056311205	6	11,99999973	0,762738854
36	chef	0,056262288	4	7,999999821	0,76910828
37	empresario	0,056257438	5	10,00000015	0,777070064

38	futbolista	0,055663653	4	7,999999821	0,78343949
39	periodista	0,054891273	4	7,999999821	0,789808917
40	guía-de-turismo	0,051452156	5	10,00000015	0,797770701
41	vendedor	0,049534921	3	5,999999866	0,802547771
42	asistente-de-vuelo	0,049348917	4	7,999999821	0,808917197
43	azafata	0,04488581	3	5,999999866	0,813694268
44	fotógrafo	0,0433002	5	10,00000015	0,821656051
45	ama-de-casa	0,042673364	3	5,999999866	0,826433121
46	farmacéutico	0,042128459	4	7,999999821	0,832802548
47	maestro	0,04146539	3	5,999999866	0,837579618
48	carnicero	0,040330008	5	10,00000015	0,845541401
49	oficina	0,039999999	2	3,999999911	0,848726115
50	piloto	0,034435578	4	7,999999821	0,855095541
51	modelo	0,034165546	5	10,00000015	0,863057325
52	alcalde	0,030181944	3	5,999999866	0,867834395
53	cirujano	0,029733239	3	5,999999866	0,872611465
54	dentista	0,028942041	3	5,999999866	0,877388535
55	redactor	0,028897626	2	3,999999911	0,880573248
56	traductor	0,027839825	3	5,999999866	0,885350318
57	carpintero	0,022265663	2	3,999999911	0,888535032
58	agricultor	0,021479744	2	3,999999911	0,891719745
59	enseñar	0,02	1	1,999999955	0,893312102
60	granjero	0,019696208	2	3,999999911	0,896496815
61	criador	0,019696208	2	3,999999911	0,899681529
62	psiquiatra	0,019516828	2	3,999999911	0,902866242
63	portero	0,01935564	2	3,999999911	0,906050955
64	taller	0,018172309	1	1,999999955	0,907643312
65	educador	0,018172309	1	1,999999955	0,909235669
66	autor	0,017675813	2	3,999999911	0,912420382
67	astrónomo	0,016511641	1	1,999999955	0,914012739
68	autónomo	0,016511641	1	1,999999955	0,915605096
69	cartero	0,016511641	1	1,999999955	0,917197452
70	mediador	0,01648229	2	3,999999911	0,920382166
71	cazador	0,015623915	2	3,999999911	0,923566879
72	director-de-cine	0,015411764	2	3,999999911	0,926751592
73	defender	0,015002731	1	1,999999955	0,928343949
74	embajador	0,015002731	1	1,999999955	0,929936306
75	militar	0,015002731	1	1,999999955	0,931528662
76	alcaldesa	0,015002731	1	1,999999955	0,933121019

77	traducción	0,013631714	1	1,999999955	0,934713376
78	trabajador	0,013631714	1	1,999999955	0,936305732
79	poeta	0,012617809	3	5,999999866	0,941082803
80	esteticista	0,012385986	1	1,999999955	0,942675159
81	contador	0,012385986	1	1,999999955	0,944267516
82	cajero	0,012385986	1	1,999999955	0,945859873
83	tienda	0,011254097	1	1,999999955	0,947452229
84	hospital	0,010225646	1	1,999999955	0,949044586
85	florero	0,010225646	1	1,999999955	0,950636943
86	repcionista	0,010225646	1	1,999999955	0,952229299
87	político	0,010225646	1	1,999999955	0,953821656
88	químico	0,010225646	1	1,999999955	0,955414013
89	despacho	0,00929118	1	1,999999955	0,957006369
90	cocinar	0,00929118	1	1,999999955	0,958598726
91	farmacia	0,00929118	1	1,999999955	0,960191083
92	plomero	0,00929118	1	1,999999955	0,961783439
93	veterinario	0,00929118	1	1,999999955	0,963375796
94	instructor	0,00929118	1	1,999999955	0,964968153
95	cantor	0,00929118	1	1,999999955	0,96656051
96	deportista	0,00929118	1	1,999999955	0,968152866
97	pastelero	0,00929118	1	1,999999955	0,969745223
98	físico	0,00929118	1	1,999999955	0,97133758
99	horario	0,008442109	1	1,999999955	0,972929936
100	artista	0,007435041	2	3,999999911	0,97611465
101	estilista	0,007258232	2	3,999999911	0,979299363
102	psicólogo	0,006969654	1	1,999999955	0,98089172
103	entrevista-de-trabajo	0,006969654	1	1,999999955	0,982484076
104	soldado	0,006969654	1	1,999999955	0,984076433
105	investigador	0,006332736	1	1,999999955	0,98566879
106	maquinista	0,006332736	1	1,999999955	0,987261146
107	general	0,006332736	1	1,999999955	0,988853503
108	presidente	0,005754021	1	1,999999955	0,99044586
109	chofer	0,005444665	2	3,999999911	0,993630573
110	restaurador	0,005228193	1	1,999999955	0,99522293
111	cosmetólogo	0,005228193	1	1,999999955	0,996815287
112	mecánico	0,004750417	1	1,999999955	0,998407643
113	oficinista	0,004750417	1	1,999999955	1

ANEXO VI – LISTADO ALFABÉTICO DE LÉXICO DISPONIBLE

(CORPUS B2_ b)

CI01, “las partes del cuerpo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	barba	0,00328245	1	1,999999955	0,991082803
2	barriga	0,11000639	13	25,999999905	0,829299363
3	bíceps	0,0103666	1	1,999999955	0,973248408
4	boca	0,52944398	39	77,99999714	0,331210191
5	brazo	0,45347556	38	75,999999905	0,431847134
6	cabellos	0,06121696	5	10,00000015	0,910828025
7	cabeza	0,56961423	36	72,00000286	0,221656051
8	cadera	0,03436467	5	10,00000015	0,943949045
9	cara	0,36953753	25	50	0,554140127
10	cejas	0,0749658	7	14,00000006	0,863694268
11	cerebro	0,00217684	1	1,999999955	0,998726115
12	cintura	0,07182431	8	15,99999964	0,87388535
13	codos	0,0695781	10	20,0000003	0,886624204
14	colita	0,00633274	1	1,999999955	0,985987261
15	corazón	0,05311658	4	7,999999821	0,915923567
16	cuello	0,25987297	27	54,00000215	0,667515924
17	cuerpo	0,02101363	2	3,999999911	0,959235669
18	culo	0,00665369	2	3,999999911	0,984713376
19	dedo	0,46233949	41	81,99999928	0,38343949
20	dientes	0,1774254	21	41,99999869	0,78089172
21	espalda	0,29661363	32	63,99999857	0,633121019
22	estómago	0,03304558	4	7,999999821	0,951592357
23	frente	0,04371168	5	10,00000015	0,931210191
24	garganta	0,01934235	3	5,999999866	0,965605096
25	glúteo	0,00746345	1	1,999999955	0,98089172
26	hombro	0,12907742	12	23,99999946	0,796178344
27	labios	0,12809826	13	25,999999905	0,812738854
28	lengua	0,18896568	20	40,00000006	0,754140127
29	mano	0,70837861	48	95,99999785	0,061146497
30	médico	0,0023632	1	1,999999955	0,997452229
31	mejilla	0,05300115	7	14,00000006	0,924840764
32	mentón	0,00386853	1	1,999999955	0,989808917
33	muelas	0,00687489	1	1,999999955	0,982165605

34	muñeca	0,06933831	8	15,99999964	0,896815287
35	músculo	0,0095491	1	1,999999955	0,977070064
36	nalgas	0,01567094	2	3,999999911	0,968152866
37	nariz	0,43172881	36	72,00000286	0,477707006
38	ojos	0,6513288	47	93,99999976	0,121019108
39	olfato	0,0095491	1	1,999999955	0,975796178
40	ombligo	0,01977235	2	3,999999911	0,961783439
41	orejas	0,32147488	30	60,00000238	0,592356688
42	paladar	0,00278516	1	1,999999955	0,993630573
43	panza	0,00633274	1	1,999999955	0,987261146
44	pecho	0,09088355	12	23,99999946	0,844585987
45	pelo	0,41362104	35	69,99999881	0,522292994
46	pestañas	0,07515361	8	15,99999964	0,85477707
47	pie	0,55654699	47	93,99999976	0,281528662
48	piel	0,06525035	6	11,99999973	0,904458599
49	pierna	0,60147297	43	86,00000143	0,175796178
50	pulgar	0,00200518	1	1,999999955	1
51	pulmones	0,0081024	1	1,999999955	0,979617834
52	puño	0,02384488	2	3,999999911	0,954140127
53	pupila	0,0103666	1	1,999999955	0,974522293
54	rastro	0,00328245	1	1,999999955	0,992356688
55	riñón	0,00879606	1	1,999999955	0,978343949
56	rodilla	0,24885885	25	50	0,699363057
57	rostro	0,03405455	2	3,999999911	0,946496815
58	sangre	0,00278516	1	1,999999955	0,994904459
59	sobrecejos	0,01563174	1	1,999999955	0,969426752
60	talón	0,00419972	1	1,999999955	0,988535032
61	tobillo	0,01335532	2	3,999999911	0,971974522
62	trasero	0,0367991	5	10,00000015	0,937579618
63	tripa	0,02250144	2	3,999999911	0,956687898
64	uñas	0,21045408	23	46,00000083	0,72866242
65	ver	0,00256552	1	1,999999955	0,996178344

CI02, “la ropa”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abrigo	0,117233962	13	25,99999905	0,742424242
2	albornoz	0,002743868	1	1,999999955	0,993265993
3	anillo	0,031649657	5	10,00000015	0,929292929
4	bañador	0,096230097	11	21,99999988	0,801346801
5	bikini	0,004627867	1	1,999999955	0,986531987
6	blusa	0,115164272	10	20,00000003	0,782828283
7	bolso	0,16805914	16	31,99999928	0,638047138
8	botas	0,073819496	8	15,99999964	0,845117845
9	bragas	0,034903485	4	7,999999821	0,917508418
10	bufanda	0,212874815	22	43,99999976	0,523569024
11	calcetines	0,203128695	21	41,99999869	0,558922559
12	calzoncillos	0,019981118	2	3,999999911	0,947811448
13	camisa	0,441626191	29	57,99999833	0,314814815
14	camiseta	0,609961331	38	75,99999905	0,063973064
15	capucha	0,016226429	1	1,999999955	0,954545455
16	chaleco	0,02	1	1,999999955	0,944444444
17	chanclas	0,018491071	3	5,999999866	0,952861953
18	chándal	0,036094394	3	5,999999866	0,910774411
19	chaqueta	0,24462299	24	47,99999893	0,449494949
20	cinturón	0,160558224	12	23,99999946	0,658249158
21	colgante	0,002005177	1	1,999999955	0,998316498
22	collar	0,006332736	1	1,999999955	0,97979798
23	colorado	0,004627867	1	1,999999955	0,988215488
24	complemento	0,009620672	1	1,999999955	0,968013468
25	corbata	0,012433324	2	3,999999911	0,964646465
26	deportivo	0,00703064	1	1,999999955	0,976430976
27	elástico	0,003381974	1	1,999999955	0,98989899
28	elegante	0,007805458	1	1,999999955	0,974747475
29	estampado	0,008665665	1	1,999999955	0,96969697
30	estirar	0,005137885	1	1,999999955	0,983164983
31	falda	0,521908998	37	74,00000095	0,198653199
32	gafas	0,054199327	7	14,00000006	0,890572391
33	gafas-de-sol	0,014998401	2	3,999999911	0,957912458
34	gemelos	0,002005177	1	1,999999955	1
35	gorro	0,086104937	10	20,00000003	0,831649832
36	guantes	0,05616384	8	15,99999964	0,878787879
37	jersey	0,177477971	15	30,00000119	0,611111111

