



UNIVERSITAT DE  
BARCELONA

## El efecto de la tarea en producciones escritas analizadas con ordenador: un estudio comparativo entre sujetos nativos y no nativos

Rosa María Lucha Cuadros

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



FACULTAD DE EDUCACIÓN

**PROGRAMA DE DOCTORADO**

**EN DIDÁCTICA DE LA LENGUA Y LA LITERATURA**

***“El efecto de la tarea en producciones  
escritas analizadas con ordenador: un estudio  
comparativo entre sujetos Nativos y no Nativos”***

---

Tesis para optar por el título de Doctor

Tesis doctoral presentada por: ROSA M<sup>a</sup> LUCHA CUADROS

Directora: DRA. LOURDES DÍAZ

Tutor: DR. JOAN PERERA

**Barcelona, octubre de 2015**

A mi familia

## AGRADECIMIENTOS

En primer lugar, deseo agradecer a la Dra. Lourdes Díaz, directora de mi tesis doctoral, su apoyo incondicional para iniciar este proyecto sobre el maravilloso mundo del español para extranjeros. Quiero expresar, igualmente, mi gratitud por confiar en mí, dedicarme su tiempo, poner a mi disposición todos sus conocimientos y orientarme en el proceso de mi tesis. Sin su apoyo, tanto en el ámbito académico como en el personal, este proyecto no habría ido posible.

A los tutores, Dra. M<sup>a</sup> Rosa Torras y Dr. Joan Perera de la Universidad de Barcelona quiero expresarles mi gratitud por estar ahí, a lo largo del desarrollo de mi tesis y por haber formado parte activa en todo el proceso de mi formación, tanto en el Máster de Formación de Profesores de ELE como en los cursos de Doctorado. Ellos han contribuido a despertar en mí el gusto por enseñar lengua e investigar sobre el uso de la misma.

En este proceso de aprendizaje e investigación también quiero agradecer la dedicación y trabajo realizado por todo el profesorado del Máster y del Doctorado.

También destacar la actitud de colaboración y eficacia del personal administrativo, facilitándome los documentos e informaciones necesarias.

No puedo dejar de recordar a todas las compañeras y compañeros de estudios del Máster y Doctorado. Me gustaría darles las gracias por haber creado un ambiente de compañerismo, de ayuda y de apoyo en momentos difíciles. Por haber compartido materiales y recursos entre todos, que hacían más fáciles nuestros aprendizajes. De esta buena relación entre compañeros, quiero destacar la amistad surgida durante el Máster con Ana, Ari, Silvia y Vanesa, y que perdura en la actualidad tanto a nivel profesional, compartiendo nuestras experiencias, como a nivel personal. Durante el Doctorado, quiero destacar la relación de amistad que surgió entre Cynthia, Ricardo Conchita, Eréndira, Yc y Zailce. Casi se puede decir que formamos una pequeña familia multicultural, en la que compartimos muy buenos momentos. A Cynthia, quiero darle las gracias por tantos momentos que hemos compartido preparando proyectos académicos y artículos para presentar en Congresos de Enseñanza del Español como Lengua Extranjera. A Ricardo, por su amistad y por su inestimable orientación técnica. A Eréndira por su aportación de material, tan importante para mi investigación. A Yc por animarme en todo momento y por su actitud positiva. A Zailce y Conchita por todos esos momentos de trabajo y diversión compartidos. ¡Gracias a todos!

Merece especial mención, Nidya, por su incalculable aportación logística, por su entusiasmo y actitud positiva que han hecho que llegue al final de mi proyecto sin rendirme.

Sin duda alguna, mi paso por la EOI de Vall d'Hebron y Drassanes, como profesora de ELE, me ha aportado una gran experiencia. Quiero agradecer al profesorado del departamento de español de ambas escuelas, la profesionalidad con la que trabajan, su entrega, su espíritu colaborador, el trabajo en equipo... Todo ello me ha servido para enriquecerme a nivel profesional y personal. Quiero expresar mi gratitud, especialmente, al departamento de español de la EOI de Vall d'Hebron por haberme facilitado la recogida de datos. Sin la colaboración y las buenas prácticas de dicho departamento, mi trabajo de investigación no hubiera sido posible. También quiero agradecer al equipo de profesores de los cursos de verano de la EOI de Drassanes el buen ambiente de trabajo que he tenido la oportunidad de compartir con ellos durante tantos años.

Por último, agradecer a mi familia todo el apoyo que me ha dado, las muestras de cariño y afecto que me ha demostrado en todo momento.

## RESUMEN

Este trabajo explora, mediante dos subgéneros diferentes de tareas escritas (una narración en pasado en primera persona - *inventar una historia* - contra una primera persona bajo el recuerdo emocional - *explicar una experiencia personal* -, cómo se desarrollan las habilidades de escritura en L2 español longitudinalmente en un entorno de enseñanza formal. La recogida de los datos de producción escrita de (i) nivel B1 L2 español de alumnos que estudiaban en la Escuela Oficial de Idiomas de Barcelona-Vall d'Hebron y de (ii) alumnos nativos estudiantes universitarios fueron transcritos y analizados con el fin de encontrar diferencias y similitudes entre dos etapas diferentes (inicio del nivel B1, final de B1 de estudiantes no nativos) y poblaciones distintas (no nativos *vs* hablantes nativos). Se diseñaron, para ello, dos tareas de elicitación que se administraron a los sujetos no nativos en dos momentos diferentes (inicio y final de curso B 1), y a un grupo nativo que fue tomado como grupo de control. El número de sujetos (casi doscientos) sitúa este estudio entre aquellos en los que los datos empíricos son el punto fuerte (como ocurre con CEDEL-2, Lozano y Mendikoetxea, 2008, 2010). El procedimiento semi-experimental implementado para la recopilación de datos entre los hablantes nativos y estudiantes de L2 español no nativos permitió dar cuenta de las similitudes y diferencias encontradas entre (i) los grupos (nativos *vs* no nativos), (ii) el tiempo transcurrido (entre el inicio y el final del nivel B1 del grupo de no nativos), (iii) así como sobre la mayor dificultad en producir una tarea respecto a la otra. Los análisis cualitativos y cuantitativos de los datos se realizaron a través de las distintas muestras del corpus y tomando en consideración los dos géneros textuales, de manera conjunta y por separado. Así, se cambia de un procedimiento estadístico a otro con el fin de capturar los mecanismos subyacentes a través de sujetos y grupos.

Esta investigación va en la línea de algunos trabajos pioneros de Manchón (2009, 2011, 2012, 2013, 2014), Polio (2001, 2003, 2012) o Larsen Freeman (1991, 1995), relacionados con el análisis de la tarea y de los géneros en inglés (principalmente), estudios casi inexistentes para L2 español, salvo por (SPOLLOC, CEDEL-2 es la excepción). Nuestro estudio hace hincapié en los problemas comunes en L1-L2 respecto a las habilidades de escritura, por ejemplo, o también las dificultades que presentan los estudiantes para escribir a través de géneros diferentes

dentro de un contexto de aprendizaje no específicamente académico-universitario y, además, multilingüe.

La organización del estudio es el siguiente. En primer lugar se presenta la Introducción. En el capítulo 1 se presenta el estado del arte. En el capítulo 2 se presentan los objetivos, las preguntas de investigación e hipótesis. En el capítulo 3 nos ocupamos de las cuestiones metodológicas. En el Capítulo 4, ofrecemos el Estudio. En el capítulo 5, el análisis de la metodología. En el capítulo 6, se presentan los resultados de la estadística descriptiva e inferencial. En el capítulo 7 se discuten los resultados. Las conclusiones se presentan en el capítulo 8, y a continuación vienen las referencias bibliográficas y anexos.

Un Índice de tablas, gráficos y figuras se puede encontrar en las páginas viii; xv y xxiii, respectivamente; con el fin de hacer más fácil encontrar los datos y los resultados resumidos.

## ABSTRACT

This work explores how writing skills in L2 Spanish develop longitudinally across two different subgenres of written tasks collected in an instructional setting, namely a 1<sup>st</sup> person narrative vs a 1<sup>st</sup> person emotional recall. Data from B1 production of L2 Spanish from learners studying at the Escola Oficial d'Idiomes de Barcelona- Vall d'Hebron have been transcribed and analyzed in order to find differences and similitudes among two different stages (onset of B1, end of B1) and populations (non-native vs native speakers). Two elicitation tasks were designed and administered to two non-native groups (whose comparability was statistically established and tested) at two different moments, and to one native group taken as control group. The number of subjects (nearly two hundred) places this study among the ones in which empirical data are a strong point (as occurs with *CEDEL-2*, Mendikoetxea and Lozano 2008, 2010). The semi-experimental procedure implemented for data gathering among both native speakers and non-native Spanish L2 learners allowed us to account for similitudes and differences found across groups and time, as well as concerning difficulties of one task over the other. Qualitative and quantitative analyses of data have been conducted across corpora and text types, both together and separately, switching from one statistical procedure to another in order to capture underlying mechanisms across subjects and groups.

This research goes along the lines of some pioneering work by Manchón (2009, 2011, 2012, 2013, 2014), Polio (2001, 2003, 2012), Larsen Freeman (1991, 1995) concerning task analyses across time and genres in English (mainly), almost non-existent for L2 Spanish, though (*SPOLLOC*, *CEDEL-2* being the exception). Our study stresses the common issues in L1-L2 written skills, and pictures B1 students' difficulties in writing across genres and time when learning in a non-specific non-academic multilingual context of instructed *plus* immersion L2 setting.

The organization of the study is as follows. First we present the Introduction. In Chapter 1 we present the State of the Art. In chapter 2 we present the objectives, research questions and hypotheses. In Chapter 3 we deal with the Methodological issues. In Chapter 4 we offer the Study. In Chapter 5, methodology analysis. In Chapter 6, we present the results of descriptive and inferential statistics. In Chapter

7 we discuss the results. Conclusions are presented in Chapter 8, and below are the bibliography and annexes.

An Index of Tables/Figures can be found in pages, viii; xv y xxiii; in order to make easier finding data and summarized results.

# ÍNDICE

ÍNDICE DE TABLAS	xi
ÍNDICE DE GRÁFICOS	xvii
ÍNDICE DE FIGURAS	xxv
<b>INTRODUCCIÓN</b>	<b>1</b>
<b>JUSTIFICACIÓN</b>	<b>5</b>

## PARTE I ESTADO DE LA CUESTIÓN

### CAPÍTULO 1 ESTADO DE LA CUESTIÓN

1.1	Antecedentes: teorías en Adquisición de Segundas Lenguas	7
1.1.1	Teorías innatistas	8
1.1.2	Teorías medioambientales	10
1.1.3	Teorías basadas en la interacción	12
1.2	Antecedentes: investigación sobre proceso, agentes y estrategias en L2	15
1.3	El desarrollo de la escritura en L2	21
1.4	El desarrollo de la escritura en español L2/LE	26
1.5	La lingüística de corpus en la investigación de segundas lenguas	33
1.6	Ser nativo de una lengua: ¿Qué lengua enseñar?	35
1.6.1	Enfoque comunicativo: ¿Cómo enseñar la lengua meta?	37
1.6.2	El enfoque transcultural: ¿Qué lengua enseñar?	40
1.6.3	La lengua meta a enseñar	41
1.7	La enseñanza mediante tareas	43
1.7.1	Definición de tarea	43
1.7.2	Ventajas de trabajar mediante tareas	45
1.7.3	Diseño de las tareas	47

**PARTE II**  
**OBJETIVOS, PREGUNTAS DE**  
**INVESTIGACIÓN Y MARCO**  
**METODOLÓGICO**

**CAPÍTULO 2 OBJETIVOS, PREGUNTAS DE INVESTIGACIÓN**  
**E HIPÓTESIS**

2.1	Objetivo general	53
2.2	Objetivos específicos	53
2.3	Preguntas de investigación	54
2.4	Hipótesis de partida	55

**CAPÍTULO 3 MARCO METODOLÓGICO**

3.1	La investigación cuantitativa	57
3.2	La investigación pseudolongitudinal	60
3.3	Las medidas de calidad general de un texto escrito	61
3.3.1	La producción escrita en el ámbito de las segundas lenguas	61
3.3.2	La fluidez	62
3.3.3	La diversidad léxica	63
3.3.4	La densidad léxica	65