38	lavar	0,006332736	1	1,999999955	0,978114478
39	maleta	0,011858027	1	1,999999955	0,966329966
40	medias	0,002743868	1	1,999999955	0,991582492
41	minifalda	0,131340444	10	20,0000003	0,7003367
42	mochila	0,056511249	5	10,00000015	0,865319865
43	pajarita	0,002226159	1	1,999999955	0,996632997
44	pantalones	0,55659759	43	86,00000143	0,136363636
45	pantufla	0,008665665	1	1,999999955	0,971380471
46	pijama	0,091781095	8	15,99999964	0,814814815
47	pulsera	0,002471495	1	1,999999955	0,994949495
48	rebeca	0,021866739	3	5,999999866	0,937710438
49	reloj	0,021400414	3	5,999999866	0,942760943
50	ropa	0,038156971	3	5,999999866	0,905723906
51	ropa-interior	0,048420023	6	11,99999973	0,900673401
52	sandalias	0,013164851	1	1,999999955	0,95959596
53	sombrero	0,214348674	22	43,99999976	0,486531987
54	sostén	0,005704109	1	1,999999955	0,981481481
55	sudadera	0,437353641	30	60,00000238	0,365319865
56	suéter	0,032630373	2	3,999999911	0,920875421
57	sujetador	0,062678814	7	14,00000006	0,856902357
58	tacones	0,116492882	14	28,00000012	0,765993266
59	tenis	0,013164851	1	1,999999955	0,961279461
60	traje	0,149300545	15	30,00000119	0,683501684
61	traje-de-baño	0,007805458	1	1,999999955	0,973063973
62	vaqueros	0,375662982	26	51,99999809	0,409090909
63	vestido	0,188340113	16	31,99999928	0,585858586
64	vestirse	0,022785522	2	3,999999911	0,932659933
65	zapatilla-de-deporte	0,005137885	1	1,999999955	0,984848485
66	zapatillas	0,121659353	12	23,99999946	0,720538721
67	zapatos	0,514442086	40	80,00000119	0,265993266

CI03, “partes de la casa (sin muebles)”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	alcoba	0,011763047	1	1,999999955	0,959619952
2	ascensor	0,003409281	1	1,999999955	0,995249406
3	aseo	0,008257459	1	1,999999955	0,973871734
4	azotea	0,009855601	1	1,999999955	0,964370546
5	balcón	0,053062402	7	14,00000006	0,883610451
6	buhardilla	0,00771309	2	3,999999911	0,978622328
7	cama	0,02	1	1,999999955	0,950118765
8	cámara	0,065276816	4	7,999999821	0,857482185
9	cantina	0,007249901	2	3,999999911	0,983372922
10	casa	0,01403966	1	1,999999955	0,957244656
11	cocina	0,684232116	48	95,99999785	0,225653207
12	comedor	0,164211139	17	34,00000036	0,681710214
13	corredor	0,009855601	1	1,999999955	0,961995249
14	cuarto	0,170653328	11	21,99999988	0,60807601
15	cuarto-de-baño	0,684572935	47	93,99999976	0,111638955
16	cuarto-de-estar	0,037559643	3	5,999999866	0,93111639
17	cuarto-trastero	0,053882025	4	7,999999821	0,866983373
18	despacho	0,005796595	1	1,999999955	0,988123515
19	dormitorio	0,04389526	3	5,999999866	0,904988124
20	entrada	0,083903477	10	20,0000003	0,847980998
21	escalera	0,11866454	18	36,00000143	0,724465558
22	espacio	0,004856644	1	1,999999955	0,990498812
23	estantería	0,02	1	1,999999955	0,947743468
24	estudio	0,035409037	6	11,99999973	0,945368171
25	garaje	0,207323298	27	54,00000215	0,532066508
26	habitación	0,508510292	34	68,00000072	0,306413302
27	ingreso	0,015652196	2	3,999999911	0,954869359
28	jardín	0,362669349	41	81,99999928	0,403800475
29	muro	0,005796595	1	1,999999955	0,985748219
30	oficina	0,008257459	1	1,999999955	0,971496437
31	pared	0,05172161	6	11,99999973	0,897862233
32	parque	0,009855601	1	1,999999955	0,966745843
33	pasillo	0,170356348	14	28,00000012	0,641330166
34	patio	0,002005177	1	1,999999955	1
35	piso	0,102793455	10	20,0000003	0,748218527
36	planta	0,008257459	1	1,999999955	0,96912114
37	portal	0,002856446	1	1,999999955	0,997624703

38	puerta	0,096483856	10	20,0000003	0,771971496
39	recibidor	0,038392711	4	7,999999821	0,923990499
40	salón	0,292330235	27	54,00000215	0,467933492
41	sótano	0,043232173	4	7,999999821	0,914489311
42	suelo	0,004069111	1	1,999999955	0,992874109
43	techo	0,196412519	21	41,99999869	0,581947743
44	terraza	0,088384591	12	23,99999946	0,824228029
45	ventana	0,092840612	10	20,0000003	0,795724466

CI04, “los muebles de la casa”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	alfombra	0,07177645	9	18,00000072	0,833333333
2	almohada	0,002337462	1	1,999999955	0,997747748
3	armario	0,476490438	42	83,99999738	0,394144144
4	bañadora	0,00586537	1	1,999999955	0,990990991
5	baño	0,021645064	3	5,999999866	0,961711712
6	cajón	0,083406389	7	14,00000006	0,790540541
7	cama	0,59248817	42	83,99999738	0,29954955
8	chimenea	0,002005177	1	1,999999955	1
9	cocina	0,055156551	4	7,999999821	0,896396396
10	cómoda	0,00586537	1	1,999999955	0,988738739
11	congelador	0,004316302	1	1,999999955	0,995495495
12	cortina	0,014807723	2	3,999999911	0,970720721
13	cuadro	0,00929118	1	1,999999955	0,984234234
14	diván	0,037602469	4	7,999999821	0,927927928
15	ducha	0,074727118	10	20,0000003	0,813063063
16	escritorio	0,061806098	8	15,99999964	0,887387387
17	espejo	0,045970064	7	14,00000006	0,912162162
18	estantería	0,14716956	18	36,00000143	0,718468468
19	fregadero	0,027987728	2	3,999999911	0,954954955
20	frigorífico	0,069919817	8	15,99999964	0,851351351
21	grifo	0,007970381	1	1,999999955	0,986486486
22	horno	0,204054072	20	40,0000006	0,644144144
23	(horno)microondas	0,01432275	2	3,999999911	0,975225225
24	lámpara	0,163142428	15	30,00000119	0,677927928
25	lavabo	0,005031571	1	1,999999955	0,993243243
26	lavadora	0,130393282	14	28,00000012	0,75
27	lavavajillas	0,069059886	8	15,99999964	0,869369369
28	mesa	0,631066442	42	83,99999738	0,204954955
29	mesilla-de-noche	0,271197557	24	47,99999893	0,599099099
30	mueble	0,012625673	1	1,999999955	0,977477477
31	nevera	0,121058576	11	21,99999988	0,774774775
32	piano	0,010830854	1	1,999999955	0,97972973
33	puerta	0,032069284	4	7,999999821	0,941441441
34	refrigerador	0,034717917	2	3,999999911	0,932432432
35	ropero	0,00929118	1	1,999999955	0,981981982
36	secadora	0,017261561	2	3,999999911	0,966216216

37	silla	0,763007104	49	98,00000191	0,11036036
38	sillón	0,279084623	27	54,00000215	0,545045045
39	sofá	0,42817679	40	80,00000119	0,484234234
40	tele(visión)	0,030936243	4	7,999999821	0,95045045
41	ventana	0,038818583	3	5,999999866	0,918918919

CI05, “alimentos y bebidas”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aceite	0,072743438	8	15,99999964	0,716470588
2	aceituna	0,055749707	5	10,00000015	0,777647059
3	agua	0,635452032	41	81,99999928	0,048235294
4	agua-tónica	0,011515941	1	1,999999955	0,947058824
5	ajo	0,027629152	5	10,00000015	0,881176471
6	albahaca	0,005516437	1	1,999999955	0,978823529
7	albaricoque	0,008738445	1	1,999999955	0,96
8	apio	0,003482437	1	1,999999955	0,994117647
9	arroz	0,123259909	11	21,99999988	0,564705882
10	asado	0,011515941	1	1,999999955	0,945882353
11	atún	0,056594711	4	7,999999821	0,756470588
12	azúcar	0,056029309	8	15,99999964	0,771764706
13	banana	0,03483187	3	5,999999866	0,834117647
14	beber	0,012786262	2	3,999999911	0,941176471
15	berenjena	0,020254388	2	3,999999911	0,903529412
16	bocadillo	0,196689829	19	37,99999952	0,371764706
17	bollo	0,006630845	1	1,999999955	0,972941176
18	bombón	0,007970381	1	1,999999955	0,963529412
19	brécol	0,013842343	1	1,999999955	0,935294118
20	burrito	0,02	1	1,999999955	0,904705882
21	butifarra	0,015176258	1	1,999999955	0,927058824
22	café	0,085591353	13	25,99999905	0,676470588
23	café-con-leche	0,004185944	1	1,999999955	0,989411765
24	café-solo	0,003818021	1	1,999999955	0,991764706
25	calabacín	0,027727311	3	5,999999866	0,875294118
26	calabaza	0,007269825	1	1,999999955	0,965882353
27	camarón	0,002897164	1	1,999999955	0,997647059
28	cangrejo	0,002642518	1	1,999999955	0,998823529
29	carne	0,095831797	9	18,00000072	0,644705882
30	cebolla	0,075653031	12	23,99999946	0,707058824
31	cerdo	0,008738445	1	1,999999955	0,961176471
32	cereales	0,016708827	2	3,999999911	0,917647059
33	cereza	0,039653819	5	10,00000015	0,816470588
34	cerveza	0,350475401	27	54,00000215	0,118823529
35	champiñón	0,003818021	1	1,999999955	0,990588235
36	chirimoya	0,005516437	1	1,999999955	0,982352941
37	chocolate	0,031912126	4	7,999999821	0,848235294

38	chorizo	0,06541273	8	15,99999964	0,725882353
39	churros	0,056321803	5	10,00000015	0,762352941
40	ciruela	0,007970381	1	1,999999955	0,964705882
41	<i>Coca-Cola</i>	0,086081639	6	11,99999973	0,661176471
42	coco	0,007269825	1	1,999999955	0,967058824
43	cococha	0,005516437	1	1,999999955	0,981176471
44	cóctel	0,006630845	1	1,999999955	0,974117647
45	comer	0,006630845	1	1,999999955	0,971764706
46	comida	0,033842344	2	3,999999911	0,836470588
47	cordero	0,015535321	2	3,999999911	0,923529412
48	croqueta	0,006630845	1	1,999999955	0,970588235
49	dulce-de-leche	0,014786473	2	3,999999911	0,931764706
50	embutido	0,01050375	1	1,999999955	0,951764706
51	empanada	0,007970381	1	1,999999955	0,962352941
52	ensalada	0,048778988	6	11,99999973	0,784705882
53	espagueti	0,016638717	1	1,999999955	0,921176471
54	espárrago	0,002897164	1	1,999999955	0,996470588
55	fresa	0,15604502	17	34,00000036	0,496470588
56	frijoles	0,013486818	2	3,999999911	0,938823529
57	frito	0,006048028	1	1,999999955	0,975294118
58	fruta	0,047499754	4	7,999999821	0,8
59	galleta	0,087273262	8	15,99999964	0,654117647
60	gazpacho	0,117847465	12	23,99999946	0,595294118
61	hambre	0,005031571	1	1,999999955	0,985882353
62	hamburguesa	0,114162713	8	15,99999964	0,62
63	harina	0,036867961	6	11,99999973	0,827058824
64	hierbabuena	0,024589323	2	3,999999911	0,885882353
65	hortaliza	0,012625673	1	1,999999955	0,944705882
66	huevo	0,057096533	7	14,00000006	0,751764706
67	jalapeño	0,013842343	1	1,999999955	0,934117647
68	jamón	0,325120628	36	72,00000286	0,191764706
69	jamón-cocido	0,01536929	2	3,999999911	0,925882353
70	judías	0,018242102	1	1,999999955	0,910588235
71	jugo	0,018242102	1	1,999999955	0,911764706
72	kiwi	0,006048028	1	1,999999955	0,977647059
73	langosta	0,002410254	1	1,999999955	1
74	leche	0,24110207	22	43,99999976	0,304705882
75	lechuga	0,028647833	4	7,999999821	0,871764706
76	legumbres	0,012625673	1	1,999999955	0,942352941
77	limón	0,137682617	13	25,99999905	0,525882353

78	manteca	0,003818021	1	1,999999955	0,992941176
79	mantequilla	0,045268916	9	18,00000072	0,810588235
80	manzana	0,383473247	33	66,00000262	0,087058824
81	mariscos	0,047801897	6	11,99999973	0,791764706
82	melocotón	0,031867865	3	5,999999866	0,851764706
83	melón	0,031536765	4	7,999999821	0,856470588
84	mojito	0,136998296	9	18,00000072	0,536470588
85	naranja	0,345326662	26	51,99999809	0,149411765
86	paella	0,263280928	21	41,99999869	0,216470588
87	palomitas	0,02	1	1,999999955	0,905882353
88	pan	0,185082033	16	31,99999928	0,414117647
89	pasta	0,181169912	15	30,00000119	0,431764706
90	pastel	0,00986605	2	3,999999911	0,954117647
91	patata	0,116352327	13	25,99999905	0,610588235
92	pavo	0,011515941	1	1,999999955	0,948235294
93	pechuga-de-pollo	0,029664798	3	5,999999866	0,867058824
94	pepino	0,016638717	1	1,999999955	0,92
95	pepito	0,005516437	1	1,999999955	0,98
96	pera	0,033388533	4	7,999999821	0,841176471
97	perejil	0,005031571	1	1,999999955	0,984705882
98	perrito-caliente	0,01063735	2	3,999999911	0,950588235
99	pescado	0,257610768	26	51,99999809	0,247058824
100	pimentón	0,004589322	1	1,999999955	0,987058824
101	pimienta	0,00968008	2	3,999999911	0,956470588
102	pimiento	0,060442969	7	14,00000006	0,743529412
103	piña	0,024075644	3	5,999999866	0,892941176
104	piña-colada	0,012625673	1	1,999999955	0,943529412
105	pincho	0,006048028	1	1,999999955	0,976470588
106	plátano	0,189315781	20	40,00000006	0,395294118
107	pollastre	0,008738445	1	1,999999955	0,958823529
108	pollo	0,121179998	14	28,00000012	0,581176471
109	postre	0,006999576	2	3,999999911	0,969411765
110	queso	0,166866481	24	47,99999893	0,476470588
111	res	0,013842343	1	1,999999955	0,936470588
112	ron	0,076130331	7	14,00000006	0,692941176
113	sal	0,060935002	8	15,99999964	0,735294118
114	salado	0,016638717	1	1,999999955	0,918823529
115	salchicha	0,018834291	3	5,999999866	0,909411765
116	salchichón	0,005031571	1	1,999999955	0,983529412

117	salmón	0,022639114	3	5,999999866	0,896470588
118	sandía	0,017584488	3	5,999999866	0,915294118
119	sangría	0,238604978	21	41,99999869	0,329411765
120	sopa	0,004185944	1	1,999999955	0,988235294
121	taco	0,047506496	3	5,999999866	0,795294118
122	tapas	0,029777484	3	5,999999866	0,863529412
123	tarta	0,015096962	2	3,999999911	0,929411765
124	té	0,07781332	7	14,00000006	0,684705882
125	tequila	0,149518877	12	23,99999946	0,510588235
126	ternera	0,013842343	1	1,999999955	0,932941176
127	tinto-de-verano	0,003482437	1	1,999999955	0,995294118
128	tomate	0,252142251	27	54,00000215	0,278823529
129	tortilla	0,113463052	12	23,99999946	0,634117647
130	uva	0,024367057	3	5,999999866	0,889411765
131	vaca	0,009580524	1	1,999999955	0,957647059
132	verdura	0,035862036	3	5,999999866	0,830588235
133	vinagre	0,02054783	4	7,999999821	0,901176471
134	vino	0,199763492	17	34,00000036	0,349411765
135	vino-blanco	0,038132474	3	5,999999866	0,82
136	vino-tinto	0,031387627	3	5,999999866	0,86
137	zanahoria	0,132242933	13	25,99999905	0,551764706
138	zumو	0,181150109	14	28,00000012	0,448235294
139	zumو-de-fruta	0,025358286	2	3,999999911	0,883529412
140	zumو-de-naranja	0,033277433	2	3,999999911	0,843529412

CI06, “objetos colocados en la mesa para la comida”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aceite	0,093990147	11	21,99999988	0,835820896
2	agua	0,057554368	5	10,00000015	0,893034826
3	beber	0,015298693	2	3,999999911	0,965174129
4	bebida	0,03242179	3	5,999999866	0,925373134
5	botella	0,34520179	30	60,00000238	0,701492537
6	centro-de-mesa	0,008164369	2	3,999999911	0,97761194
7	comer	0,020468483	3	5,999999866	0,945273632
8	comida	0,005960808	1	1,999999955	0,982587065
9	copa	0,057141647	4	7,999999821	0,902985075
10	cubierto	0,017367754	3	5,999999866	0,960199005
11	cuchara	0,600198209	43	86,00000143	0,432835821
12	cucharilla	0,110618323	8	15,99999964	0,808457711
13	cuchillo	0,76044482	46	92,00000167	0,114427861
14	ensaladera	0,033419218	2	3,999999911	0,917910448
15	fruta	0,003672951	1	1,999999955	0,997512438
16	jarra	0,042891916	4	7,999999821	0,912935323
17	mantel	0,477595717	39	77,99999714	0,626865672
18	navaja	0,017719747	1	1,999999955	0,952736318
19	pan	0,029163193	3	5,999999866	0,932835821
20	papel	0,018244579	2	3,999999911	0,950248756
21	pasta	0,009673755	1	1,999999955	0,967661692
22	pimienta	0,092433512	10	20,00000003	0,860696517
23	plato	0,696026623	44	87,99999952	0,223880597
24	plato-hondo	0,026290569	2	3,999999911	0,937810945
25	sal	0,112899065	11	21,99999988	0,788557214
26	sentarse	0,004679076	1	1,999999955	0,990049751
27	servilleta	0,247670174	24	47,99999893	0,76119403
28	tenedor	0,565229416	39	77,99999714	0,529850746
29	utensilio	0,008570824	1	1,999999955	0,970149254
30	vajilla	0,007593641	1	1,999999955	0,980099502
31	vaso	0,620776772	41	81,99999928	0,325870647
32	vinagre	0,061547745	8	15,99999964	0,880597015
33	vino	0,008570824	1	1,999999955	0,972636816
34	vino-seco	0,0052812	1	1,999999955	0,987562189
35	vino-tinto	0,005960808	1	1,999999955	0,985074627

CI07, “la cocina y sus utensilios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abrebotellas	0,011930292	3	5,999999866	0,927083333
2	abrelatas	0,005051436	2	3,999999911	0,989583333
3	afilador	0,018868959	2	3,999999911	0,875
4	agua	0,003754687	1	1,999999955	1
5	balanza	0,01733133	2	3,999999911	0,881944444
6	batidor	0,068684436	7	14,00000006	0,663194444
7	batidora	0,013164851	1	1,999999955	0,909722222
8	botella	0,008665665	1	1,999999955	0,951388889
9	cacerola	0,095629901	8	15,99999964	0,600694444
10	cafetera	0,02	1	1,999999955	0,861111111
11	calentar	0,005704109	1	1,999999955	0,975694444
12	cazo	0,038141645	4	7,999999821	0,75
13	cazuela	0,037048217	4	7,999999821	0,774305556
14	cocinar	0,029757585	4	7,999999821	0,805555556
15	colador	0,099093094	9	18,00000072	0,572916667
16	cortar	0,010331975	2	3,999999911	0,944444444
17	cubierto	0,054526702	4	7,999999821	0,71875
18	cuchara	0,224970534	18	36,00000143	0,381944444
19	cucharilla	0,015308793	2	3,999999911	0,899305556
20	cucharón	0,048147786	5	10,00000015	0,736111111
21	cuchillo	0,205783516	16	31,99999928	0,4375
22	cuenca	0,068861157	5	10,00000015	0,638888889
23	embudo	0,016226429	1	1,999999955	0,885416667
24	escurridor	0,003754687	1	1,999999955	0,996527778
25	espátula	0,00703064	1	1,999999955	0,96875
26	esponja	0,028012257	3	5,999999866	0,826388889
27	espumadera	0,00703064	1	1,999999955	0,965277778
28	freír	0,008665665	1	1,999999955	0,958333333
29	frigorífico	0,037080728	3	5,999999866	0,760416667
30	fuego	0,028861156	3	5,999999866	0,815972222
31	grifo	0,010680926	1	1,999999955	0,9375
32	hervidor	0,010680926	1	1,999999955	0,930555556
33	horno	0,445441127	27	54,00000215	0,09375
34	(horno)microondas	0,110641956	10	20,0000003	0,541666667
35	lavavajillas	0,065022886	6	11,99999973	0,704861111
36	licuadora	0,019346589	2	3,999999911	0,868055556
37	mantel	0,016226429	1	1,999999955	0,892361111

38	mesa	0,027170362	4	7,999999821	0,840277778
39	mezclador	0,013164851	1	1,999999955	0,913194444
40	mezclar	0,005704109	1	1,999999955	0,979166667
41	molde	0,008665665	1	1,999999955	0,947916667
42	molinillo	0,004627867	1	1,999999955	0,993055556
43	navaja	0,016226429	1	1,999999955	0,888888889
44	nevera	0,142611116	11	21,99999988	0,475694444
45	olla	0,270124078	18	36,00000143	0,256944444
46	olla-a-presión	0,020854296	2	3,999999911	0,857638889
47	paella	0,065068699	6	11,99999973	0,684027778
48	pala-de-freír	0,008665665	1	1,999999955	0,954861111
49	pelador	0,010680926	1	1,999999955	0,934027778
50	plato	0,23930271	18	36,00000143	0,319444444
51	rallador	0,033937994	3	5,999999866	0,784722222
52	refrigerador	0,033164848	2	3,999999911	0,791666667
53	rodillo	0,014435614	2	3,999999911	0,90625
54	sacacorchos	0,021400414	3	5,999999866	0,850694444
55	sartén	0,400939256	29	57,99999833	0,194444444
56	servilleta	0,005704109	1	1,999999955	0,972222222
57	silla	0,005704109	1	1,999999955	0,982638889
58	taza	0,00703064	1	1,999999955	0,961805556
59	tenedor	0,126008689	9	18,00000072	0,506944444
60	trapo	0,013164851	1	1,999999955	0,916666667
61	vaso	0,074722193	6	11,99999973	0,621527778