### PARTE III

## ESTUDIO Y METODOLOGÍA DE ANÁLISIS

### CAPÍTULO 4 EL ESTUDIO

4.1	El contexto	67
4.1.1	Los participantes	67
4.1.1.1	Los sujetos Nativos	67
4.1.1.2	Los sujetos No Nativos	68
4.1.2	Las instituciones	73
4.1.2.1	La Universitat Pompeu Fabra	73
4.1.2.2	La Escuela Oficial de Idioma	74
4.1.2.3	El contexto del Español como Lengua Extranjera (ELE) en las EEOOI	75
4.2	Corpus y Recogida de datos	75
4.2.1	Corpus y Recogida de datos de Sujetos Nativos	77
4.2.2	Corpus y Recogida de datos de Sujetos No Nativos	78

### CAPÍTULO 5 METODOLOGÍA DE ANÁLISIS

5.1	Instrumentos para el análisis lingüístico	81
5.1.1	AntConc	81
5.1.2	Paramtext	83
5.2	Instrumentos para el análisis estadístico	84
5.2.1	G-Stat 2.0: software estadístico de acceso libre	85
5.2.1.1	Fases del análisis estadístico	87

**PARTE IV**  
**RESULTADOS, ANÁLISIS Y DISCUSIÓN**  
**RESULTADOS y ANÁLISIS**

**CAPÍTULO 6**

6.1	Estadística descriptiva	103
6.1.1	Estadística descriptiva para la muestra «F» por grupos	104
6.1.2	Estadística descriptiva para la muestra «1D» por grupos	131
6.1.3	Estadística descriptiva para el M1 y M2 del grupo de NN	161
6.1.4	Estadística descriptiva para la muestra «F» por estadios	189
6.1.5	Estadística descriptiva para la muestra «1D»	215
6.2	Resultados y análisis de la «Estadística inferencial»	245
6.2.1	Resultados y análisis de la comparación del texto «F» y del texto «1D»	245
6.2.2	Resultados y análisis de la comparación de los dos momentos de recogida, M1 y M2, del grupo de NN	295
6.2.3	Resultados y análisis de comparar el grupo NN y el grupo N	337

**CAPÍTULO 7 DISCUSIÓN**

7.1	Comparación de las muestras del texto «F» y del texto «1D»	357
7.2	Comparación del grupo NN- M1 y M2	361
7.3	Comparación del grupo NN y el grupo N: fluidez, diversidad léxica y densidad léxica en las muestras «F» y «1D»	367

**CAPÍTULO 8 CONCLUSIONES** 373

**REFERENCIAS BIBLIOGRÁFICAS** 381

<b>ANEXOS</b>	i
ANEXO 1 Plantilla para escribir el texto ‘F’	ii
ANEXO 2 Plantilla para escribir el texto ‘1D’	iii

## ÍNDICE DE TABLAS

Tabla 1: Clasificación Teorías de ASL Fuente: Larsen-Freeman y Long (1991) .....	8
Tabla 2. Concepto de ‘Competencia comunicativa’ Adaptado de Widdowson (1995) .....	36
Tabla 3. Género de los sujetos de la muestra en valores absolutos. ....	67
Tabla 4. Estudios de los sujetos de la muestra. ....	70
Tabla 5. Nivel de estudios de los sujetos de la muestra según el género. ....	71
Tabla 6. Acceso al nivel Intermedio B-1. ....	72
Tabla 7. Total de producciones de los sujetos Nativos. ....	77
Tabla 8. Total de producciones de los sujetos NN en el M1. ....	79
Tabla 9. Total de producciones de los sujetos NN en el M2. ....	79
Tabla 10. Total de producciones de los sujetos NN en el M1 y M2. ....	79
Tabla 11. Clasificación de las muestras por tipo de texto .....	88
Tabla 12. Clasificación de las muestras por estadios (nivel de lengua) .....	88
Tabla 13. Resultados de la prueba de <i>Shapiro-Wilk</i> para el grupo NN en M1 y M2 .....	89
Tabla 14. Resultados de la prueba de <i>Shapiro-Wilk</i> para NN .....	90
Tabla 15. Clasificación de las muestras .....	95
Tabla 16. Clasificación de las muestras por grupo y tipo de texto. ....	103
Tabla 17. Clasificación de las muestras por M1-M2 del grupo de NN. ....	103
Tabla 18. Clasificación de las muestras por estadios (nivel de lengua) y tipo de texto. ....	104
Tabla 19. Totales de la muestra «F» .....	104
Tabla 20. Estadísticos para la variable « <i>Tokens</i> » NN – N .....	105
Tabla 21. Tabla de estadísticos para la variable « <i>Types</i> » NN – N .....	108
Tabla 22. Tabla de estadísticos para la variable «Verbos» NN – N .....	111
Tabla 23. Tabla de estadísticos para la variable «Sustantivos» NN – N .....	113
Tabla 24. Tabla de estadísticos para la variable «Adjetivos» NN – N .....	115
Tabla 25. Tabla de estadísticos para la variable «Adverbios» NN – N .....	118
Tabla 26. Tabla de estadísticos para la variable «Pronombres» NN – N .....	121

Tabla 27. Tabla de estadísticos para la variable «Preposiciones» NN – N .....	124
Tabla 28. Tabla de estadísticos para la variable «Artículos» NN – N .....	126
Tabla 29. Tabla de estadísticos para la variable «Conjunciones» NN – N .....	129
Tabla 30. Tabla de totales de la muestra «NN-1D / N-1D» .....	132
Tabla 31. Estadísticos para la variable «Tokens» NN-1D / N-1D .....	132
Tabla 32. Estadísticos para la variable «Types» NN-1D / N-1D .....	135
Tabla 33. Estadísticos para la variable «Verbos» NN-1D / N-1D .....	138
Tabla 34. Estadísticos para la variable «Sustantivos» NN-1D / N-1D .....	141
Tabla 35. Estadísticos para la variable «Adjetivos» NN-1D / N-1D .....	144
Tabla 36. Estadísticos para la variable «Adverbios» NN-1D / N-1D .....	147
Tabla 37. Estadísticos para la variable «Pronombres» NN-1D / N-1D .....	150
Tabla 38. Estadísticos para la variable «Preposiciones» NN-1D / N-1D .....	153
Tabla 39. Estadísticos para la variable « Artículos» NN-1D / N-1D .....	156
Tabla 40. Estadísticos para la variable «Conjunciones» NN-1D / N-1D .....	159
Tabla 41. Totales de los «M-1» y «M-2» NN-M1 / NN-M2 .....	162
Tabla 42. Estadísticos para la variable «Tokens» NN-M1 / NN-M2 .....	162
Tabla 43. Estadísticos para la variable «Types» NN-M1 / NN-M2 .....	165
Tabla 44. Estadísticos para la variable «Verbos» NN-M1 / NN-M2 .....	168
Tabla 45. Estadísticos para la variable «Sustantivos» NN-M1 / NN-M2 .....	171
Tabla 46. Estadísticos para la variable «Adjetivos» NN-M1 / NN-M2 .....	173
Tabla 47. Estadísticos para la variable «Adverbios» NN-M1 / NN-M2 .....	176
Tabla 48. Estadísticos para la variable «Pronombres» NN-M1 / NN-M2 .....	179
Tabla 49. Estadísticos para la variable «Preposiciones» NN-M1 / NN-M2 .....	181
Tabla 50. Estadísticos para la variable «Artículos» NN-M1 / NN-M2 .....	184
Tabla 51. Estadísticos para la variable «Conjunciones» NN-M1 / NN-M2 .....	187
Tabla 52. Totales de las variables para las muestras «F» NN-M1-F / NN-M2-F / N-F .....	190
Tabla 53. Estadísticos para la variable «Tokens» NN-M1-F / NN-M2-F / N-F .....	193
Tabla 54. Estadísticos para la variable «Types» NN-M1-F / NN-M2-F / N-F .....	195

Tabla 55. Estadísticos para la variable «Verbos» NN-M1-F / NN-M2-F / N-F .....	195
Tabla 56. Estadísticos para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F .....	198
Tabla 57. Estadísticos para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F .....	201
Tabla 58. Estadísticos para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F .....	203
Tabla 59. Estadísticos para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F .....	206
Tabla 60. Estadísticos para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F .....	208
Tabla 61. Estadísticos para la variable «Artículos» NN-M1-F / NN-M2-F / N-F .....	211
Tabla 62. Estadísticos para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F .....	213
Tabla 63. Totales de las variables para las muestras «1D» NN-M1-1D / NN-M2-1D / N-1D .....	216
Tabla 64. Estadísticos para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D .....	217
Tabla 65. Estadísticos para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D .....	220
Tabla 66. Estadísticos para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D .....	222
Tabla 67. Estadísticos para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D .....	226
Tabla 68. Estadísticos para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D .....	227
Tabla 69. Estadísticos para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D .....	230
Tabla 70. Estadísticos para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D .....	233
Tabla 71. Estadísticos para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D .....	236
Tabla 72. Estadísticos para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D .....	239
Tabla 73. Estadísticos para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D .....	242
Tabla 74. Las 50 palabras más frecuentes. M-F / M-1D .....	247
Tabla 75. Las 50 palabras más frecuentes con lista de <i>stopwords</i> . M-F / M-1D .....	249
Tabla 76. Lista de agrupamientos del posesivo «mi» M-F / M-1D .....	250
Tabla 77. Lista de agrupamientos del posesivo átono «mis» M-F / M-1D .....	252
Tabla 78. Lista de agrupamientos del pronombre complemento «me» M-F / M-1D .....	252
Tabla 79. Lista de agrupamientos del pronombre «nos» M-F / M-1D .....	254
Tabla 80. Lista de agrupamientos del pronombre sujeto «yo» M-F / M-1D .....	255
Tabla 81. Lista de agrupamientos del sustantivo «día» M-F / M-1D .....	256
Tabla 82. Lista de agrupamientos del sustantivo «casa» M-F / M-1D .....	257

Tabla 83. Lista de agrupamientos del sustantivo «cosas» M-F / M-1D .....	258
Tabla 84. Lista de agrupamientos del sustantivo «amigos» M-F / M-1D .....	259
Tabla 85. Lista de agrupamientos del sustantivo «tiempo» M-F / M-1D .....	259
Tabla 86. Lista de agrupamientos del sustantivo «gente» M-F / M-1D .....	260
Tabla 87. <i>Output</i> de la herramienta <i>Concordance</i> para el adjetivo «mi» M-F / M-1D .....	262
Tabla 88. <i>Output</i> de la herramienta <i>Concordance</i> para el adjetivo «mis» M-F / M-1D .....	263
Tabla 89. <i>Output</i> de la herramienta <i>Concordance</i> para el pronombre «me» M-F / M-1D .....	264
Tabla 90. <i>Output</i> de la herramienta <i>Concordance</i> para el pronombre «nos» M-F / M-1D .....	265
Tabla 91. <i>Output</i> de la herramienta <i>Concordance</i> para el pronombre «yo» M-F / M-1D .....	266
Tabla 92. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «día» M-F / M-1D .....	267
Tabla 93. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «casa» M-F / M-1D .....	268
Tabla 94. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «cosas» M-F / M-1D .....	269
Tabla 95. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «amigos» M-F / M-1D .....	270
Tabla 96. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «tiempo» M-F / M-1D .....	271
Tabla 97. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «gente» M-F / M-1D .....	272
Tabla 98. <i>Output</i> de la herramienta <i>WordList..</i> NN – N .....	276
Tabla 99. <i>Output</i> de la herramienta <i>WordList</i> con <i>stopwords</i> NN – N .....	277
Tabla 100. Lista de agrupamientos del posesivo «mi» NN – N .....	278
Tabla 101. Lista de agrupamientos del pronombre complemento «me» NN – N .....	280
Tabla 102. Lista de agrupamientos del pronombre sujeto «yo» NN – N .....	281
Tabla 103. Lista de agrupamientos del sustantivo «casa» NN – N .....	282
Tabla 104. Lista de agrupamientos del sustantivo «momento» NN – N .....	283
Tabla 105. Lista de agrupamientos del sustantivo «hombre» NN – N .....	284
Tabla 106. Lista de agrupamientos del sustantivo «miedo» NN – N .....	285
Tabla 107. Lista de agrupamientos del sustantivo «tiempo» NN – N .....	286
Tabla 108. <i>Output</i> de la herramienta <i>Concordance</i> para el adjetivo «mi» NN – N .....	288
Tabla 109. <i>Output</i> de la herramienta <i>Concordance</i> para el pronombre «me» NN – N .....	289
Tabla 110. <i>Output</i> de la herramienta <i>Concordance</i> para el pronombre «yo» NN – N .....	289

Tabla 111. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «casa» NN – N .....	290
Tabla 112. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «momento» NN – N .....	291
Tabla 113. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «hombre» .....	292
Tabla 114. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «miedo» NN – N .....	293
Tabla 115. <i>Output</i> de la herramienta <i>Concordance</i> para el sustantivo «tiempo» NN – N .....	293
Tabla 116. Prueba de Shapiro-Wilk de <i>tokens</i> para el M1 y el M2 .....	298
Tabla 117. Totales de <i>Tokens</i> y <i>Types</i> M-1 /M-2 .....	300
Tabla 118. Prueba de Shapiro-Wilk del índice Uber U para el M1 y el M2 .....	300
Tabla 119. Totales y media de palabras léxicas y palabras funcionales .....	302
Tabla 120. Prueba de Shapiro-Wilk para PL y PF en el M1 y el M2 .....	303
Tabla 121. Totales de <i>Tokens</i> y <i>Types</i> del M1. Muestras «F» y «1D» .....	306
Tabla 122. Medias de <i>Tokens</i> y <i>Types</i> del M1. Muestras «F» y «1D» .....	307
Tabla 123. Prueba de <i>Shapiro-Wilk</i> para las muestras «F» y «1D» en el M1 .....	307
Tabla 124. Prueba de <i>Shapiro-Wilk</i> de las muestras «F» y «1D» para el M1 .....	308
Tabla 125. Totales de palabras léxicas y palabras funcionales para el M1 .....	310
Tabla 126. Media de palabras léxicas y palabras funcionales para el M1 .....	310
Tabla 127. Prueba de <i>Shapiro-Wilk</i> para la muestra «F» .....	311
Tabla 128. Prueba de <i>Shapiro-Wilk</i> para la muestra «1D» .....	311
Tabla 129. Totales de <i>Tokens</i> y <i>Types</i> de las muestras «F» y «1D» en el M2 .....	313
Tabla 130. Medias de <i>Tokens</i> y <i>Types</i> del M2. Muestras F y 1D .....	314
Tabla 131. Prueba de <i>Shapiro-Wilk</i> para la variable <i>Tokens</i> de las muestras «F» y «1D» en el M2 .....	314
Tabla 132. Prueba de <i>Shapiro-Wilk</i> para el M2. Muestras F y 1D .....	316
Tabla 133. Totales de palabras léxicas y palabras funcionales para el M2 .....	317
Tabla 134. Media de palabras léxicas y palabras funcionales para el M2 .....	318
Tabla 135. Prueba de <i>Shapiro-Wilk</i> para la muestra «F» M2 .....	318
Tabla 136. Prueba de <i>Shapiro-Wilk</i> para la muestra «1D» M2 .....	319
Tabla 137. Total de <i>tokens</i> de las muestras «F» y «1D» en el M1 y el M2 .....	321

Tabla 138. Media de <i>tokens</i> de las muestras «F» y «1D» en el M1 y el M2 .....	321
Tabla 139. Prueba de Shapiro-Wilk para las muestras «F» y «1D» en el M1 y el M2 .....	322
Tabla 140. Total de <i>tokens</i> y <i>types</i> de las muestras «F» y «1D» en el M1 y el M2 .....	325
Tabla 141. Índice Uber U para las muestras «F» y «1D» en el M1 y el M2 .....	325
Tabla 142. Prueba de <i>Shapiro-Wilk</i> de las muestras «F» y «1D» para el M1 y el M2 .....	325
Tabla 143. Total de PL de las muestras «F» y «1D» para el M1 y el M2 .....	330
Tabla 144. Media de PL de las muestras «F» y «1D» para el M1 y el M2 .....	330
Tabla 145. Prueba de <i>Shapiro-Wilk</i> para la muestra «F» y «1D» en el M1 y el M2 .....	331
Tabla 146. Número total de <i>tokens</i> por nivel de lengua .....	337
Tabla 147. Media de <i>tokens</i> por nivel de lengua .....	338
Tabla 148. Prueba de Shapiro-Wilk para <i>tokens</i> por nivel de lengua .....	339
Tabla 149. Número total de <i>tokens</i> y <i>types</i> de la muestra «F» por nivel de lengua .....	341
Tabla 150. Cálculo del índice Uber U para la muestra «F» por nivel de lengua .....	341
Tabla 151. Prueba de Shapiro-Wilk para índice Uber U de la muestra «F» por nivel de lengua .....	342
Tabla 152. Totales y medias de PL de la muestra «F» por nivel de lengua .....	344
Tabla 153. Prueba de Shapiro-Wilk para la densidad léxica de la muestra «F» por nivel de lengua.....	344
Tabla 154. Número total de <i>tokens</i> por nivel de lengua .....	346
Tabla 155. Media de <i>tokens</i> por nivel de lengua .....	347
Tabla 156. Prueba de Shapiro-Wilk para la fluidez de la muestra «1D» por nivel de lengua .....	347
Tabla 157. Número total de <i>tokens</i> y <i>types</i> de la muestra «1D» por nivel de lengua .....	349
Tabla 158. Cálculo del índice Uber U para la muestra «1D» por nivel de lengua .....	350
Tabla 159. Prueba de Shapiro-Wilk para la diversidad léxica de la muestra «1D» por nivel de lengua	351
Tabla 160. Totales y medias de PL de la muestra «1D» por nivel de lengua .....	353
Tabla 161. Prueba de Shapiro-Wilk para la densidad léxica de la muestra «1D» por nivel de lengua .....	354

## ÍNDICE DE GRÁFICOS

Gráfico 1. Género de los sujetos de la muestra en porcentaje. ....	68
Gráfico 2. Edad de los sujetos Nativos de la muestra. ....	68
Gráfico 3. Género de los sujetos NN de la muestra en porcentaje. ....	69
Gráfico 4. Edad de los sujetos NN de la muestra. ....	69
Gráfico 5. Distribución de número de sujetos por nacionalidad. ....	70
Gráfico 6. Nivel de estudios de los sujetos de la muestra en porcentaje. ....	71
Gráfico 7. Nivel de estudios de los sujetos de la muestra según el género en porcentaje. ....	71
Gráfico 8. Nivel de estudios de los sujetos de la muestra según el género en porcentaje. ....	72
Gráfico 9. Vía de acceso al nivel Intermedio B-1 en porcentaje. ....	73
Gráfico 10. Gráfico de dispersión para la variable «Tokens» grupos NN y N ....	106
Gráfico 11. Diagrama de Cajas para la variable «Tokens» grupos NN y N ....	106
Gráfico 12. Histograma para la variable «Tokens» grupos NN y N ....	107
Gráfico 13. Gráfico de dispersión para la variable «Types» NN – N ....	109
Gráfico 14. Diagrama de cajas para la variable «Types» NN – N ....	109
Gráfico 15. Histogramas para la variable «Types» NN – N ....	110
Gráfico 16. Gráfico de dispersión para la variable «Verbos» NN – N ....	111
Gráfico 17. Diagrama de cajas para la variable «Verbos» NN – N ....	112
Gráfico 18. Histograma para la variable «Verbos» NN – N ....	112
Gráfico 19. Gráfico de dispersión para la variable «Sustantivos» NN – N ....	114
Gráfico 20. Diagrama de cajas para la variable «Sustantivos» NN – N ....	114
Gráfico 21. Histograma para la variable «Sustantivos» NN – N ....	115
Gráfico 22. Gráfico de dispersión para la variable «Adjetivos» NN – N ....	116
Gráfico 23. Diagrama de Cajas para la variable «Adjetivos» NN – N ....	117
Gráfico 24. Histograma para la variable «Adjetivos» NN – N ....	118
Gráfico 25. Gráfico de dispersión para la variable «Adverbios» NN – N ....	119
Gráfico 26. Diagrama de Cajas para la variable «Adverbios» NN – N ....	120
Gráfico 27. Histograma para la variable «Adverbios» NN – N ....	120

Gráfico 28. Gráfico de dispersión para la variable «Pronombres» NN – N .....	122
Gráfico 29. Diagrama de Cajas para la variable «Pronombres» NN – N .....	122
Gráfico 30. Histograma para la variable «Pronombres» NN – N .....	123
Gráfico 31. Gráfico de dispersión para la variable «Preposiciones» NN – N .....	124
Gráfico 32. Diagrama de Cajas para la variable «Preposiciones» NN – N .....	125
Gráfico 33. Histograma para la variable «Preposiciones» NN – N .....	126
Gráfico 34. Gráfico de dispersión para la variable «Artículos» NN – N .....	127
Gráfico 35. Diagrama de Cajas para la variable «Artículos» NN – N .....	128
Gráfico 36. Histograma para la variable «Artículos» NN – N .....	128
Gráfico 37. Gráfico de dispersión para la variable «Conjunciones» NN – N .....	130
Gráfico 38. Diagrama de Cajas para la variable «Conjunciones» NN – N .....	130
Gráfico 39. Histograma para la variable «Conjunciones» NN – N .....	131
Gráfico 40. Gráfico de dispersión para la variable «Tokens» NN-1D / N-1D .....	133
Gráfico 41. Diagrama de Cajas para la variable «Tokens» NN-1D / N-1D .....	134
Gráfico 42. Histograma para la variable «Tokens» NN-1D / N-1D .....	135
Gráfico 43. Gráfico de dispersión para la variable «Types» NN-1D / N-1D .....	136
Gráfico 44. Diagrama de Cajas para la variable «Types» » NN-1D / N-1D .....	137
Gráfico 45. Histograma para la variable «Types» NN-1D / N-1D .....	138
Gráfico 46. Gráfico de dispersión para la variable «Verbos» NN-1D / N-1D .....	139
Gráfico 47. Diagrama de Cajas para la variable «Verbos» NN-1D / N-1D .....	140
Gráfico 48. Histograma para la variable «Verbos» NN-1D / N-1D .....	141
Gráfico 49. Gráfico de dispersión para la variable «Sustantivos» NN-1D / N-1D .....	142
Gráfico 50. Diagrama de Cajas para la variable «Sustantivos» NN-1D / N-1D .....	143
Gráfico 51. Histograma para la variable «Sustantivos» NN-1D / N-1D .....	143
Gráfico 52. Gráfico de dispersión para la variable «Adjetivos» NN-1D / N-1D .....	145
Gráfico 53. Diagrama de Cajas para la variable «Adjetivos» NN-1D / N-1D .....	145
Gráfico 54. Histograma para la variable «Adjetivos» NN-1D / N-1D .....	146
Gráfico 55. Gráfico de dispersión para la variable «Adverbios» NN-1D / N-1D .....	148

Gráfico 56. Diagrama de Cajas para la variable «Adverbios» NN-1D / N-1D .....	148
Gráfico 57. Histograma para la variable «Adverbios» NN-1D / N-1D .....	149
Gráfico 58. Gráfico de dispersión para la variable «Pronombres» NN-1D / N-1D .....	151
Gráfico 59. Diagrama de Cajas para la variable «Pronombres» NN-1D / N-1D .....	151
Gráfico 60. Histograma para la variable «Pronombres» NN-1D / N-1D .....	152
Gráfico 61. Gráfico de dispersión para la variable « Preposiciones» NN-1D / N-1D .....	154
Gráfico 62. Diagrama de Cajas para la variable « Preposiciones» NN-1D / N-1D .....	154
Gráfico 63. Histograma para la variable « Preposiciones» NN-1D / N-1D .....	155
Gráfico 64. Gráfico de dispersión para la variable « Artículos» NN-1D / N-1D .....	157
Gráfico 65. Diagrama de Cajas para la variable « Artículos» NN-1D / N-1D .....	157
Gráfico 66. Histograma para la variable «Artículos» NN-1D / N-1D .....	158
Gráfico 67. Gráfico de dispersión para la variable « Conjunciones» NN-1D / N-1D .....	160
Gráfico 68. Diagrama de Cajas para la variable « Conjunciones» NN-1D / N-1D .....	160
Gráfico 69. Histograma para la variable « Conjunciones» NN-1D / N-1D .....	161
Gráfico 70. Gráfico de dispersión para la variable «Tokens» NN-M1 / NN-M2 .....	163
Gráfico 71. Diagrama de Cajas para la variable «Tokens» NN-M1 / NN-M2 .....	164
Gráfico 72. Histograma para la variable «Tokens» NN-M1 / NN-M2 .....	164
Gráfico 73. Gráfico de dispersión para la variable «Types» NN-M1 / NN-M2 .....	166
Gráfico 74. Diagrama de Cajas para la variable «Types» NN-M1 / NN-M2 .....	166
Gráfico 75. Histograma para la variable «Types» NN-M1 / NN-M2 .....	167
Gráfico 76. Gráfico de dispersión para la variable «Verbos» NN-M1 / NN-M2 .....	169
Gráfico 77. Diagrama de Cajas para la variable «Verbos» NN-M1 / NN-M2 .....	169
Gráfico 78. Histograma para la variable «Verbos» NN-M1 / NN-M2 .....	170
Gráfico 79. Gráfico de dispersión para la variable «Sustantivos» NN-M1 / NN-M2 .....	171
Gráfico 80. Diagrama de Cajas para la variable «Sustantivos» NN-M1 / NN-M2 .....	172
Gráfico 81. Histograma para la variable «Sustantivos» NN-M1 / NN-M2 .....	173
Gráfico 82. Gráfico de dispersión para la variable «Adjetivos» NN-M1 / NN-M2 .....	174
Gráfico 83. Diagrama de Cajas para la variable «Adjetivos» NN-M1 / NN-M2 .....	175