CI08, “la escuela: muebles y materiales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	actividad	0,007463451	1	1,999999955	0,959361393
2	alumno	0,083195001	14	28,00000012	0,712626996
3	aprender	0,037623931	6	11,99999973	0,843251089
4	aprobar	0,008327321	1	1,999999955	0,95065312
5	aprobar-examen	0,006689199	1	1,999999955	0,968069666
6	apuntes	0,068062894	7	14,00000006	0,748911466
7	armario	0,061703347	4	7,999999821	0,779390421
8	asignatura	0,018875049	4	7,999999821	0,899854862
9	aula	0,021933107	2	3,999999911	0,894049347
10	aula-magna	0,012905287	1	1,999999955	0,931785196
11	bachillerato	0,006689199	1	1,999999955	0,965166909
12	bolí(grafo)	0,624877334	42	83,99999738	0,06095791
13	bolso	0,013474131	2	3,999999911	0,912917271
14	borrador	0,188381001	17	34,00000036	0,586357039
15	borrar	0,004815899	1	1,999999955	0,982583454
16	calculador	0,0110055	2	3,999999911	0,940493469
17	carpeta	0,004316302	1	1,999999955	0,986937591
18	cátedra	0,017925216	1	1,999999955	0,90275762
19	ciencia	0,006689199	1	1,999999955	0,966618287
20	clase	0,074912056	9	18,00000072	0,738751814
21	colegio	0,007463451	1	1,999999955	0,955007257
22	color	0,04371915	5	10,00000015	0,809869376
23	computador	0,00929118	1	1,999999955	0,949201742
24	cuaderno	0,40448308	33	66,00000262	0,323657475
25	curso	0,013700644	2	3,999999911	0,910014514
26	deberes	0,007335746	2	3,999999911	0,96371553
27	despacho	0,023637723	5	10,00000015	0,891146589
28	diccionario	0,026486574	4	7,999999821	0,87808418
29	director	0,004815899	1	1,999999955	0,978229318
30	ejercicio	0,029524386	3	5,999999866	0,856313498
31	enseñanza	0,00929118	1	1,999999955	0,946298984
32	enseñar	0,066228643	7	14,00000006	0,759071118
33	entregar	0,003868532	1	1,999999955	0,98838897
34	escribir	0,12941815	17	34,00000036	0,637155298
35	escuchar	0,028645176	4	7,999999821	0,865021771
36	escuela	0,00929118	1	1,999999955	0,947750363
37	estuche	0,379247427	28	56,00000024	0,400580552

38	estudiante	0,133570641	18	36,00000143	0,612481858
39	estudiar	0,226635858	28	56,00000024	0,531204644
40	examen	0,062445763	10	20,00000003	0,773584906
41	goma	0,023949379	4	7,999999821	0,883889695
42	goma-de-borrar	0,052397802	4	7,999999821	0,79245283
43	historia	0,01501652	2	3,999999911	0,905660377
44	hoja	0,005995267	1	1,999999955	0,972423803
45	informe	0,002785156	1	1,999999955	0,994194485
46	lápiz	0,499719381	37	74,00000095	0,223512337
47	leer	0,123478919	17	34,00000036	0,661828737
48	libro	0,481130153	36	72,00000286	0,275761974
49	literatura	0,003107528	1	1,999999955	0,992743106
50	luz	0,003868532	1	1,999999955	0,989840348
51	maestro	0,007463451	1	1,999999955	0,960812772
52	manual	0,017925216	1	1,999999955	0,901306241
53	mapa	0,005373323	1	1,999999955	0,97532656
54	marcador	0,0495185	7	14,00000006	0,802612482
55	matemática	0,013458719	2	3,999999911	0,915820029
56	mesa	0,388972521	25	50	0,359941945
57	mochila	0,323707402	26	51,99999809	0,438316401
58	nota	0,013458719	2	3,999999911	0,918722787
59	oficina	0,01439903	1	1,999999955	0,907111756
60	opinar	0,008327321	1	1,999999955	0,952104499
61	opinión	0,002237269	1	1,999999955	0,997097242
62	ordenador	0,209623486	21	41,99999869	0,561683599
63	ordenar	0,007463451	1	1,999999955	0,956458636
64	página	0,008327321	1	1,999999955	0,953555878
65	pantalla	0,004316302	1	1,999999955	0,985486212
66	papel	0,118069962	9	18,00000072	0,674891147
67	papelera	0,029491721	2	3,999999911	0,859216255
68	pastel	0,011566503	1	1,999999955	0,937590711
69	pensar	0,00929118	1	1,999999955	0,944847605
70	perchero	0,013099012	3	5,999999866	0,928882438
71	pintar	0,004815899	1	1,999999955	0,979680697
72	pintura	0,012905287	1	1,999999955	0,930333817
73	pizarra	0,560122848	39	77,99999714	0,169811321
74	pluma	0,007463451	1	1,999999955	0,957910015
75	prestar-atención	0,005995267	1	1,999999955	0,970972424
76	profe(sor)	0,278022587	36	72,00000286	0,490566038
77	puerta	0,013159713	2	3,999999911	0,921625544

78	pupitre	0,05820176	5	10,00000015	0,786647315
79	química	0,002237269	1	1,999999955	0,998548621
80	registro	0,027304316	2	3,999999911	0,867924528
81	regla	0,037804585	5	10,00000015	0,834542816
82	repasar	0,005373323	1	1,999999955	0,973875181
83	repetir	0,006689199	1	1,999999955	0,969521045
84	reseña	0,003467214	1	1,999999955	0,991291727
85	resumen	0,002496226	1	1,999999955	0,995645864
86	revisar	0,004815899	1	1,999999955	0,976777939
87	rotulador	0,078340262	9	18,00000072	0,725689405
88	sacapuntas	0,109751455	12	23,99999946	0,692307692
89	secretaría	0,012307751	3	5,999999866	0,936139332
90	silla	0,560440719	36	72,00000286	0,113207547
91	subrayar	0,010811167	2	3,999999911	0,943396226
92	tabla-periódica	0,004815899	1	1,999999955	0,981132075
93	tableta	0,03977184	6	11,99999973	0,827285922
94	tijeras	0,031283285	6	11,99999973	0,851959361
95	tiza	0,043310452	6	11,99999973	0,818577649
96	tomar-apuntes	0,013151759	2	3,999999911	0,924528302
97	tutoría	0,004316302	1	1,999999955	0,984034833
98	valoración	0,002005177	1	1,999999955	1
99	ventana	0,027042652	3	5,999999866	0,872278665

CI09, “iluminación y calefacción”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abanico	0,016756885	1	1,999999955	0,914201183
2	acondicionar	0,006918464	1	1,999999955	0,982248521
3	aire	0,004856644	1	1,999999955	0,99704142
4	aire-acondicionado	0,089713357	7	14,00000006	0,671597633
5	apagar	0,091260664	8	15,99999964	0,650887574
6	araña	0,02807932	2	3,999999911	0,837278107
7	bombilla	0,254878402	20	40,0000006	0,363905325
8	calefacción	0,124975942	10	20,0000003	0,597633136
9	calefactor	0,02	1	1,999999955	0,869822485
10	calentador	0,02	1	1,999999955	0,872781065
11	calentar	0,085136369	7	14,00000006	0,692307692
12	caliente	0,015832158	2	3,999999911	0,920118343
13	calor	0,243989661	24	47,99999893	0,434911243
14	calorífico	0,02	1	1,999999955	0,875739645
15	candela	0,017559642	2	3,999999911	0,908284024
16	candelabro	0,011763047	1	1,999999955	0,946745562
17	chimenea	0,057646122	5	10,00000015	0,816568047
18	claro	0,01166674	2	3,999999911	0,952662722
19	electricidad	0,053551454	5	10,00000015	0,831360947
20	electricista	0,018681511	2	3,999999911	0,890532544
21	eléctrico	0,005796595	1	1,999999955	0,99112426
22	encender	0,114825211	10	20,0000003	0,627218935
23	energía-eléctrica	0,014062519	3	5,999999866	0,934911243
24	estrella	0,017559642	2	3,999999911	0,902366864
25	estufa	0,01811306	2	3,999999911	0,896449704
26	farola	0,074488804	6	11,99999973	0,73964497
27	frío	0,058550335	7	14,00000006	0,801775148
28	fuego	0,330727935	24	47,99999893	0,304733728
29	hilo-eléctrico	0,01403966	1	1,999999955	0,940828402
30	hogar	0,016756885	1	1,999999955	0,911242604
31	horno	0,071512699	6	11,99999973	0,75739645
32	(horno)microondas	0,006918464	1	1,999999955	0,976331361
33	iluminado	0,018910699	3	5,999999866	0,884615385
34	iluminar	0,134498641	12	23,99999946	0,568047337
35	insolación	0,008257459	1	1,999999955	0,967455621

36	lámina	0,009855601	1	1,999999955	0,964497041
37	lámpara	0,804322839	47	93,99999976	0,139053254
38	led	0,023526095	2	3,999999911	0,860946746
39	linterna	0,020166166	2	3,999999911	0,866863905
40	luminoso	0,004069111	1	1,999999955	1
41	luz	0,515237451	32	63,99999857	0,233727811
42	madera	0,025914831	4	7,999999821	0,849112426
43	mechero	0,082267769	10	20,0000003	0,721893491
44	neón	0,023895262	2	3,999999911	0,855029586
45	nevera	0,005796595	1	1,999999955	0,98816568
46	oscuro	0,006918464	1	1,999999955	0,979289941
47	panel-solar	0,008257459	1	1,999999955	0,970414201
48	radiador	0,167096436	15	30,00000119	0,532544379
49	refrescador	0,01403966	1	1,999999955	0,937869822
50	refriar	0,011763047	1	1,999999955	0,943786982
51	sol	0,208063886	18	36,00000143	0,48816568
52	sombra	0,010987575	2	3,999999911	0,958579882
53	termo(sifón)	0,006918464	1	1,999999955	0,985207101
54	ultravioleta	0,009855601	1	1,999999955	0,961538462
55	vela	0,061925936	8	15,99999964	0,781065089
56	ventana	0,006918464	1	1,999999955	0,973372781
57	ventilador	0,015652196	2	3,999999911	0,926035503
58	verano	0,004856644	1	1,999999955	0,99408284