Gráfico 84. Histograma para la variable «Adjetivos» NN-M1 / NN-M2 .....	175
Gráfico 85. Gráfico de dispersión para la variable «Adverbios» NN-M1 / NN-M2 .....	177
Gráfico 86. Diagrama de Cajas para la variable «Adverbios» NN-M1 / NN-M2 .....	177
Gráfico 87. Histograma para la variable «Adverbios» NN-M1 / NN-M2 .....	178
Gráfico 88. Gráfico de dispersión para la variable «Pronombres» NN-M1 / NN-M2 .....	179
Gráfico 89. Diagrama de Cajas para la variable «Pronombres» NN-M1 / NN-M2 .....	180
Gráfico 90. Histograma para la variable «Pronombres» NN-M1 / NN-M2 .....	181
Gráfico 91. Gráfico de dispersión para la variable «Preposiciones» NN-M1 / NN-M2 .....	182
Gráfico 92. Diagrama de Cajas para la variable «Preposiciones» NN-M1 / NN-M2 .....	183
Gráfico 93. Histograma para la variable «Preposiciones» NN-M1 / NN-M2 .....	183
Gráfico 94. Gráfico de dispersión para la variable «Artículos» NN-M1 / NN-M2 .....	185
Gráfico 95. Diagrama de Cajas para la variable «Artículos» NN-M1 / NN-M2 .....	185
Gráfico 96. Histograma para la variable «Artículos» NN-M1 / NN-M2 .....	186
Gráfico 97. Gráfico de dispersión para la variable «Conjunciones» NN-M1 / NN-M2 .....	187
Gráfico 98. Diagrama de Cajas para la variable «Conjunciones» NN-M1 / NN-M2 .....	188
Gráfico 99. Histograma para la variable «Conjunciones» NN-M1 / NN-M2 .....	181
Gráfico 100. Gráfico de dispersión para la variable «Tokens» NN-M1-F / NN-M2-F / N-F .....	189
Gráfico 101. Diagrama de Cajas para la variable «Tokens» NN-M1-F / NN-M2-F / N-F .....	191
Gráfico 102. Histograma para la variable «Tokens» NN-M1-F / NN-M2-F / N-F .....	192
Gráfico 103. Gráfico de dispersión para la variable «Types» NN-M1-F / NN-M2-F / N-F .....	193
Gráfico 104. Diagrama de Cajas para la variable «Types» NN-M1-F / NN-M2-F / N-F .....	194
Gráfico 105. Histograma para la variable «Types» NN-M1-F / NN-M2-F / N-F .....	195
Gráfico 106. Gráfico de dispersión para la variable «Verbos» NN-M1-F / NN-M2-F / N-F .....	196
Gráfico 107. Diagrama de cajas para la variable «Verbos» NN-M1-F / NN-M2-F / N-F .....	196
Gráfico 108. Histograma para la variable «Verbos» NN-M1-F / NN-M2-F / N-F .....	196
Gráfico 109. Gráfico de dispersión para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F .....	199
Gráfico 110. Diagrama de cajas para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F .....	199
Gráfico 111. Histograma para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F .....	200

Gráfico 112. Gráfico de dispersión para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F .....	202
Gráfico 113. Diagrama de cajas para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F .....	202
Gráfico 114. Histograma para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F .....	203
Gráfico 115. Gráfico de dispersión para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F .....	204
Gráfico 116. Diagrama de cajas para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F .....	204
Gráfico 117. Histograma para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F .....	205
Gráfico 118. Gráfico de dispersión para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F .....	206
Gráfico 119. Diagrama de cajas para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F .....	207
Gráfico 120. Histograma para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F .....	208
Gráfico 121. Gráfico de dispersión para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F .....	209
Gráfico 122. Diagrama de cajas para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F .....	210
Gráfico 123. Histograma para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F .....	211
Gráfico 124. Gráfico de dispersión para la variable «Artículos» NN-M1-F / NN-M2-F / N-F .....	212
Gráfico 125. Diagrama de cajas para la variable «Artículos» NN-M1-F / NN-M2-F / N-F .....	212
Gráfico 126. Histograma para la variable «Artículos» NN-M1-F / NN-M2-F / N-F .....	212
Gráfico 127. Gráfico de dispersión para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F .....	214
Gráfico 128. Diagrama de cajas para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F .....	215
Gráfico 129. Histograma para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F .....	215
Gráfico 130. Gráfico de dispersión para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D .....	218
Gráfico 131. Diagrama de cajas para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D .....	218
Gráfico 132. Histograma para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D .....	219
Gráfico 133. Gráfico de dispersión para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D .....	221
Gráfico 134. Diagrama de cajas para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D .....	221
Gráfico 135. Histograma para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D .....	221
Gráfico 136. Gráfico de dispersión para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D .....	223
Gráfico 137. Diagrama de cajas para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D .....	223
Gráfico 138. Histograma para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D .....	224
Gráfico 139. Gráfico de dispersión para la variable «Sustantivos» NN-M1-1D/NN-M2-1D N-1D .....	225

Gráfico 140. Diagrama de cajas para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D .....	226
Gráfico 141. Histograma para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D .....	227
Gráfico 142. Gráfico de dispersión para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D .....	228
Gráfico 143. Diagrama de cajas para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D .....	229
Gráfico 144. Histograma para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D .....	229
Gráfico 145. Gráfico de dispersión para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D .....	231
Gráfico 146. Diagrama de cajas para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D.....	232
Gráfico 147. Histograma para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D .....	232
Gráfico 148. Gráfico de dispersión para la variable «Pronombres» NN-M1-1D/NN-M2-1D/N-1D.....	234
Gráfico 149. Diagrama de cajas para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D .....	235
Gráfico 150. Histograma para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D .....	235
Gráfico 151. Gráfico de dispersión para la variable «Preposiciones» NN-M1-1D/NN-M2-1D/N-1D.....	237
Gráfico 152. Diagrama de cajas para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D.....	237
Gráfico 153. Histograma para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D .....	238
Gráfico 154. Gráfico de dispersión para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D .....	240
Gráfico 155. Diagrama de cajas para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D .....	240
Gráfico 156. Histograma para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D .....	241
Gráfico 157. Gráfico de dispersión para la variable «Conjunciones» NN-M1-1D/NN-M2-1D/N-1D.....	243
Gráfico 158. Diagrama de cajas para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D.....	243
Gráfico 159. Histograma para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D .....	244
Gráfico 160. Totales del M1 y del M2 .....	297
Gráfico 161. Media de <i>Tokens</i> y <i>Types</i> . M1 M2 .....	298
Gráfico 162. Media de palabras léxicas y palabras funcionales .....	302
Gráfico 163. Totales de palabras léxicas y palabras funcionales .....	302
Gráfico 164. Media de palabras léxicas y palabras funcionales .....	303
Gráfico 165. Totales de palabras léxicas y palabras funcionales para el M1 .....	310
Gráfico 166. Media de palabras léxicas y palabras funcionales para el M1 .....	310
Gráfico 167. Totales de <i>Tokens</i> y <i>Types</i> de las muestras «F» y «1D» en el M2 .....	313

Gráfico 168. Medias de <i>Tokens</i> y <i>Types</i> de las muestras «F» y «1D» en el M2 .....	314
Gráfico 169. Totales de palabras léxicas y palabras funcionales para el M2 .....	317
Gráfico 170. Media de palabras léxicas y palabras funcionales para el M2 .....	318
Gráfico 171. Total de <i>tokens</i> de las muestras «F» y «1D» en el M1 y el M2 .....	321
Gráfico 172. Media de <i>tokens</i> de las muestras «F» y «1D» en el M1 y el M2 .....	322
Gráfico 173. Total de PL de las muestras «F» y «1D» para el M1 y el M2 .....	330
Gráfico 174. Media de PL de las muestras «F» y «1D» para el M1 y el M2 .....	330
Gráfico 175. Total de <i>tokens</i> por nivel de lengua .....	338
Gráfico 176. Media de <i>tokens</i> por nivel de lengua .....	338
Gráfico 177. Total de <i>tokens</i> y <i>types</i> de la muestra «F» por estadios por nivel de lengua .....	341
Gráfico 178. Totales y medias de PL de la muestra «F» por nivel de lengua .....	344
Gráfico 179. Totales y medias de <i>Tokens</i> de la muestra «1D» por nivel de lengua .....	347
Gráfico 180. Totales de <i>Tokens</i> y <i>Types</i> de la muestra «1D» por nivel de lengua .....	350
Gráfico 181. Totales y medias de PL de la muestra «1D» por nivel de lengua .....	353



## ÍNDICE FIGURAS

Figura 1. Competencia lingüística. Relaciones jerárquicas Fuente: Adaptado de Bachman (1990) .....	39
Figura 2. Medida de diversidad léxica Uber U .....	65
Figura 3. Ejemplo de <i>output</i> de la herramienta <i>Wordlist</i> . .....	81
Figura 4. Ejemplo de <i>output</i> de la herramienta <i>Concordance</i> . .....	82
Figura 5. Ejemplo de <i>output</i> de la herramienta <i>Clusters</i> . .....	82
Figura 6. Ejemplo de <i>output</i> de la herramienta Métrica. ....	83
Figura 7. Ejemplo de <i>output</i> de la herramienta Morfología. ....	84
Figura 8. Ejemplo de <i>output</i> de la herramienta Vocabulario. ....	84
Figura 9. Ejemplo del software G-Stat 2.0. ....	90
Figura 10. Ejemplo de <i>output</i> de cálculo de ajuste a la normalidad .....	90
Figura 11. Ejemplo de <i>output</i> de cálculo igualdad de varianzas .....	91
Figura 12. Ejemplo de <i>output</i> de cálculo igualdad de varianza .....	91
Figura 13. Ejemplo de <i>output</i> de estadísticos .....	92
Figura 14: Ejemplo de «diagrama de caja» .....	94
Figura 15. Ejemplo de la prueba t-Student .....	96
Figura 16. Ejemplo de <i>output</i> de la prueba t-Student .....	96
Figura 17. Ejemplo de <i>output</i> de la prueba de Mann-Whitney .....	97
Figura 18. Ejemplo de <i>output</i> de la prueba de Mann-Whitney .....	97
Figura 19. Ejemplo de la prueba ANOVA .....	98
Figura 20. Ejemplo de <i>output</i> de la prueba ANOVA .....	98
Figura 21. Ejemplo de <i>output</i> de la prueba ANOVA .....	99
Figura 22. Ejemplo de <i>output</i> de la prueba ANOVA- Comparaciones múltiples .....	99
Figura 23. Ejemplo de la prueba de <i>Kruskal-Wallis</i> .....	100
Figura 24. Ejemplo de <i>output</i> de la prueba de <i>Kruskal-Wallis</i> 94 .....	101
Figura 25. Ejemplo de <i>output</i> de la prueba de <i>Kruskal-Wallis</i> - Comparaciones múltiples .....	101
Figura 26. Test de Mann-Whitney (Wilcoxon) .....	299

Figura 27. Prueba F-Snedecor .....	300
Figura 28. Prueba t-Student de varianzas iguales .....	301
Figura 29. Densidad léxica en la muestra del M1 .....	304
Figura 30. Densidad léxica en la muestra del M2 .....	304
Figura 31. Densidad léxica entre la muestra del M1 y del M2 .....	305
Figura 32. Prueba de F-Snedecor .....	307
Figura 33. Prueba T-Student para varianzas diferentes .....	308
Figura 34. Prueba de F-Snedecor .....	309
Figura 35. Prueba de t-Student para varianzas iguales .....	309
Figura 36. Prueba U Mann-Whitney para la muestra «F» .....	312
Figura 37. Prueba U Mann-Whitney para la muestra «1D» .....	312
Figura 38. Prueba U Mann-Whitney para las muestras «F» y «1D» del M1 .....	313
Figura 39. Prueba de F-Snedecor para el contraste de varianzas .....	315
Figura 40. Prueba T-Student para varianzas diferentes .....	315
Figura 41. Prueba de F-Snedecor para el contraste de varianzas .....	316
Figura 42. Prueba de t-Student para varianzas iguales .....	316
Figura 43. Prueba U Mann-Whitney para la muestra «F» .....	319
Figura 44. Prueba U Mann-Whitney para la muestra «1D» .....	319
Figura 45. Prueba U Mann-Whitney para las muestras «F» y «1D» del M2 .....	320
Figura 46. Prueba Anova para fluidez de las muestras «F» y «1D» en el M1 y el M2 .....	323
Figura 47. Prueba de Comparaciones múltiples para la fluidez de las muestras en el M1 y el M2 .....	324
Figura 48. Prueba Anova para la diversidad léxica de las muestras «F» y «1D» en el M1 y el M2 .....	326
Figura 49. Prueba: Comparaciones múltiples para la diversidad léxica de las muestras en el M1 y el M2.....	327
Figura 50. Prueba F-Snedecor: contraste de varianzas de la muestra «F» en el M1 y el M2 .....	328
Figura 51. Prueba t-Student de varianzas iguales para la muestra «F» entre el M1 y el M2 .....	328
Figura 52. Prueba F-Snedecor: contraste de varianzas de la muestra «1D» en el M1 y el M2 .....	329
Figura 53. Prueba t-Student de varianzas iguales para la muestra «1D» entre el M1 y el M2 .....	329

Figura 54. Prueba Kruskal-Wallis para la densidad léxica de las muestras «F» y «1D» en el M1 y el M2	.331
Figura 55. Prueba: Comparaciones múltiples para la densidad léxica de las muestras «F» y «1D» en el M1 y el M2	.....333
Figura 56. Prueba Mann-Whitney para la densidad léxica de la muestra «F» entre el M1 y el M2	.....333
Figura 57. Prueba Mann-Whitney para la densidad léxica de la muestra «1D» entre el M1 y el M2	.....334
Figura 58. Test Anova para la fluidez de las muestras NN-M1-F, NN-M2-F, N-F	.....339
Figura 59. Test Anova-«comparaciones múltiples» para la fluidez de las muestras NN-M1-F, NN-M2-F, N-F	.....340
Figura 60. Test Kruskal-Wallis para la diversidad léxica de las muestras NN-M1-F, NN-M2-F, N-F	.....342
Figura 61. Test Kruskal-Wallis-«comparaciones múltiples» para la diversidad léxica de las muestras NN-M1-F, NN-M2-F, N-F	.....343
Figura 62. Test Kruskal-Wallis para la densidad léxica de las muestras NN-M1-F, NN-M2-F, N-F	.....345
Figura 63. Test Kruskal-Wallis-«comparaciones múltiples» para la densidad léxica de las muestras NN-M1-F, NN-M2-F, N-F	.....346
Figura 64. Test Anova para la fluidez de las muestras NN-M1-1D, NN-M2-1D, N-1D	.....348
Figura 65. Test Anova-«comparaciones múltiples» para la fluidez de las muestras NN-M1-1D, NN-M2-1D, N-1D	.....349
Figura 66. Test Kruskal-Wallis para la diversidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D	.351
Figura 67. Test Kruskal-Wallis-«comparaciones múltiples» para la diversidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D	.....352
Figura 68. Test Kruskal-Wallis para la densidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D	..354
Figura 69. Test Kruskal-Wallis-«comparaciones múltiples» para la densidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D	.....355



## INTRODUCCIÓN

Habitualmente los productos escritos en una L2, ya sean de traducción o composición, han sido objeto de estudio y análisis en el marco de la lingüística contrastiva y el análisis de errores, por ejemplo. Pero a partir de la década de los 70 se llamó la atención sobre (i) la ausencia de estudios centrados en el carácter comunicativo del uso de la lengua, (ii) la ausencia de estudios centrados en las características textuales propias o (iii) la inexistencia de estudios sobre procesos de escritura y procesos seguidos por los escritores en L1 y L2 en sí.

En los estudios sobre procesos de escritura, ha resultado muy productivo la aplicación de metodologías de enseñanza como el enfoque por tareas, que han permitido superar la artificialidad comunicativa de la mayoría de actividades de producción escrita en metodologías anteriores. El uso de técnicas como el '*focus on form*', por ejemplo, han permitido llevar a cabo interesantes estudios de investigación-acción que han arrojado mucha luz sobre los procesos de escritura y las posibilidades de mejora de la competencia con técnicas procedentes de la psicolingüística aplicada a la enseñanza de lenguas. A partir de estas técnicas hemos centrado la atención en cómo se desarrolla la L2 (desde la perspectiva de la adquisición) y atendiendo a lo que consideramos que es la competencia: una suma de subcompetencias (léxica, sintáctica, discursiva), que son las que vamos a medir en el trabajo mediante los indicadores seleccionados y con las herramientas elegidas para cada caso. En cuanto a las herramientas, hemos adoptado la perspectiva práctica del profesor-investigador y hemos elegido las que se encuentran disponibles en la red, de libre acceso, cambiando y combinando recursos y ámbito de aplicación en un ejercicio de poner puertas al campo y obtener una imagen de la destreza objeto de estudio en sus dos momentos de cata.

En este sentido, creemos que lo novedoso de esta investigación es la aproximación a la destreza escrita en español L2 y a la competencia escritora dentro un contexto no académico, como puede ser la universidad. Destacar también que se ha llevado a cabo un estudio longitudinal con el que ver, precisamente, el desarrollo de esta competencia como un componente más en un curso de lengua extranjera. Es, además, particularmente novedoso abordar el estudio de este desarrollo de la competencia escrita **no en un curso ni un taller de escritura** específicos, sino en un ambiente de enseñanza general. Creemos que haber elegido este contexto resulta particularmente revelador para ver cómo se desarrolla la

competencia escritora ‘no experta’ de alumnos de L2 y compararla con la producción de nativos que tampoco son expertos.

A continuación, pasamos a detallar cómo se estructura la tesis. La hemos dividido en cuatro partes precedidas de la ‘introducción’ y la ‘justificación’. Cada parte está dividida en capítulos.

La parte I está formada por el capítulo uno dedicado al ‘estado de la cuestión’. En este capítulo hacemos una aproximación, en primer lugar, a las teorías de ASL (Adquisición de Segundas Lenguas) para ofrecer unas pinceladas sobre las tres teorías más destacadas del siglo XX, como son las ‘innatistas’, las ‘medioambientales’ o las ‘de interacción’. En la sección dos de este capítulo abordamos el tema de la investigación sobre ‘proceso, agentes y estrategias en L2’. Con la intención de delimitar el tema, en la siguiente sección focalizamos en la expresión escrita en el ámbito de L2, y presentamos algunos de los trabajos que se están llevando a cabo sobre estudios del ‘proceso’ de escribir, otros, como el de Sasaki (2009) interesada en las estancias en el exterior o el de Verspoor y Smiskova, que se interesan en las ‘chunks’ como medida para estudiar el desarrollo del producto escrito. En la sección siguiente, hacemos un recorrido sobre las investigaciones que han focalizado en el desarrollo de la competencia escritora en español. Y encontramos trabajos tanto cuantitativos como cualitativos y que abordan el tema desde diferentes perspectivas, desde el proceso como también desde el producto final. También dedicamos una sección a la ‘lingüística de corpus’ para explicar la relevancia de trabajar con este tipo de material. La sección 1.6 la dedicamos para contextualizar qué se entiende hoy en día sobre ‘ser nativo’ de una lengua y, por tanto, qué lengua enseñar en el aula de ELE. Finalmente, presentamos la sección dedicada a la ‘enseñanza mediante tareas’ y cómo se ha ido desarrollando este concepto desde que naciera en la década de los ochenta hasta nuestros días.

La parte II está formada por los capítulos dos y tres. En el primero desarrollamos los ‘objetivos, preguntas de investigación e hipótesis’. El capítulo tres está dedicado al marco metodológico en el que nos hemos basado en esta tesis. En él hablamos de la metodología cuantitativa; de qué se entiende por investigación pseudolongitudinal, y de las medidas de calidad general del texto (fluidez, diversidad léxica y densidad léxica).

En la parte III, compuesta por los capítulos cuatro y cinco, focalizamos en nuestro ‘estudio’ y, así, en el capítulo cuatro hablamos del contexto (participantes e instituciones), así como del corpus y de cómo se llevó a cabo la compilación de datos. El capítulo cinco lo

dedicamos a explicar la metodología de análisis seguida en la investigación, así como el software informático (Paramtext; AntConc) que hemos utilizado para ayudarnos en el análisis a nivel lingüístico, como para realizar a posteriori el análisis estadístico (G-Stat 2.0).

En la parte IV agrupamos los capítulos seis y siete. Estos capítulos están dedicados a la presentación de los *resultados y análisis*, capítulo seis, y a la discusión de los mismos en el capítulo 7. En el capítulo seis dedicamos una sección a presentar el análisis descriptivo de los datos, mientras que en la segunda sección vamos dando respuesta a las preguntas de investigación a partir de realizar un tratamiento estadístico inferencial.

Finalizamos la tesis con el capítulo ocho dedicado a las conclusiones a las que hemos llegado. Tomando las hipótesis de partida como punto de referencia, vamos viendo con la ayuda de los resultados y análisis del capítulo siete si se confirman o no.

Cierra la tesis las ‘referencias bibliográficas’ y los ‘anexos’.



## JUSTIFICACIÓN

El estudio de la competencia escritora dentro del ámbito de la Adquisición de Segundas Lenguas, y en particular en el mundo de ELE (Español como Lengua Extranjera) ha supuesto un hándicap en el campo de la didáctica, tanto para alumnos como para docentes. De Haan y van Esch (2005: 101) dicen sobre la escritura en España: “In Spanish schools there is no real training in writing and students are not taught to compose, not even in their mother tongue (Murphy, personal communication).” Así, no es extraño encontrar a estudiantes quejándose de las actividades de escritura que les mandan realizar, como a docentes quejarse por los ‘malos’ resultados.

Dentro de este contexto, nace nuestro interés por investigar el desarrollo de la competencia lingüística focalizando en la destreza de la expresión escrita en el ámbito de ELE. El objetivo es llegar a comprender mejor el complejo entramado que es el campo de la competencia escritora. Consideramos que si llegamos a poder entender mejor este proceso, será más fácil poder ofrecer al alumno también una mejor práctica.

Con ese objetivo en mente, mejorar las prácticas de docentes y alumnos en torno a la producción escrita, decidimos iniciar esta investigación. De entre las distintas maneras de abordar el tema del progreso de la competencia escritora, decidimos centrarnos (i) en el análisis de la fluidez, la diversidad léxica y la densidad léxica en dos tipos de producciones escritas por estudiantes de español como L2 (lengua dos) de nivel B1 (según el MCER) en dos momentos de su proceso de aprendizaje (inicio y final de curso) con el objetivo de comprobar si se producía progreso durante el período de su instrucción académica; (ii) así como en el estudio de listas de frecuencias, agrupamientos léxicos y concordancias con la finalidad de caracterizar géneros textuales. Asimismo, creímos relevante observar las posibles similitudes o diferencias en los resultados de las producciones escritas por los sujetos no nativos y nativos. Identificar el léxico o las estructuras que utiliza cada uno de los grupos a la hora de elaborar dos géneros textuales diferentes y comparar esos resultados, puede ayudar a trabajar con *input* real dentro del aula de ELE.

Por todo ello, la relevancia de este trabajo de investigación radica (i) en estudiar los temas ya mencionados a partir de un corpus de aprendices formado por producciones escritas de estudiantes de español L1 y L2, (ii) en que la muestra correspondiente al grupo de L2 español corresponde a un grupo multicultural y plurilingüe (algo poco habitual en la mayoría de las investigaciones de este tipo), (iii) en que la muestra correspondiente al grupo de L2

español seguía un curso general de ELE y no específico, es decir, no centrado en el tratamiento de la expresión escrita, (iv) en que se trata de un estudio pseudolongitudinal (normalmente son investigaciones transversales) y (v) en que se trabaja con software informático de acceso libre tanto para el análisis lingüístico como el estadístico (esto facilita que estudiantes o investigadores de lingüística aplicada u otras áreas afines y/o docentes de L2/LE puedan realizar sus propios trabajos de investigación).