CI10, “la ciudad”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	acera	0,008131393	1	1,999999955	0,964788732
2	aeropuerto	0,054216936	7	14,00000006	0,754929577
3	albergue	0,008131393	1	1,999999955	0,961971831
4	alcalde	0,018096749	1	1,999999955	0,908450704
5	alcaldía	0,012130613	1	1,999999955	0,93943662
6	andar	0,02084025	2	3,999999911	0,892957746
7	antiguo	0,004931939	1	1,999999955	0,988732394
8	aparcar	0,00365367	1	1,999999955	0,998591549
9	apartamento	0,023338107	2	3,999999911	0,884507042
10	árbol	0,070514552	6	11,99999973	0,670422535
11	(auto)bús	0,253651202	27	54,00000215	0,333802817
12	autovía	0,027350849	2	3,999999911	0,856338028
13	avenida	0,115746886	8	15,99999964	0,583098592
14	ayuntamiento	0,128914341	13	25,99999905	0,571830986
15	banco	0,041409049	5	10,00000015	0,815492958
16	bar	0,081786118	12	23,99999946	0,661971831
17	barco	0,004931939	1	1,999999955	0,985915493
18	barrio	0,112609319	7	14,00000006	0,621126761
19	biblioteca	0,04648364	5	10,00000015	0,801408451
20	bici(cleta)	0,1130739	9	18,00000072	0,611267606
21	café	0,011820192	2	3,999999911	0,943661972
22	calle	0,88570106	50	100	0,070422535
23	caminar	0,038327079	3	5,999999866	0,835211268
24	capital	0,005450636	1	1,999999955	0,983098592
25	carretera	0,195171684	12	23,99999946	0,416901408
26	carro	0,00403793	1	1,999999955	0,992957746
27	casa	0,205742136	19	37,99999952	0,4
28	casco-antiguo	0,009931706	1	1,999999955	0,956338028
29	castro	0,012130613	1	1,999999955	0,936619718
30	catedral	0,008986579	1	1,999999955	0,957746479
31	cementerio	0,00365367	1	1,999999955	0,997183099
32	centro	0,029632729	2	3,999999911	0,846478873
33	centro-comercial	0,010976233	1	1,999999955	0,945070423
34	centro-histórico	0,006657422	1	1,999999955	0,971830986
35	cine	0,115698859	11	21,99999988	0,598591549
36	coche	0,301854372	25	50	0,184507042
37	colegio	0,010976233	1	1,999999955	0,946478873

38	comer	0,00403793	1	1,999999955	0,995774648
39	comisaría	0,00403793	1	1,999999955	0,994366197
40	comprar	0,012130613	1	1,999999955	0,935211268
41	conocer	0,004931939	1	1,999999955	0,987323944
42	correr	0,010901271	2	3,999999911	0,952112676
43	cruce	0,057482496	7	14,00000006	0,733802817
44	cruzar	0,050927822	5	10,00000015	0,777464789
45	dar-una-vuelta	0,010976233	1	1,999999955	0,947887324
46	derecha	0,018918285	2	3,999999911	0,901408451
47	despacho	0,018470034	2	3,999999911	0,907042254
48	disco(teca)	0,042612225	5	10,00000015	0,808450704
49	edificio	0,098870285	9	18,00000072	0,633802817
50	escuela	0,267535418	22	43,99999976	0,295774648
51	esquina	0,023106847	2	3,999999911	0,890140845
52	establecimiento	0,015488982	2	3,999999911	0,916901408
53	estación	0,245425075	28	56,00000024	0,373239437
54	estadio	0,010976233	1	1,999999955	0,949295775
55	estanco	0,013406401	1	1,999999955	0,923943662
56	estatua	0,017060071	2	3,999999911	0,914084507
57	estudiar	0,006657422	1	1,999999955	0,970422535
58	explorar	0,004462603	1	1,999999955	0,991549296
59	farmacia	0,056668267	8	15,99999964	0,745070423
60	gimnasio	0,006023884	1	1,999999955	0,978873239
61	habitación	0,023106847	2	3,999999911	0,887323944
62	habitante	0,008986579	1	1,999999955	0,95915493
63	hospital	0,159847409	18	36,00000143	0,484507042
64	iglesia	0,382334292	27	54,00000215	0,149295775
65	ir-de-compras	0,009931706	1	1,999999955	0,954929577
66	izquierda	0,018918285	2	3,999999911	0,904225352
67	jardín	0,085557312	8	15,99999964	0,645070423
68	letrero	0,007357589	1	1,999999955	0,966197183
69	librería	0,006657422	1	1,999999955	0,969014085
70	manzana	0,014816364	1	1,999999955	0,922535211
71	mercado	0,19052574	14	28,00000012	0,436619718
72	metro	0,052684359	6	11,99999973	0,763380282
73	monumento	0,142489821	11	21,99999988	0,553521127
74	moto(cicleta)	0,012130613	1	1,999999955	0,94084507
75	museo	0,149430767	19	37,99999952	0,511267606
76	negocio	0,041311264	4	7,999999821	0,821126761
77	oficio	0,018096749	1	1,999999955	0,909859155

78	país	0,02	1	1,999999955	0,897183099
79	palacio	0,060984746	6	11,99999973	0,723943662
80	panadería	0,020262009	2	3,999999911	0,895774648
81	parada	0,062388927	10	20,0000003	0,715492958
82	parque	0,290621251	28	56,00000024	0,264788732
83	parquímetro	0,005450636	1	1,999999955	0,98028169
84	paseante	0,012130613	1	1,999999955	0,933802817
85	pasear	0,049926862	6	11,99999973	0,785915493
86	paseo	0,012681305	2	3,999999911	0,928169014
87	paso-de-cebra	0,066771328	5	10,00000015	0,701408451
88	pastelería	0,013406401	1	1,999999955	0,925352113
89	peatón	0,024754172	2	3,999999911	0,881690141
90	peluquería	0,007357589	1	1,999999955	0,967605634
91	piso	0,026593886	5	10,00000015	0,870422535
92	planta	0,006253994	2	3,999999911	0,974647887
93	playa	0,006023884	1	1,999999955	0,977464789
94	plaza	0,461875796	29	57,99999833	0,111267606
95	policía	0,040155392	7	14,00000006	0,830985915
96	pueblo	0,018096749	1	1,999999955	0,911267606
97	puerto	0,005450636	1	1,999999955	0,984507042
98	quiosco	0,012681305	2	3,999999911	0,930985915
99	rascacielos	0,066918164	8	15,99999964	0,694366197
100	redonda	0,012130613	1	1,999999955	0,932394366
101	restaurante	0,149250537	19	37,99999952	0,538028169
102	rincón	0,029072981	2	3,999999911	0,853521127
103	río	0,004462603	1	1,999999955	0,990140845
104	semáforo	0,160329968	16	31,99999928	0,45915493
105	señal-de-tráfico	0,02538435	3	5,999999866	0,878873239
106	supermercado	0,052303996	5	10,00000015	0,770422535
107	taxi	0,035862818	3	5,999999866	0,843661972
108	teatro	0,014845178	3	5,999999866	0,921126761
109	tienda	0,299393058	29	57,99999833	0,225352113
110	trabajar	0,029310055	3	5,999999866	0,850704225
111	trabajo	0,009931706	1	1,999999955	0,953521127
112	tráfico	0,026946979	2	3,999999911	0,863380282
113	tren	0,047735348	6	11,99999973	0,794366197
114	tribunal	0,002449129	1	1,999999955	1
115	turista	0,012130613	1	1,999999955	0,938028169
116	universidad	0,070020616	9	18,00000072	0,683098592
117	urbano	0,02	1	1,999999955	0,898591549

118	vía	0,037584949	3	5,999999866	0,83943662
119	viajar	0,008131393	1	1,999999955	0,963380282
120	vida-nocturna	0,006023884	1	1,999999955	0,976056338
121	villa	0,005450636	1	1,999999955	0,981690141
122	visitar	0,02643091	3	5,999999866	0,874647887
123	vivir	0,027239017	3	5,999999866	0,86056338
124	zona-peatonal	0,008986579	1	1,999999955	0,96056338

CI11, “el campo”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	a-pie	0,004847364	1	2,040816285	0,97761194
2	abeja	0,007460967	1	2,040816285	0,952736318
3	agricultor	0,087898403	7	14,28571492	0,594527363
4	agricultura	0,013259112	1	2,040816285	0,902985075
5	agua	0,039910052	4	8,163265139	0,748756219
6	aire	0,013259112	1	2,040816285	0,900497512
7	al-aire-libre	0,022769878	2	4,081632569	0,828358209
8	animal	0,384502292	27	55,10203838	0,213930348
9	arar	0,006461975	1	2,040816285	0,957711443
10	árbol	0,560603023	34	69,38775778	0,084577114
11	ave	0,003636185	1	2,040816285	0,995024876
12	azada	0,005596744	1	2,040816285	0,967661692
13	bici(cleta)	0,005596744	1	2,040816285	0,970149254
14	bosque	0,020408163	1	2,040816285	0,843283582
15	burro	0,04298909	4	8,163265139	0,731343284
16	caballo	0,102199025	12	24,48979616	0,554726368
17	cabra	0,015308909	1	2,040816285	0,883084577
18	campesino	0,179148316	11	22,44897932	0,44278607
19	campo	0,208890676	12	24,48979616	0,330845771
20	canal	0,004847364	1	2,040816285	0,980099502
21	cantina	0,004847364	1	2,040816285	0,975124378
22	casa	0,067686468	4	8,163265139	0,639303483
23	cerdo	0,039483272	5	10,20408198	0,76119403
24	cereales	0,009946144	1	2,040816285	0,927860697
25	cereza	0,007460967	1	2,040816285	0,950248756
26	chivo	0,004847364	1	2,040816285	0,982587065
27	conejo	0,016408119	2	4,081632569	0,873134328
28	cortar	0,033667274	2	4,081632569	0,776119403
29	cortijo	0,189752907	17	34,69387889	0,373134328
30	cultivación	0,026289994	2	4,081632569	0,803482587
31	cultivador	0,022967545	2	4,081632569	0,823383085
32	cultivar	0,125214204	9	18,36734712	0,502487562
33	cultivo	0,007460967	1	2,040816285	0,945273632
34	espantapájaros	0,013922942	2	4,081632569	0,888059701
35	finca	0,011483773	1	2,040816285	0,92039801
36	flor	0,293008149	17	34,69387889	0,256218905
37	fruta	0,074415453	7	14,28571492	0,629353234

38	fruto	0,008614398	1	2,040816285	0,94278607
39	galera	0,009946144	1	2,040816285	0,932835821
40	gallina	0,077377662	7	14,28571492	0,611940299
41	gallo	0,022769878	2	4,081632569	0,833333333
42	ganso	0,008614398	1	2,040816285	0,940298507
43	gato	0,013259112	1	2,040816285	0,895522388
44	granja	0,012812721	2	4,081632569	0,912935323
45	granjero	0,040300451	3	6,12244904	0,73880597
46	grano	0,020156274	2	4,081632569	0,848258706
47	hacienda	0,020408163	1	2,040816285	0,84079602
48	hierba	0,231443837	18	36,73469424	0,300995025
49	hoja	0,006461975	1	2,040816285	0,960199005
50	hormiga	0,002362416	1	2,040816285	0,997512438
51	huerto	0,008614398	1	2,040816285	0,937810945
52	insecto	0,044471521	5	10,20408198	0,721393035
53	instrumento	0,007460967	1	2,040816285	0,955223881
54	irrigación	0,009946144	1	2,040816285	0,930348259
55	jardín	0,055759355	3	6,12244904	0,68159204
56	maíz	0,066700958	7	14,28571492	0,656716418
57	mariposa	0,017847592	3	6,12244904	0,860696517
58	mosca	0,002046099	1	2,040816285	1
59	mosquito	0,011483773	1	2,040816285	0,915422886
60	naturaleza	0,046243619	3	6,12244904	0,699004975
61	oveja	0,032762613	4	8,163265139	0,793532338
62	paisaje	0,006461975	1	2,040816285	0,962686567
63	pájaro	0,005596744	1	2,040816285	0,972636816
64	parque	0,017675597	1	2,040816285	0,86318408
65	pasear	0,032913689	3	6,12244904	0,78358209
66	pato	0,009946144	1	2,040816285	0,925373134
67	pavo	0,004198323	1	2,040816285	0,992537313
68	perro	0,095713869	9	18,36734712	0,577114428
69	planta	0,044841055	4	8,163265139	0,708955224
70	plantar	0,007460967	1	2,040816285	0,947761194
71	pollo	0,047423877	4	8,163265139	0,691542289
72	pozo	0,013259112	1	2,040816285	0,905472637
73	pradería	0,013259112	1	2,040816285	0,89800995
74	puerco	0,013922942	2	4,081632569	0,893034826
75	rastro	0,02009817	2	4,081632569	0,853233831
76	ratón	0,015542889	2	4,081632569	0,878109453
77	red	0,008614398	1	2,040816285	0,935323383