Así, el primer germen de la presente tesis lo encontramos en la memoria final que la doctoranda realizó para completar el Máster de Formación de Profesores de ELE de la Universidad de Barcelona, que fue evolucionando en el DEA y ha culminado con la realización de este trabajo de investigación.

Durante este tiempo fuimos seleccionando y delimitando los temas a analizar; decidimos centrarnos, finalmente, en el estudio de la calidad general del texto y en el análisis de listas de frecuencias, agrupamientos y concordancias. De ahí que nuestro objetivo general sea: “investigar la fluidez, la diversidad léxica y la densidad léxica en dos tipos de producciones escritas por estudiantes de español como L2 (lengua dos) de nivel B1 (según el MCER) en dos momentos de su proceso de aprendizaje (inicio y final de curso) y compararlas con el estadio último de competencia nativa (grupo de control, grupo nativo).” Con ello, nos dispusimos a extraer las características principales de los dos tipos de expresión escrita que los sujetos nativos y no nativos produjeron, (a) inventar una historia y (b) explicar una experiencia personal. Asimismo, analizamos si los sujetos nativos y no nativos utilizaban el mismo tipo de léxico o no en función del tipo de texto que debían escribir. También nos interesaba indagar si en los sujetos no nativos había progreso en las producciones escritas en cuanto a la calidad general del texto entre los dos momentos de su proceso de aprendizaje (inicio y final de curso). Otro aspecto que analizamos fue si el texto en el que se siguió una instrucción específica con enseñanza mediante tareas presentaba mejores resultados de calidad del texto que en los que no se realizó instrucción específica. Finalmente, comparamos los resultados de las producciones de los sujetos no nativos y los nativos para comprobar si había similitudes o diferencias entre ellos.

De esta manera, pretendemos aportar al ámbito de la investigación y al de la didáctica de ELE nuestros hallazgos en el análisis de la calidad del texto y del comportamiento de las listas de frecuencias, agrupamientos y concordancias dependiendo del género textual y de si se trata de un ‘escritor’ nativo o no nativo.

# PARTE I

## ESTADO DE LA CUESTIÓN

### CAPÍTULO 1

#### ESTADO DE LA CUESTIÓN

##### 1.1 Antecedentes: teorías en Adquisición de Segundas Lenguas

Los estudios realizados en torno a las diferentes teorías y enfoques de la ASL han sido bastantes a lo largo del siglo XX. En palabras de Larsen-Freeman y Long (1991) el número de teorías propuestas sobre ASL han sido no menos de 40. Pero para ayudar a moverse por el intrincado mundo de la ASL en los últimos años se han publicado varios manuales en los que se recopilan trabajos de los investigadores más relevantes de dicho ámbito. Uno de ellos, por ejemplo, es el publicado por Doughty y Long (2003), *The Handbook of Second Language Acquisition*. En él se reúne a los teóricos más notables en el ámbito de la investigación sobre ASL, convirtiéndolo en un manual de referencia. A lo largo del libro se trata el tema desde la perspectiva biológica y cognitiva y se abarca, asimismo, temas como (i) los procesos que se producen en la ASL (DeKeyser, Hulstijn o Long); (ii) métodos de investigación (Norris y Ortega, Chaudron) o (iii) las limitaciones biológicas y psicológicas en la adquisición de segundas lenguas (Dörnyei y Skehan, Robinson o Pienemann). Otros manuales son los publicados por Peter Robinson y Nick C. Ellis (2008) y, más recientemente, por Julia Herschensohn y Martha Young-Scholten (2013). En este último, las editoras han reunido a autores jóvenes que tratan a lo largo de 31 capítulos seis temas básicos en el ámbito de la ASL como: (i) teoría y práctica, (ii) factores internos, (iii) factores externos, (iv) factores biológicos, (v) propiedades de los sistemas de interlengua y (vi) modelos de desarrollo. También encontramos publicaciones en las que se hace un recorrido histórico como el que presenta Rod Ellis (2005) y Rod Ellis y Natsuko Shintani (2014). También investigadores españoles se han preocupado por el tema de la ASL. Ya en 1988 y 1989, Javier Zanón publica en la revista *Cable* un trabajo en el que hace un breve recorrido por las teorías de ASL. Juan Mayor (1994), dentro del congreso de ASELE, presenta una visión algo más amplia de las distintas teorías de ASL. Francisco Moreno (2007), por su lado, realiza una aproximación al tema desde la perspectiva sociolingüística y muestra un breve recorrido sobre las teorías interaccionistas, ambientalistas y variacionistas.

Aquí vamos a presentar brevemente la clasificación que Larsen-Freeman y Long realizan en el año 1991 sobre las distintas teorías que se han propuesto a lo largo del siglo XX. Estos investigadores las dividen en tres grandes grupos, delimitando, de este modo, un campo tan amplio y controvertido como es el de ASL. Estos tres grupos son:

Teorías de ASL
Teorías innatistas
Teorías medioambientales
Teorías interaccionistas

Tabla 1: Clasificación Teorías de ASL Fuente: Larsen-Freeman y Long (1991)

### 1.1.1 Teorías innatistas

Las teorías innatistas parten de la idea de que la adquisición de la lengua se genera gracias a una dotación biológica innata que hace posible su aprendizaje.

Chomsky, uno de los representantes más conocidos de esta teoría y de los primeros en difundirla, habla en la década de los 60 de la existencia de una gramática universal, es decir, de un conocimiento innato de universales básicos como por ejemplo las categorías sintácticas o los rasgos fonológicos distintivos. Este carácter universal es el que hace que la adquisición de lenguas, tanto L1 como L2 o LE, siga un proceso similar (Baralo, 1996; MacWhinney, 2005). Así, la autora hace hincapié en el hecho de que los niños “cometen los mismos ‘errores’ y los superan en la misma etapa, independientemente de que convivan con adultos que les corrijan o no les corrijan.” (Baralo, 1996:124). También MacWhinney (*op. cit.*) ejemplifica esta similitud explicando que cuando se aprenden nuevas palabras en L2/LE se suele utilizar las mismas estrategias que cuando aprendemos nuevas palabras en L1. Es por ello que Marta Baralo considera, como ya apuntaba Chomsky, que existe una gramática universal compuesta por una serie de elementos con la información básica e imprescindible para que se produzca la adquisición de una lengua. Esos elementos, que según la autora son universales e innatos, poseen una información que no está incluida en el *input* que reciben los niños. Baralo parte de esta idea para intentar explicar cómo se realiza la adquisición de una L2 o LE y de qué manera se genera la interlengua del aprendiente:

hay una infinitud de frases posibles y el hablante de L2 sólo tiene acceso a un *input* muy reducido, pero aun así, hace generalizaciones y construye un sistema de conocimiento más allá de los datos del *input*. El hablante de LE no se limita a repetir e imitar lo que escucha, construye enunciados propios, característicos de la interlengua (IL). (Baralo, 1996:124)

No obstante, la autora asegura que, aun habiendo diferencias claras en el proceso de adquisición de una L1 y una L2/LE, también existe una semejanza que se ve reforzada por el hecho de que en ambos casos hay cierta creatividad en las producciones de un niño adquiriendo su L1 y en la de los aprendientes de L2/LE.

Siguiendo la línea innatista, en los años 80 ve la luz la Hipótesis del *input* expuesta por Krashen en 1985. Según esta teoría, la adquisición de una L2/LE se produce gracias a un proceso natural a través del suministro constante de *input* comprensible (caudal lingüístico comprensible). Para favorecer este proceso de adquisición, el *input* debe poseer rasgos gramaticales que estén en un nivel algo superior al nivel de interlengua del aprendiente (I+1). Así, según Krashen, la adquisición se produce mediante una exposición continua a un *input* comprensible y gracias a un filtro afectivo bajo. Solo de esta manera, insiste el autor, el aprendiente podrá desarrollar su competencia lingüística. En este contexto, el *output* se ve como una consecuencia de la adquisición y no como causa. (Krashen, 1985; McLaughlin, 1991; Ellis, 2001; Ellis y Shintani, 2014).

Con anterioridad a la hipótesis del *input*, Krashen ya había presentado la ‘Hipótesis del monitor’. Para explicar su teoría, el autor señala que cuando un adulto aprende una L2/LE puede seguir dos caminos:

- (i) El de la adquisición, es decir, se accede a la lengua de modo similar a como lo hacen los niños, de una manera subconsciente y natural. En esta misma línea MacWhinney (2005) explica que cuando en L2/LE se desea crear una oración nueva, por ejemplo, se recurre al mismo tipo de estrategias a las que recurren los niños cuando aprenden su lengua materna.
- (ii) El del aprendizaje; en esta modalidad se accede a la lengua meta de forma consciente y habitualmente en contextos formales. Partiendo de esta distinción, el autor justifica el hecho de que “la producción de los adultos en L2 es posible gracias al sistema adquirido, actuando el aprendizaje sólo como un monitor que, cuando las condiciones lo permiten, inspecciona, y a veces altera, el educto del sistema adquirido.” (Krashen, 1977:144-145). Eso significa que para que el monitor pueda actuar, el aprendiente tiene que conocer las reglas de la lengua meta, es lo que ha estudiado en un contexto formal, y tiene que tener tiempo para pensar, reflexionar sobre lo que está produciendo. Es decir, cuando la tarea que realiza le permite reflexionar y percibir que ha cometido un error es cuando el ‘monitor’ hace su función. Krashen señala las tareas de expresión escrita como las que tienen más posibilidades de que el monitor actúe ya que el aprendiente tiene tiempo para repasar lo que está escribiendo. Pero también indica que en una conversación lenta el monitor puede funcionar igualmente (Krashen, 1977).

### ***1.1.2 Teorías medioambientales***

Los investigadores que siguen las teorías medioambientales consideran que en el proceso de aprendizaje de una segunda lengua es más relevante el entorno en el que se desenvuelve el aprendiente y las experiencias vividas que la capacidad innata que posee cada individuo, como defendían las teorías innatistas. Es decir, que factores psicológicos de carácter afectivo como la motivación o la autoestima pueden influir en que se produzca un resultado positivo o negativo a la hora de aprender una lengua diferente a la lengua materna (Larsen-Freeman y Long, 1991; Alcón, 2002).

Las teorías más destacadas dentro de esta corriente las encontramos en la «Hipótesis de Pidginización» y en el «Modelo de Asimilación Cultural de Schumann». En dichas teorías, adquiere gran importancia las actitudes sociológicas y psicológicas tanto del grupo como del individuo ante la sociedad y su lengua meta.

Schumann (1978) defiende en su teoría que cuando existe una distancia psicológica y social entre el individuo y la cultura de la lengua meta el sujeto no puede procesar el aprendizaje de la L2 y no es capaz de superar los primeros estadios de la misma. Es decir, que su competencia lingüística se queda en lo que él llama *pidginización*, con un uso de estructuras gramaticales simples, sin complejidad.

Schumann (*op. cit.*) desarrolla esta teoría para dar respuesta al proceso de adquisición de una L2 por parte de emigrantes. El autor explica que el grado de desarrollo y el nivel final alcanzado por el aprendiente dependen de la capacidad de adaptación que tenga a la cultura de la lengua meta. Es lo que él llama ‘aculturación’. Así, cuanto mayor sea el grado de aculturación, mayor será el nivel final de la L2 que alcance el aprendiente.

Para Schumann (1986) existen dos tipos de ‘aculturación’: (i) en el primer tipo los sujetos están dispuestos no solo a integrarse en la cultura de la lengua meta, sino también a adoptar sus costumbres y valores; (ii) mientras que en el segundo se produce integración social, pero no hay interés en ir más allá, es decir, en adoptar, como sí hace el primer grupo, las costumbres y valores de la sociedad de la lengua meta. A pesar de esa diferencia, el autor considera que en ambos casos se consigue la integración social así como la adquisición de la L2, aunque, como se ha dicho más arriba, el nivel final alcanzado sea diferente.

Pero para Schumann (1978) hay también dos factores relevantes que van a influir en el proceso de aprendizaje de la L2 y que explican las diferencias existentes en los resultados finales del nivel alcanzado por cada individuo. Estos factores son los sociales y los

psicológicos y determinan la distancia social y psicológica entre el aprendiente y el grupo de la lengua meta. Con ‘distancia social’ se refiere a la posición del aprendiente, como miembro de un grupo, respecto al grupo de la lengua meta. Mientras que en ‘distancia psicológica’ agrupa cuatro elementos de carácter individual e índole emocional: (i) choque lingüístico, (ii) choque cultural, (iii) motivación y (iv) permeabilidad del ego.

En cuanto a la distancia social, Schumann (*op. cit.*) enumera ocho factores:

- (i) **Dominio social:** el tipo de relación que se establece entre el grupo del aprendiente y el grupo de la lengua meta puede ser (a) de dominio, (b) de subordinación o (c) neutra. Es decir, si la lengua de uno de los dos grupos es política, cultural, económica o técnicamente superior, la distancia social será mayor y, por tanto, la adquisición de la L2 resultará más difícil. En cambio, si la distancia es menor, hay una relación neutra, el contacto entre ambos grupos aumentará y eso favorecerá la adquisición de la lengua meta.
- (ii) **Patrón de integración:** es la manera de posicionarse ante el grupo de la lengua meta: (a) conservando sus costumbres y valores; (b) adoptando las costumbres y valores del grupo de la lengua meta, y (c) adoptando las costumbres y valores del grupo de la lengua meta al mismo tiempo que conservan sus costumbres y valores. En el primer caso, la adquisición de la L2 se dificultará, mientras que en los casos dos y tres se facilita.
- (iii) **Recinto:** hace referencia a los espacios sociales como escuelas, iglesias, lugares de trabajo, etc. Si estos espacios se comparten por parte de los dos grupos, la adquisición de la L2 se facilita.
- (iv) **Cohesión:** si el grupo del aprendiente es un grupo cohesionado y se mantiene separado del grupo de la lengua meta, las posibilidades de adquisición de la L2 se reducen.
- (v) **Tamaño:** si el tamaño del grupo social del aprendiente no es muy grande, más posibilidades de contacto habrá con el grupo de la lengua meta y, por tanto, la adquisición se favorecerá.
- (vi) **Congruencia:** si las dos culturas son similares, es más probable que se produzca contacto social y, de esta manera, el aprendizaje de L2 se facilita.

- (vii) **Actitud:** cuando la actitud del grupo del aprendiente hacia el grupo de la lengua meta es positiva, se facilita la adquisición de la L2.
- (viii) **Tiempo previsto de residencia:** si el grupo del aprendiente tiene previsto permanecer en la sociedad del grupo de la lengua meta bastante tiempo, sentirá la necesidad de aprender la L2 para facilitar los contactos y la comunicación.

Como se acaba de ver, tanto la distancia social como la psicológica, junto con el grado de aculturación, son factores, según Schumann, básicos para que la adquisición de la L2 pueda superar los primeros estadios y no quede en su fase de pidginización.

A raíz de la teoría de Schumann, los estudios en ASL adoptan una nueva perspectiva reconociendo la influencia de los factores sociales y psicológicos, así como los factores individuales o de grupo en el aprendizaje de segundas lenguas. Este nuevo punto de vista, permitió explicar de una forma más coherente el proceso de adquisición de segundas lenguas. A pesar de ello, es una teoría que no ha tenido demasiado apoyo por parte de investigadores empíricos ya que consideran que el concepto de ‘aculturación’ es complejo para poder ser estudiado pues no existen medidas fiables y válidas que computen la distancia social y psicológica a la que hace referencia Schumann (Larsen-Freeman y Long, 1991; Saville- Troike, 2006).

### ***1.1.3 Teorías basadas en la interacción***

Con las teorías interaccionistas se produce un cambio de perspectiva, se pasa de considerar al aprendiente como individuo que se limita a interiorizar competencias lingüísticas a verlo como locutor que interactúa. Así, encontramos dos posturas como son las de Piaget (1979) y Vygotsky (1978) que defienden la base de la interacción en el proceso de adquisición de una lengua, pero desde puntos de vista diferentes. Mientras que Piaget (*op. cit.*) considera que esa interacción se produce entre el niño y los objetos que le rodean, que puede tocar y observar; para Vygotsky (*op. cit.*), la interacción se origina dentro de un acto social, es decir, cuando el niño se relaciona con sus padres u otros individuos cercanos. La diferencia básica entre ambos, como se observa, reside en el tipo de interacción que se establece: para el primero es una interacción física y, en cambio, para Vygotsky es una interacción social. Sin duda, ha sido el interaccionismo social el que ha tenido un recorrido más amplio.

Williams y Burden (1999: 48) señalan que:

Para los teóricos del interaccionismo social, los niños nacen en un mundo sociable y el aprendizaje se realiza a través de la interacción con otras personas. Desde que nacemos interactuamos con los demás en nuestra vida diaria y a través de estas interacciones damos sentido al mundo.

Con estas palabras, los autores ponen de manifiesto el cambio de registro en cuanto a la manera de entender la adquisición de segundas lenguas.

Los investigadores Williams y Burden (*op. cit.*) establecen como precursores del interaccionismo social a los psicólogos Vygotsky y Feuerstein. Ambos desarrollaron sus teorías en la primera mitad del siglo XX. No obstante, Williams y Burden apuntan que no ha sido hasta finales de la década de los ochenta que los investigadores han empezado a interesarse en la influencia de la teoría del interaccionismo social en la enseñanza de idiomas y a considerar las ideas propuestas por Vygotsky.

En la teoría de Vygotsky (1978) sobre el aprendizaje de lenguas, aparecen dos conceptos esenciales: (i) la mediación y (ii) la Zona de Desarrollo Próximo (ZDP). Según Vygotsky, el aprendiente, ante la realización de una tarea, afronta dos niveles de desarrollo: (i) nivel de desarrollo real y (ii) nivel de desarrollo potencial. En el primero, el aprendiente exhibe lo que es capaz de hacer de manera independiente en ese momento. Mientras que en el segundo, el aprendiente muestra lo que será capaz de hacer en el futuro. Para ir del nivel de desarrollo real al potencial es necesario, según Vygotsky, pasar por la ZDP. Esta ZDP es el estadio que se encuentra un nivel por encima de la competencia que tiene el aprendiente. Es decir, es el nivel que separa el nivel de desarrollo real y el nivel de desarrollo potencial. Para que se produzca aprendizaje en esa ZDP, el autor considera imprescindible la figura del ‘mediador’, del experto; una persona que al poseer un mayor conocimiento de la lengua va a ayudar al aprendiente a alcanzar el siguiente nivel al que aspira.

Pero la figura del ‘mediador’ también la encontramos en los postulados de Feuerstein. Para él, la interacción, entendida como ‘mediación’, consigue producir cambios en la estructura cognitiva del aprendiente. Es decir, esa mediación por parte del ‘experto’ proporciona al aprendiente las herramientas, estrategias necesarias para resolver la tarea y que en un futuro pueda utilizarlas para solucionar tareas similares. (Ellis, R. 1994; Williams y Burden, 1999; Muñoz, 2000).

Dentro de las teorías del interaccionismo social también destaca la ‘Hipótesis del Interaccionismo’ de Michael Long (1983). El autor introduce en su teoría la importancia del

*input*. Pero para Long lo relevante no es tanto que el *input* que se aporte sea comprensible, como que se haga comprensible; y para que esto suceda el ‘mediador’ tiene tres vías: (i) simplificar el *input*, (ii) presentarlo dentro de un contexto claro y significativo o (iii) alterar la estructura de la interacción. Es lo que Long llama ‘interacción modificada’. El mediador, a través de la negociación de significado, va a ir adaptando el *input* a las necesidades del aprendiente. Pues para Long (op. cit.) no es suficiente el *input* que se aporta, aunque éste esté modificado, para que se produzca adquisición, sino que es fundamental que el aprendiente participe de forma activa en la interacción con el mediador negociando el significado y así favoreciendo el desarrollo de su interlengua. Eso significa que cuantas más interacciones significativas realice el aprendiente, con su correspondiente negociación de significado, más ocasiones de aprendizaje tendrá.

Otras teorías que, al igual que la ‘Hipótesis de la interacción’ de Long, se asientan en los principios de la interacción son: (i) la hipótesis del *output* comprensible de Swain y (ii) la interacción desde el procesamiento de la información (Swain 1998, Swain 2000, Swain y Lapkin 2001). Estas teorías siguen la idea que propone Rod Ellis (1994) sobre dos aspectos que, según el autor, están estrechamente relacionados entre ellos y que intervienen activamente en la adquisición:

Aspecto interpersonal	Aspecto intrapersonal
Cuando la actividad comunicativa se produce cara a cara	Procesamiento mental que se produce durante el acto comunicativo

De la teoría de Swain habría que destacar el valor que la autora otorga al *output* como fuente de aprendizaje ya que, según ella, el *input* no es suficiente para que se produzca adquisición de una L2/LE. Swain insiste en la necesidad de que el aprendiente produzca un *output* comprensible para hacerse entender por su interlocutor, mediador. Esto le va a permitir generar un discurso más y mejor elaborado al tiempo que favorece el desarrollo de su interlengua. Otro elemento positivo que aporta la ‘hipótesis del *output*’ es que, de esta manera, el aprendiente toma conciencia de sus lagunas al reconocer los elementos problemáticos que se materializan en sus producciones (Swain 1998, Swain 2000, Swain y Lapkin 2001).

## 1.2 Antecedentes: investigación sobre proceso, agentes y estrategias en L2

Dicen Long y Richards (1990) en un prólogo a Kroll (1990) que el proceso de la escritura y la naturaleza del discurso escrito suscitaron un interés renovado en la década de los setenta, atrayendo a especialistas en educación, lingüistas, lingüistas aplicados y profesores. En Inglaterra, especialistas como Britton (1975) investigaron a los adolescentes en su proceso de escritura (planificación, borradores, toma de decisiones y heurística empleados). En EEUU, equipos mixtos de investigadores y educadores como Emig, Murray y Graves (1971) fundaron la *escuela del proceso* cuyo rastro se encuentra en la bibliografía sobre teoría y práctica de la escritura<sup>1</sup>. A partir de esta escuela, se resalta que la escritura es un *proceso recursivo* más que lineal y que hay una *variación* significativamente importante en el comportamiento escritor individual, tanto en la L1 como en la L2.

En esta línea, en la bibliografía especializada de los ochenta y noventa<sup>2</sup> aparecen investigaciones que focalizan en el proceso de la escritura y ya no tanto en el producto final. Valerie Arndt (1987:257) señala este viraje en los estudios sobre escritura: “Over the last decade or so, the change of emphasis in writing research from product to process has centred attention on the composing activities through which initial ideas and meaning evolve into written texts.”

Así, los trabajos que se publican en esa década inciden en el proceso de la escritura desde diferentes perspectivas. Por ejemplo, Ann Raimes (1985; 1987) centra su investigación en los procesos y estrategias de escritura que utilizan sujetos nativos y no nativos y concluye que existen muchas similitudes entre unos y otros. Mientras que Zamel (1983:165) insiste en que el proceso de la tarea de escribir es recursivo y no lineal: “This research has revealed that composing is a non-linear, exploratory, and generative process whereby writers discover and reformulate their ideas as they attempt to approximate meaning.” Otra manera de aproximarse al proceso de la escritura es a través de las revisiones. Siguiendo esta idea, Faigley y Witte (1981:400) dicen al respecto:

---

<sup>1</sup> Programático fue también el trabajo de Applebee, A. 1982, recogido en Nystrand (Ed.) en un volumen que es un clásico. Applebee, A. 1982. “Writing and learning in school settings”. En: Nystrand, M. (Ed.) *What Writers Know: the Language, Process and Structure of Written Discourse*, New York: Academic Press, pp. 365-382.

<sup>2</sup> Entre las muchas revistas dedicadas a la ASL en general y en especial a la escritura, destacamos *ELT Journal*, *Research in the Teaching of English*, *Language Learnign*, *College Composition and Communication*, *TESOL Quarterly* o *Applied Linguistics*.

The view of revision as a complex activity is recent. For many years teachers saw revision as copy editing, a tidying up activity aimed at eliminating surface errors in grammar, punctuation, spelling, and diction. The tidying up view of revision presupposes the three stage linear model of composing- consisting of prewriting, writing, and rewriting activities- articulated by Rohman and Wlecke in the mid-60's.