78	res	0,009946144	1	2,040816285	0,922885572
79	río	0,11026711	9	18,36734712	0,524875622
80	rural	0,004198323	1	2,040816285	0,990049751
81	sendero	0,030617818	2	4,081632569	0,798507463
82	soja	0,004198323	1	2,040816285	0,987562189
83	tierra	0,406464636	25	51,02040768	0,146766169
84	toro	0,005596744	1	2,040816285	0,965174129
85	trabajar	0,023205256	2	4,081632569	0,81840796
86	tractor	0,15208064	15	30,6122452	0,480099502
87	tranquilidad	0,022769878	2	4,081632569	0,838308458
88	trigo	0,034487631	4	8,163265139	0,771144279
89	uva	0,004198323	1	2,040816285	0,985074627
90	vaca	0,182809025	17	34,69387889	0,415422886
91	vegetación	0,015308909	1	2,040816285	0,880597015
92	verde	0,025154179	4	8,163265139	0,813432836
93	verdura	0,064711459	7	14,28571492	0,674129353
94	verme	0,011483773	1	2,040816285	0,917910448
95	villa	0,013259112	1	2,040816285	0,907960199
96	vino	0,016895296	2	4,081632569	0,868159204

CI12, “medios de transporte”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	a-pie	0,134974703	14	28,00000012	0,74137931
2	andar	0,018677335	3	5,999999866	0,906130268
3	asiento	0,003563461	1	1,999999955	0,990421456
4	atraso	0,004750417	1	1,999999955	0,978927203
5	(auto)bús	0,666665137	46	92,00000167	0,180076628
6	autovía	0,009048271	2	3,999999911	0,946360153
7	avión	0,633473039	49	98,00000191	0,363984674
8	barco	0,231958166	27	54,00000215	0,681992337
9	bici(cleta)	0,599127591	48	95,99999785	0,455938697
10	billete	0,036617614	7	14,00000006	0,837164751
11	caballo	0,053501382	8	15,99999964	0,795019157
12	caminar	0,006332736	1	1,999999955	0,959770115
13	camión	0,041634168	4	7,999999821	0,812260536
14	camioneta	0,002315168	1	1,999999955	1
15	canoa	0,006332736	1	1,999999955	0,957854406
16	caravana	0,003563461	1	1,999999955	0,988505747
17	carne-de-conducir	0,026464157	5	10,00000015	0,873563218
18	carro	0,104649633	6	11,99999973	0,752873563
19	ciclomotor	0,02550108	3	5,999999866	0,879310345
20	coche	0,843886614	48	95,99999785	0,091954023
21	coger	0,046200946	5	10,00000015	0,804597701
22	conducir	0,083559535	14	28,00000012	0,779693487
23	conductor	0,008313878	2	3,999999911	0,95210728
24	confortable	0,003086329	1	1,999999955	0,994252874
25	correr	0,011817545	2	3,999999911	0,932950192
26	despacio	0,005401497	2	3,999999911	0,977011494
27	embarcación	0,016884219	2	3,999999911	0,913793103
28	estación	0,017322086	1	1,999999955	0,908045977
29	funicular	0,002673083	1	1,999999955	0,998084291
30	gasolina	0,007311748	1	1,999999955	0,955938697
31	globo-aerostático	0,012428233	3	5,999999866	0,929118774
32	globo-dirigible	0,003563461	1	1,999999955	0,992337165
33	góndola	0,011254097	1	1,999999955	0,942528736
34	helicóptero	0,032096397	6	11,99999973	0,848659004
35	llevar	0,007311748	1	1,999999955	0,954022989
36	metro	0,272853822	31	62,00000048	0,58045977

37	monopatín	0,012993929	1	1,999999955	0,923371648
38	moto(cicleta)	0,298715442	34	68,00000072	0,521072797
39	motor	0,002673083	1	1,999999955	0,996168582
40	nave	0,030609636	4	7,999999821	0,856321839
41	nave-espacial	0,006172657	2	3,999999911	0,971264368
42	navegar	0,006332736	1	1,999999955	0,961685824
43	navío	0,011254097	1	1,999999955	0,940613027
44	parapente	0,004114356	1	1,999999955	0,986590038
45	parar	0,014774845	2	3,999999911	0,917624521
46	pasajero	0,01131504	3	5,999999866	0,938697318
47	patinetes	0,028085528	4	7,999999821	0,863984674
48	pedalear	0,004114356	1	1,999999955	0,98467433
49	perder	0,005484809	1	1,999999955	0,973180077
50	rápido	0,006236544	2	3,999999911	0,96743295
51	salir	0,02	1	1,999999955	0,896551724
52	subir	0,017322086	1	1,999999955	0,909961686
53	submarino	0,006332736	1	1,999999955	0,963601533
54	tándem	0,013926919	2	3,999999911	0,921455939
55	taxi	0,26460433	26	51,99999809	0,630268199
56	tomar	0,022412691	3	5,999999866	0,888888889
57	tractor	0,022508195	2	3,999999911	0,883141762
58	transportar	0,008442109	1	1,999999955	0,948275862
59	transporte-público	0,004750417	1	1,999999955	0,980842912
60	tranvía	0,145628154	17	34,00000036	0,714559387
61	tren	0,652358174	47	93,99999976	0,270114943
62	triciclo	0,004114356	1	1,999999955	0,982758621
63	velero	0,019494442	2	3,999999911	0,900383142
64	viajar	0,03878564	6	11,99999973	0,823754789
65	yate	0,020108305	3	5,999999866	0,894636015

CI13, “trabajos del campo y del jardín”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	agricultor	0,044765636	3	5,999999866	0,743961353
2	agricultura	0,02	1	1,999999955	0,855072464
3	agua	0,01439903	1	1,999999955	0,90821256
4	arar	0,095892325	8	15,99999964	0,54589372
5	árbol	0,032229085	3	5,999999866	0,77294686
6	arreglar	0,010366603	1	1,999999955	0,93236715
7	caballero	0,007463451	1	1,999999955	0,975845411
8	campesino	0,136261523	8	15,99999964	0,468599034
9	campo	0,059999999	3	5,999999866	0,700483092
10	cavar	0,062595688	5	10,00000015	0,685990338
11	cerdo	0,002785155	1	1,999999955	0,995169082
12	césped	0,007463451	1	1,999999955	0,966183575
13	coger	0,019608459	3	5,999999866	0,874396135
14	cortar	0,405599326	29	57,99999833	0,140096618
15	crecer	0,014615178	3	5,999999866	0,898550725
16	criador	0,041862484	3	5,999999866	0,758454106
17	cultivación	0,02	1	1,999999955	0,850241546
18	cultivar	0,193461567	13	25,99999905	0,376811594
19	dar-agua	0,087212503	11	21,99999988	0,599033816
20	erradicar	0,02	1	1,999999955	0,845410628
21	escavar	0,007463451	1	1,999999955	0,971014493
22	extirpar	0,005373323	1	1,999999955	0,985507246
23	fatiga	0,010366603	1	1,999999955	0,942028986
24	fertilizar	0,007737064	2	3,999999911	0,961352657
25	flor	0,106400467	8	15,99999964	0,507246377
26	florero	0,010366603	1	1,999999955	0,951690821
27	floristería	0,003868532	1	1,999999955	0,990338164
28	granjero	0,086628109	6	11,99999973	0,628019324
29	hierba	0,077896349	7	14,00000006	0,661835749
30	huerto	0,010746646	2	3,999999911	0,927536232
31	ingeniero	0,02	1	1,999999955	0,859903382
32	irrigación	0,01439903	1	1,999999955	0,903381643
33	irrigar	0,005373323	1	1,999999955	0,980676329
34	jardín	0,02879806	2	3,999999911	0,782608696
35	jardinería	0,010366603	1	1,999999955	0,946859903
36	jardinero	0,201662749	16	31,99999928	0,314009662
37	obrero	0,020733206	2	3,999999911	0,835748792

38	plantar	0,258907467	20	40,0000006	0,236714976
39	recoger	0,023585768	5	10,00000015	0,826086957
40	regar	0,139385074	11	21,99999988	0,429951691
41	sembrar	0,047851015	6	11,99999973	0,729468599
42	tierra	0,025373323	2	3,999999911	0,801932367
43	trabajar	0,017830055	2	3,999999911	0,884057971
44	vaquero	0,02879806	2	3,999999911	0,792270531
45	vender	0,02	1	1,999999955	0,84057971
46	verdura	0,002785155	1	1,999999955	1
47	viña	0,010366603	1	1,999999955	0,937198068
48	vino	0,002785155	1	1,999999955	1
49	zapar	0,012836775	2	3,999999911	0,917874396

CI14, “los animales”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abeja	0,03317504	2	3,999999911	0,88340192
2	águila	0,050771013	5	10,00000015	0,823045267
3	alce	0,028897626	2	3,999999911	0,893004115
4	almeja	0,014491913	2	3,999999911	0,927297668
5	alondra	0,010225646	1	1,999999955	0,95473251
6	anchoa	0,00929118	1	1,999999955	0,961591221
7	anguila	0,008442109	1	1,999999955	0,969821674
8	araña	0,028449399	3	5,999999866	0,897119342
9	armadillo	0,022673363	2	3,999999911	0,906721536
10	atún	0,019091146	3	5,999999866	0,914951989
11	avestruz	0,013631714	1	1,999999955	0,930041152
12	avispa	0,009675879	2	3,999999911	0,957475995
13	babosa	0,005228193	1	1,999999955	0,986282579
14	bacalao	0,004750417	1	1,999999955	0,987654321
15	ballena	0,03200243	5	10,00000015	0,890260631
16	bisonte	0,013631714	1	1,999999955	0,92866941
17	buey	0,01349241	3	5,999999866	0,934156379
18	búfalo	0,006505391	2	3,999999911	0,976680384
19	búho	0,00294193	1	1,999999955	0,994513032
20	buitre	0,003237816	1	1,999999955	0,993141289
21	burro	0,173005566	20	40,0000006	0,589849108
22	caballito-de-mar	0,002428804	1	1,999999955	1
23	caballo	0,544799984	40	80,00000119	0,242798354
24	cabra	0,067868851	11	21,99999988	0,781893004
25	cabrón	0,008974832	2	3,999999911	0,964334705
26	cachorro	0,010225646	1	1,999999955	0,951989026
27	camello	0,011296357	2	3,999999911	0,942386831
28	canguro	0,011254097	1	1,999999955	0,945130316
29	caracol	0,042023767	5	10,00000015	0,864197531
30	cebra	0,083228625	9	18,00000072	0,766803841
31	cerdo	0,285189211	28	56,00000024	0,401920439
32	chivo	0,011254097	1	1,999999955	0,946502058
33	ciervo	0,008442109	1	1,999999955	0,968449931
34	cocodrilo	0,015411764	2	3,999999911	0,919067215
35	conejo	0,274814904	30	60,00000238	0,443072702
36	cordero	0,053967878	6	11,99999973	0,801097394
37	culebra	0,014814789	2	3,999999911	0,924554184

38	delfin	0,046756063	7	14,00000006	0,851851852
39	elefante	0,128512338	12	23,99999946	0,694101509
40	escarabajo	0,006332736	1	1,999999955	0,982167353
41	gallina	0,206976339	22	43,99999976	0,562414266
42	gallo	0,045859437	4	7,999999821	0,85733882
43	gato	0,809534907	47	93,99999976	0,131687243
44	gaviota	0,011254097	1	1,999999955	0,943758573
45	gorila	0,008442109	1	1,999999955	0,967078189
46	hámster	0,134284481	12	23,99999946	0,650205761
47	hicotea	0,005754021	1	1,999999955	0,983539095
48	hipopótamo	0,006969654	1	1,999999955	0,972565158
49	hormiga	0,086216152	7	14,00000006	0,754458162
50	insecto	0,008442109	1	1,999999955	0,965706447
51	jirafa	0,222832724	20	40,00000006	0,50617284
52	koala	0,002673083	1	1,999999955	0,995884774
53	león	0,322645724	33	66,00000262	0,327846365
54	letargo	0,006332736	1	1,999999955	0,979423868
55	lobo	0,048646823	4	7,999999821	0,842249657
56	mariposa	0,114002235	12	23,99999946	0,710562414
57	mejillón	0,003563461	1	1,999999955	0,991769547
58	mirlo	0,004316302	1	1,999999955	0,989026063
59	mono	0,132017732	20	40,00000006	0,677640604
60	mosca	0,011083152	2	3,999999911	0,950617284
61	mosquito	0,088295676	10	20,00000003	0,744855967
62	muflón	0,012385986	1	1,999999955	0,935528121
63	murciélago	0,009500833	2	3,999999911	0,960219479
64	nutria	0,024632445	3	5,999999866	0,903978052
65	oso	0,03973313	7	14,00000006	0,873799726
66	oveja	0,059501048	8	15,99999964	0,792866941
67	pájaro	0,546356559	41	81,99999928	0,187928669
68	paloma	0,052645512	5	10,00000015	0,807956104
69	papagayo	0,007670632	1	1,999999955	0,971193416
70	pato	0,026737288	2	3,999999911	0,899862826
71	pavo	0,02	1	1,999999955	0,910836763
72	perezoso	0,002673083	1	1,999999955	0,998628258
73	perro	0,903004229	49	98,00000191	0,067215364
74	pez	0,291956514	26	51,99999809	0,36351166
75	pollo	0,161385313	19	37,99999952	0,615912209
76	puerco	0,011254097	1	1,999999955	0,9478738
77	pulpo	0,015045201	2	3,999999911	0,9218107

78	rana	0,035344303	5	10,00000015	0,880658436
79	rato	0,003563461	1	1,999999955	0,990397805
80	ratón	0,249433324	26	51,99999809	0,478737997
81	reptil	0,005228193	1	1,999999955	0,984910837
82	salmón	0,006332736	1	1,999999955	0,978052126
83	sapo	0,011986933	2	3,999999911	0,939643347
84	serpiente	0,112488836	15	30,00000119	0,731138546
85	sierpe	0,016511641	1	1,999999955	0,916323731
86	tarántula	0,002673083	1	1,999999955	0,997256516
87	ternero	0,012385986	1	1,999999955	0,936899863
88	tiburón	0,051833957	6	11,99999973	0,816186557
89	tigre	0,211527824	19	37,99999952	0,53223594
90	toro	0,049334254	5	10,00000015	0,836762689
91	tortuga	0,15969108	13	25,99999905	0,633744856
92	trucha	0,010225646	1	1,999999955	0,953360768
93	vaca	0,378292769	29	57,99999833	0,282578875
94	verme	0,006332736	1	1,999999955	0,98079561
95	yegua	0,021677164	2	3,999999911	0,909465021
96	zoo	0,006969654	1	1,999999955	0,9739369
97	zorro	0,050499231	5	10,00000015	0,829903978

CI15, “juegos y distracciones”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	aburrir	0,018172309	1	1,999999955	0,933884298
2	ajedrez	0,052730907	5	10,00000015	0,776859504
3	alcohol	0,026614418	2	3,999999911	0,888429752
4	amigo	0,031115472	3	5,999999866	0,859504132
5	aprender	0,00929118	1	1,999999955	0,975206612
6	atracción	0,028987389	3	5,999999866	0,878099174
7	bailar	0,212835938	15	30,00000119	0,353305785
8	balón	0,113802955	10	20,0000003	0,636363636
9	baloncesto	0,311291516	21	41,99999869	0,146694215
10	bar	0,029291181	2	3,999999911	0,871900826
11	beber	0,03489418	3	5,999999866	0,832644628
12	beisbol	0,016511641	1	1,999999955	0,946280992
13	bici(cleta)	0,006332736	1	1,999999955	0,987603306
14	cantar	0,04703629	5	10,00000015	0,799586777
15	cartas	0,266988039	21	41,99999869	0,32231405
16	chatear	0,006332736	1	1,999999955	0,989669421
17	chico	0,016511641	1	1,999999955	0,94214876
18	ciclismo	0,013631714	1	1,999999955	0,960743802
19	cine	0,157209069	12	23,99999946	0,469008264
20	cocinar	0,018924728	2	3,999999911	0,923553719
21	colorar	0,018172309	1	1,999999955	0,935950413
22	comer	0,075790167	7	14,00000006	0,73553719
23	competición	0,011254097	1	1,999999955	0,966942149
24	consola	0,011254097	1	1,999999955	0,97107438
25	correr	0,177313745	16	31,99999928	0,444214876
26	dado	0,011254097	1	1,999999955	0,969008264
27	damas	0,036139909	3	5,999999866	0,818181818
28	deporte	0,128030851	10	20,0000003	0,597107438
29	descansar	0,015002731	1	1,999999955	0,954545455
30	dibujar	0,062068444	5	10,00000015	0,766528926
31	disco(teca)	0,08409442	6	11,99999973	0,72107438
32	distraer	0,03317504	2	3,999999911	0,845041322
33	divertir	0,030143354	2	3,999999911	0,863636364
34	dormir	0,002428804	1	1,999999955	1
35	entrenarse	0,016112741	2	3,999999911	0,950413223
36	escondite	0,024678258	3	5,999999866	0,909090909
37	escribir	0,025637412	3	5,999999866	0,894628099

38	escuchar-música	0,133220613	9	18,00000072	0,55785124
39	esquiar	0,025228376	2	3,999999911	0,902892562
40	<i>Facebook</i>	0,018172309	1	1,999999955	0,931818182
41	fiesta	0,007670632	1	1,999999955	0,98553719
42	fotografía	0,022922892	2	3,999999911	0,91322314
43	fútbol	0,311100811	21	41,99999869	0,190082645
44	gimnasio	0,018223751	2	3,999999911	0,92768595
45	golf	0,010225646	1	1,999999955	0,973140496
46	guitarra	0,00929118	1	1,999999955	0,979338843
47	instrumento-musical	0,025228376	2	3,999999911	0,898760331
48	internet	0,018172309	1	1,999999955	0,929752066
49	ir-de-compras	0,049707759	6	11,99999973	0,789256198
50	juego-de-mesa	0,073251463	6	11,99999973	0,747933884
51	jugar	0,473836452	29	57,99999833	0,059917355
52	juguete	0,013631714	1	1,999999955	0,962809917
53	leer	0,326024234	21	41,99999869	0,103305785
54	libro	0,296230674	21	41,99999869	0,233471074
55	marioneta	0,016511641	1	1,999999955	0,944214876
56	<i>Monopoly</i>	0,02	1	1,999999955	0,917355372
57	móvil	0,194394991	13	25,99999905	0,411157025
58	muñeca	0,087431885	7	14,00000006	0,698347107
59	nadar	0,145325914	13	25,99999905	0,518595041
60	natación	0,100342184	9	18,00000072	0,683884298
61	navegar	0,00929118	1	1,999999955	0,977272727
62	ordenador	0,195679307	15	30,00000119	0,384297521
63	pádel	0,008442109	1	1,999999955	0,981404959
64	parque	0,016961811	2	3,999999911	0,940082645
65	pasear	0,120822459	9	18,00000072	0,615702479
66	paseo	0,02	1	1,999999955	0,915289256
67	pelí(cula)	0,066198327	4	7,999999821	0,756198347
68	pelota	0,131815135	9	18,00000072	0,576446281
69	peluche	0,008442109	1	1,999999955	0,983471074
70	periódico	0,031929422	4	7,999999821	0,853305785
71	pintar	0,046348095	4	7,999999821	0,80785124
72	piscina	0,034748875	4	7,999999821	0,840909091
73	<i>Play-Station</i>	0,084660284	5	10,00000015	0,708677686
74	playa	0,012385986	1	1,999999955	0,964876033
75	practicar	0,015002731	1	1,999999955	0,952479339
76	redes-sociales	0,112856261	6	11,99999973	0,648760331

77	salir-con-amigos	0,028215909	3	5,999999866	0,884297521
78	serie	0,030143354	2	3,999999911	0,867768595
79	surf	0,02	1	1,999999955	0,919421488
80	tapear	0,005754021	1	1,999999955	0,995867769
81	teatro	0,035757661	4	7,999999821	0,826446281
82	tele(visión)	0,272362143	22	43,99999976	0,27892562
83	tenis	0,151045844	11	21,99999988	0,491735537
84	tocar	0,039999999	2	3,999999911	0,811983471
85	viajar	0,013927179	2	3,999999911	0,958677686
86	videojuegos	0,133351162	10	20,0000003	0,539256198
87	visitar	0,006179745	2	3,999999911	0,993801653
88	volar	0,003563461	1	1,999999955	0,997933884
89	voleibol	0,102000244	8	15,99999964	0,665289256

CI16, “profesiones y oficios”

Nº	Vocablo	ID	Frecuencia	% Aparición	Frec. Acumulada
1	abogado	0,491881609	30	60,00000238	0,114649682
2	actor	0,281964391	24	47,99999893	0,305732484
3	actriz	0,056981489	6	11,99999973	0,746815287
4	agricultor	0,021479744	2	3,999999911	0,891719745
5	albañil	0,057614364	4	7,999999821	0,737261146
6	alcalde	0,030181944	3	5,999999866	0,867834395
7	alcaldesa	0,015002731	1	1,999999955	0,933121019
8	ama-de-casa	0,042673364	3	5,999999866	0,826433121
9	arquitecto	0,056401234	4	7,999999821	0,753184713
10	artista	0,007435041	2	3,999999911	0,97611465
11	asistente-de-vuelo	0,049348917	4	7,999999821	0,808917197
12	astrónomo	0,016511641	1	1,999999955	0,914012739
13	autónomo	0,016511641	1	1,999999955	0,915605096
14	autor	0,017675813	2	3,999999911	0,912420382
15	azafata	0,04488581	3	5,999999866	0,813694268
16	bailarín	0,113824271	13	25,99999905	0,597133758
17	barista	0,056311205	6	11,99999973	0,762738854
18	bombero	0,115266412	8	15,99999964	0,576433121
19	cajero	0,012385986	1	1,999999955	0,945859873
20	camarero	0,449185342	38	75,99999905	0,175159236
21	campesino	0,080714539	8	15,99999964	0,670382166
22	cantante	0,121199481	13	25,99999905	0,546178344
23	cantor	0,00929118	1	1,999999955	0,96656051
24	carnicero	0,040330008	5	10,00000015	0,845541401
25	carpintero	0,022265663	2	3,999999911	0,888535032
26	cartero	0,016511641	1	1,999999955	0,917197452
27	cazador	0,015623915	2	3,999999911	0,923566879
28	chef	0,056262288	4	7,999999821	0,76910828
29	chofer	0,005444665	2	3,999999911	0,993630573
30	cirujano	0,029733239	3	5,999999866	0,872611465
31	cocinar	0,00929118	1	1,999999955	0,958598726
32	cocinero	0,293964416	29	57,99999833	0,267515924
33	contador	0,012385986	1	1,999999955	0,944267516
34	cosmetólogo	0,005228193	1	1,999999955	0,996815287
35	criador	0,019696208	2	3,999999911	0,899681529
36	defender	0,015002731	1	1,999999955	0,928343949
37	dentista	0,028942041	3	5,999999866	0,877388535

38	deportista	0,00929118	1	1,999999955	0,968152866
39	despacho	0,00929118	1	1,999999955	0,957006369
40	director	0,09380886	8	15,99999964	0,648089172
41	director-de-cine	0,015411764	2	3,999999911	0,926751592
42	doctor	0,416348994	29	57,99999833	0,22133758
43	educador	0,018172309	1	1,999999955	0,909235669
44	embajador	0,015002731	1	1,999999955	0,929936306
45	empleado	0,097376049	8	15,99999964	0,635350318
46	empresario	0,056257438	5	10,00000015	0,777070064
47	enfermero	0,195621982	13	25,99999905	0,385350318
48	enseñar	0,02	1	1,999999955	0,893312102
49	entrenador	0,098051094	8	15,99999964	0,622611465
50	entrevista-de-trabajo	0,006969654	1	1,999999955	0,982484076
51	escritor	0,127062872	8	15,99999964	0,50477707
52	esteticista	0,012385986	1	1,999999955	0,942675159
53	estilista	0,007258232	2	3,999999911	0,979299363
54	estudiante	0,148478016	13	25,99999905	0,431528662
55	farmacéutico	0,042128459	4	7,999999821	0,832802548
56	farmacia	0,00929118	1	1,999999955	0,960191083
57	físico	0,00929118	1	1,999999955	0,97133758
58	florero	0,010225646	1	1,999999955	0,950636943
59	fontanero	0,082016304	6	11,99999973	0,657643312
60	fotógrafo	0,0433002	5	10,00000015	0,821656051
61	futbolista	0,055663653	4	7,999999821	0,78343949
62	general	0,006332736	1	1,999999955	0,988853503
63	granjero	0,019696208	2	3,999999911	0,896496815
64	guía-de-turismo	0,051452156	5	10,00000015	0,797770701
65	horario	0,008442109	1	1,999999955	0,972929936
66	hospital	0,010225646	1	1,999999955	0,949044586
67	ingeniero	0,061955404	5	10,00000015	0,73089172
68	instructor	0,00929118	1	1,999999955	0,964968153
69	investigador	0,006332736	1	1,999999955	0,98566879
70	jardinero	0,145388886	12	23,99999946	0,472929936
71	jefe	0,068372279	5	10,00000015	0,722929936
72	juez	0,130046636	12	23,99999946	0,492038217
73	maestro	0,04146539	3	5,999999866	0,837579618
74	maquinista	0,006332736	1	1,999999955	0,987261146
75	mecánico	0,004750417	1	1,999999955	0,998407643
76	mediador	0,01648229	2	3,999999911	0,920382166

77	médico	0,179588169	16	31,99999928	0,410828025
78	militar	0,015002731	1	1,999999955	0,931528662
79	modelo	0,034165546	5	10,00000015	0,863057325
80	músico	0,076955706	12	23,99999946	0,689490446
81	obrero	0,099162839	8	15,99999964	0,609872611
82	oficina	0,039999999	2	3,999999911	0,848726115
83	oficinista	0,004750417	1	1,999999955	1
84	panadero	0,122084573	13	25,99999905	0,525477707
85	pastelero	0,00929118	1	1,999999955	0,969745223
86	peluquero	0,222062781	22	43,99999976	0,340764331
87	periodista	0,054891273	4	7,999999821	0,789808917
88	pescador	0,076532155	8	15,99999964	0,702229299
89	piloto	0,034435578	4	7,999999821	0,855095541
90	pintor	0,076465681	8	15,99999964	0,714968153
91	plomero	0,00929118	1	1,999999955	0,961783439
92	poeta	0,012617809	3	5,999999866	0,941082803
93	policía	0,148127049	14	28,00000012	0,453821656
94	político	0,010225646	1	1,999999955	0,953821656
95	portero	0,01935564	2	3,999999911	0,906050955
96	presidente	0,005754021	1	1,999999955	0,99044586
97	profesor	0,709641576	42	83,99999738	0,066878981
98	psicólogo	0,006969654	1	1,999999955	0,98089172
99	psiquiatra	0,019516828	2	3,999999911	0,902866242
100	químico	0,010225646	1	1,999999955	0,955414013
101	repcionista	0,010225646	1	1,999999955	0,952229299
102	redactor	0,028897626	2	3,999999911	0,880573248
103	restaurador	0,005228193	1	1,999999955	0,99522293
104	secretario	0,198214173	15	30,00000119	0,364649682
105	soldado	0,006969654	1	1,999999955	0,984076433
106	taller	0,018172309	1	1,999999955	0,907643312
107	tienda	0,011254097	1	1,999999955	0,947452229
108	trabajador	0,013631714	1	1,999999955	0,936305732
109	trabajar	0,120336145	11	21,99999988	0,563694268
110	traducción	0,013631714	1	1,999999955	0,934713376
111	traductor	0,027839825	3	5,999999866	0,885350318
112	vendedor	0,049534921	3	5,999999866	0,802547771
113	veterinario	0,00929118	1	1,999999955	0,963375796

ANEXO VII – DATOS DE RIQUEZA LÉXICA

(CORPUS GENERAL)

Texto	Informante	<i>Tokens</i>	<i>Types</i>	<i>Tokens</i> léxicos	Hápax
1	11200 101	100	64	48	52
2	11200 102	100	53	55	35
3	21200 103	100	66	56	46
4	21100 104	100	62	50	50
5	21200 105	100	61	59	48
6	11200 106	100	57	55	38
7	11200 107	100	60	59	42
8	11200 108	100	60	55	42
9	11200 109	100	60	55	42
10	21100 110	100	64	57	47
11	11200 111	100	62	59	47
12	11200 112	100	65	60	44
13	11200 113	100	63	52	45
14	11200 114	100	59	55	39
15	11200 115	100	68	63	49
16	11200 116	100	53	52	30
17	11200 117	100	65	53	48
18	11200 118	100	61	43	42
19	11200 119	100	53	51	36
20	11200 120	100	66	50	58
21	11200 121	100	59	53	39
22	11200 122	100	65	52	51
23	11200 123	100	59	52	39
24	11200 124	100	66	55	52
25	11200 125	100	71	56	60
26	11200 126	100	67	55	51
27	11100 127	100	59	55	42
28	11100 128	100	56	60	34
29	11100 129	100	58	55	41
30	11200 130	100	64	60	41
31	11200 131	100	62	53	58
32	11200 132	100	75	64	58
33	11200 133	100	64	53	51
34	11100 134	100	61	48	44

35	11200 135	100	64	56	51
36	11100 136	100	69	66	49
37	11200 137	100	59	51	38
38	11200 138	100	69	54	52
39	11200 139	100	63	50	47
40	11100 140	100	54	49	32
41	11100 141	100	63	61	41
42	11200 142	100	70	58	55
43	11200 143	100	61	60	43
44	11200 144	100	62	59	45
45	11200 145	100	57	54	40
46	11200 146	100	66	60	49
47	11100 147	100	66	60	51
48	11200 148	100	64	57	47
49	11200 149	100	65	55	52
50	11200 150	100	61	54	42
51	12200 051	100	67	61	49
52	22100 052	100	54	50	33
53	22100 053	100	62	50	45
54	12100 054	100	60	54	40
55	12200 055	100	59	53	45
56	22200 056	100	62	53	44
57	12100 057	100	55	48	38
58	12200 058	100	65	63	47
59	22200 059	100	71	57	57
60	22200 060	100	66	56	48
61	12200 061	100	61	48	44
62	12200 062	100	67	60	51
63	12200 063	100	62	53	47
64	22200 064	100	64	55	48
65	22200 065	100	61	58	42
66	12200 066	100	63	55	42
67	22200 067	100	54	50	34
68	12200 068	100	66	57	51
69	12200 069	100	68	59	53
70	12200 070	100	53	49	34
71	12200 071	100	65	52	51
72	12200 072	100	64	52	50
73	12200 073	100	67	52	50
74	12200 074	100	57	50	39
75	12200 075	100	73	56	56

76	12200 076	100	66	55	45
77	12200 077	100	70	61	51
78	12200 078	100	58	46	41
79	12100 079	100	56	55	34
80	12200 080	100	59	58	42
81	12200 081	100	72	59	57
82	12100 082	100	67	59	50
83	12200 083	100	64	58	48
84	12200 084	100	67	52	52
85	12200 085	100	61	54	42
86	12200 086	100	65	57	46
87	12200 087	100	63	49	48
88	12200 088	100	66	53	52
89	12200 089	100	65	48	50
90	12200 090	100	61	53	43
91	12200 091	100	63	52	49
92	12200 092	100	58	54	38
93	12200 093	100	60	57	40
94	12200 094	100	69	57	51
95	12100 095	100	67	58	50
96	12100 096	100	66	55	51
97	12200 097	100	67	54	52
98	12200 098	100	59	56	40
99	12200 099	100	67	57	49
100	22100 100	100	61	50	44

ANEXO VIII – DATOS DE RIQUEZA LÉXICA
(CORPUS B2_ b)

Texto	Informante	<i>Tokens</i>	<i>Types</i>	<i>Tokens léxicos</i>	Hápax
1	22200 001	100	57	58	35
2	22200 002	100	46	42	29
3	12200 003	100	67	52	55
4	22100 004	100	69	56	55
5	12100 005	100	59	54	43
6	12200 006	100	60	60	36
7	12200 007	100	59	55	40
8	12200 008	100	62	51	46
9	12200 009	100	55	54	35
10	12200 010	100	63	53	49
11	12200 011	100	62	51	43
12	12200 012	100	62	52	46
13	12200 013	100	61	54	44
14	12200 014	100	66	58	47
15	12200 015	100	58	57	41
16	12200 016	100	61	62	45
17	12200 017	100	70	58	55
18	12200 018	100	51	53	44
19	12200 019	100	62	53	46
20	12200 020	100	60	56	44
21	12200 021	100	60	51	44
22	12200 022	100	67	56	56
23	12200 023	100	62	55	46
24	12200 024	100	60	53	39
25	12100 025	100	60	55	40
26	12200 026	100	58	55	40
27	12200 027	100	68	55	56
28	12200 028	100	62	55	41
29	12200 029	100	67	55	53
30	12200 030	100	66	56	54
31	12200 031	100	62	50	44
32	12200 032	100	57	47	39
33	12200 033	100	60	56	43
34	12200 034	100	60	58	42
35	12200 035	100	62	58	44
36	12200 036	100	56	50	37

37	12200 037	100	68	54	56
38	12200 038	100	61	60	41
39	12200 039	100	60	53	42
40	12200 040	100	67	56	50
41	12200 041	100	60	58	46
42	12200 042	100	62	55	43
43	22100 043	100	58	52	45
44	12200 044	100	55	54	35
45	12200 045	100	63	57	48
46	22200 046	100	60	57	41
47	12200 047	100	61	61	45
48	12200 048	100	60	57	42
49	12200 049	100	60	57	48
50	12200 050	100	63	57	47