Particularmente interesantes resultan también los trabajos sobre estadios de desarrollo de la competencia discursiva a través de la escritura en Reino Unido y en Australia derivados de la perspectiva funcionalista de Halliday y Hasan (1985). Progresivamente, el interés de la investigación cubre tanto la investigación per se (valga la redundancia) como la evaluación y el asesoramiento a lo largo de los programas; esto es, la interacción entre investigación y enseñanza (*researching, assessing L2 teaching and learning*). En relación con esto, cuatro trabajos importantes sobre diseño sofisticado para la recogida de datos empíricos fueron los de Hillocks (1986), Lauer (1988), McClelland (1985) o North (1987), que sentaron las bases de la discusión sobre proceso (y procesos) de escritura y tipos de escritores, y que ya se citan en Kroll en el mismo año 1990, como punto de referencia de las investigaciones de esa década. También para entonces se habían celebrado y publicado trabajos, programáticos para esta línea de investigación y sus implicaciones didácticas, en *TESOL*, como los de Raimes (1983, 1987) y los de Zamel (1987). Los 90 propiamente dichos dejaron trabajos sobre aprendizaje de segundas lenguas de carácter más filosófico de más envergadura, como los de la relación entre la escritura en L1 y L2 (Johns, A. M. 1990, Friedlander 1990), el propio proceso de escritura (Rowe, A. 1990) o las conexiones entre lectura y escritura (Reid 1990; Fathman y Whalley 1990; Campbell 1990), en EEUU, Brasil e Israel, por ejemplo. El énfasis en la metodología (en particular en la más innovadora en ese momento, como era la del focus-on-form), fue también objeto de estudio (Fathman y Whalley 1990) en California. Tony Silva, uno de los fundadores y organizadores de los Simposios sobre Second Language Writing, también en 1990, ofreció ya una panorámica programática de la investigación sobre escritura en L2 que daba la pauta de lo que recogerían congresos y publicaciones especializadas en esa década y aun después (Silva 1990:11-23): “The L2 writer, the L2 reader, the L2 text, and the L1 context”.

La anteriormente mencionada Kroll, investigadora y editora reconocida en los 90 en EEUU y también en Europa, retoma en 2003 la tarea de edición y actualización panorámica de la investigación en un nuevo trabajo, partiendo de la afirmación de que los aprendices de ese momento, 2003, son ya la siguiente generación de escritores en segunda lengua. Y en relación a eso titula la introducción del volumen: “*Teaching the next generation of second language writers*” (Kroll 2003:1).

Los cinco bloques temáticos que vertebran su volumen (frente a las dos partes del anterior) dan una idea de la proliferación de líneas de investigación que se ha producido (y que hemos rastreado y apuntado a partir de las revistas sobre adquisición de L2, lingüística aplicada y específicamente de Writing). Para nuestros propósitos, la relación entre escritura y gramática (Liz Hampt-Lyons 2003) y los géneros en ESL/EFL (Johns, 2003) son los más interesantes, si bien es el trabajo de Polio (cap.2) sobre “Research on Second Language Writing: An overview of what we investigate and how”, donde se repasa la investigación empírica hasta esa fecha, retomando los puntos que ya Silva y Kroll pusieron de manifiesto en el 90 tan certeramente. Polio<sup>3</sup> (2003) ofrece una recopilación exhaustiva (expresada en cuatro valiosas tablas de síntesis de la bibliografía publicada sobre el tema), a saber: sobre los textos producidos por los aprendices (*foci and techniques*) (*op. cit.*: 40-41), sobre el proceso (*op. cit.*: 45-46), sobre los participantes (*op. cit.*: 51-52) y sobre el contexto social en que ocurre el aprendizaje de la escritura (*op. cit.*: 56).

En relación con los aspectos *longitudinales*, estos son raramente objeto de estudio específico en los trabajos de esa época (donde son los contextos de enseñanza y la tipología de alumnos y textos el foco, junto con la grabación en audio del proceso, una técnica generalizada y ampliamente reflejada en la bibliografía, ampliamente centrada más en el estudio de las estrategias que en la valoración de los productos en sí o en el estudio de la evolución de un sujeto a lo largo del tiempo).

Las diferencias en la L1 y en las tareas investigadas, sin embargo, sí se abordan en Zamel (1982) en un trabajo particularmente interesante, dentro de la línea de estudio de estrategias mencionada. Aunque no estudia la composición en la L1 *per se*, uno de sus sujetos la incorpora en su proceso de escritura en L2; curiosamente, el sujeto más competente en la escritura en L2; no así los otros. Estos incluso declararon en la grabación que ir comentando lo que hacían en la L1 sería como si tuvieran activados dos cerebros al mismo tiempo (coincidiendo con los hallazgos de Johnson 1987). El tema de la redacción y el recurso o no a la L1 como estrategia, y su relación con otras estrategias, es un factor importante en nuestra investigación, puesto que algunos autores<sup>4</sup> coincidieron en que ciertos temas suscitaron más uso de la L1 en la composición escrita en la L2 que otros. En particular, los relacionados con lo más personal, lo que implicaba más las vivencias y opiniones del

---

<sup>3</sup> Para una visión detallada de la investigación sobre escritura en L2/LE en EEUU, la propia Polio recomienda una fuente de gran utilidad de referencias on line: CLEAR. CLEAR hace referencia a “Center for Language Education and Research”.

<sup>4</sup> Por ejemplo: Burtoff (1983), Lay (1982) y Johnson (1987), respectivamente

alumno, frente a los temas clásicos de redacción de las universidades (a favor o en contra de una idea, afirmación, opinión de un texto, etc.). De esta serie de investigaciones retuvimos la importancia de comparar dos tipos de textos (estímulos).

Cumming (1986), también en el ámbito de la investigación de la escritura, en este caso con bilingües o casi bilingües, realizó un estudio con adultos franco-canadienses del que reportó que sus seis estudiantes adultos usaron su L1 para generar contenidos en las tres tareas de escritura solicitadas, que eran: personal, expositiva y académica, respectivamente. Cumming usó grabaciones audio, notas observacionales y cuestionarios, etc. Sobre el uso de la L1, observó que los escritores inexpertos usaban consistentemente la L1 para generar ideas, mientras que los expertos la usaban también para comprobar el estilo (en esto diverge de Johson 1987). Friendlander (1990) también recoge que en sus experimentos los aprendices usan la L1 para generar contenido y concluye que eso incide en una mejor planificación y redacción. También apunta que la traducción, que ocurre con frecuencia, no limita a los aprendices, ni en tiempo ni en calidad, mientras producen sus textos en la L2. En relación con el recurso a la L1 y el tipo de texto, no extrae diferencias notables. Sí son interesantes las medidas de fluidez según el tipo de texto y el nivel de competencia, que tomaremos en consideración en nuestro estudio también.

En 2008, destacan, por un lado, los trabajos de Leki, Cumming y Silva sobre el estado de la cuestión de la disciplina hasta ese momento, y, por otro lado, los de Lourdes Ortega. La autora puso de relieve la necesidad de abordar la escritura en L2 desde distintos contextos, resaltando el contexto no académico frente a la escritura curricular de *colleges* y universidades, recogiendo observaciones y críticas previas (Cumming 2001; Leki *et al.* 2006, 2008; Silva y Brice 2004). La manera en que se enseña y aprende la escritura varía según los contextos, edades, condiciones materiales y prácticas sociales, que no siempre coinciden en los contextos de lengua extranjera (LE/ FL *foreign language*) con los de SL (lengua segunda, con inmersión). Señala, resumiendo ese debate, Manchón (2009: 2), que en Lengua extranjera (LE/FL) muchos estudiantes o aprendices en contexto de instrucción coinciden con los universitarios (L2) en que todos tienen que *aprender a escribir* (textos formales, académicos, artículos, reseñas, etc.). Los procesos serán, pues, en parte, comparables. Pero en otros casos, en LE (FL) se *escribe para aprender* la lengua, lo que constituye un proceso bien distinto. Todavía más distinto es, destaca, los casos de CLIL (Content and Language Integrated Learning) en que se usa la lengua extranjera como *vehículo para aprender los contenidos*. El panorama, diverso, ha de abordarse también de

manera distinta, puesto que las metas de los aprendices y el papel que desempeña la lengua en cada tipo de práctica o situación también lo son.

Sea como fuere, el sustrato común que recorre todos los trabajos del volumen de Manchón (2009) es la declaración explícita de la importancia de **situar las prácticas escritoras** para entenderlas y explicarlas, y para que tengan sentido las implicaciones pedagógicas que se deriven de las investigaciones respectivas. No menos común como sustrato es el carácter empírico de todos los trabajos que avalan las revisiones teóricas y las implicaciones didácticas.

Por su lado, Leki, Cumming y Silva (*op. cit.*), en su revisión del estado de la cuestión sobre la disciplina, hacen hincapié en la eclosión que se produce en la década de los noventa acerca de los estudios sobre escritura en L2. Se produce la aparición de revistas especializadas en el tema –en la que ellos mismos están implicados—como el *Journal of Second Language Writing* (1992), o la difusión o diseminación de datos y resultados a través de congresos especializados, como el Symposium on Second Language Writing, cofundado por Tony Silva y Paul Matsuda en 1998. También Ortega (2009) llama la atención sobre este creciente interés en los estudios sobre LE/FL, como la mencionada revista, *Journal of Second Language Writing* (que en 2008 sacó a la luz un número especial editado por Manchón y De Haan). En ese mismo año 2008, el *Symposium on Second Language Writing* versó sobre “Second language Writing instruction: Principles and practices”, así, en plural. Manchón (2009) dice, textualmente “the organisers recognized that SL research has overshadowed work on L2 writing done in the foreign language (FL) context (...) a situation that the *Symposium* attempted to remedy (<http://www.sslw2008.org/>). Igualmente importante fue la aparición de colecciones o publicaciones en serie especializadas en escritura en L2 en el seno de editoriales relevantes, como Parlor Press o the University of Michigan Press.

A partir de entonces, los trabajos y estudios aparecidos en distintos foros y canales son muchísimos para los 25 años de disciplina. Además de los muy conocidos y relevantes de panorámica como los de Hyland (2002/2009), el ya mencionado de Leki, Cumming y Silva (2008), Kroll (1992, 2003), existen muchos otros editados (es decir, con múltiples autores y sobre un aspecto relevante como hilo conductor) sobre el inglés, en las dos décadas siguientes. Ejemplos como los de Silva y Matsuda (2010) sobre aspectos teóricos, Matsuda y Silva (2005) sobre aspectos metodológicos de la investigación; Matsuda, Cox, Jordan, Ortmeier-Hooper (2006) o Kasten (2010) sobre aspectos pedagógicos. Sobre la relación con

otras disciplinas o bien, la escritura en enfoques interdisciplinarios, Belcher e Hirvela (2001, 2008); Charles- Peccorari- Hunston 2009; Manchón (2011, 2013). Algunos trabajos y volúmenes han proporcionado perspectivas críticas de un único aspecto relevante, como por ejemplo el *feedback* en la expresión escrita en L2 (Hyland y Hyland 2006), estudios de género (Tardy 2011), poblaciones o grupos meta específicos, sean escritores de inglés como lengua extranjera (*foreign language contexts*), como Byrnes, Maxim y Norris (2010); Cimasko y Reichelt (2011), Manchón (2009); académico, como en Lillis y Curry (2010); o adolescentes, como Cummins (2014) u Ortmeier Hooper y Enricht (2011). En la mayoría de los casos, se trata de los resultados de proyectos de investigación colectivos, institucionales, que implican equipos de investigación, en ocasiones numerosos y con larga trayectoria, cuyos trabajos se difunden en volúmenes editados que sintetizan la investigación de un período y la relacionan con la anterior y con la de los otros equipos. Así, por ejemplo, Cumming (2006) sobre el proyecto de escritura académica que lideró en la Universidad de Toronto; Manchón (2009), sobre programas realizados en distintos lugares del mundo que hubieran tenido influencia sobre la escritura en L2; y el de Byrnes, Maxim y Norris (2010) en que recogen los resultados de su investigación de una década en Georgetown sobre los resultados del diseño e implementación de un currículo de escritura basada en géneros, destinado a estudiantes de grado (*undergraduate*) en la especialidad de alemán, cuyos resultados también presentan.

Manchón (2012) aborda el desarrollo de las competencias que integran la competencia (o destreza) escrita en inglés L2, entre las que destaca la revisión sobre la metodología de la investigación longitudinal que, nuevamente, propone Polio (2012). Manchón, Roca de Larios y Murphy (2009) y Manchón y Roca de Larios (2011) son, quizá los que más se han ocupado en nuestro país –pero sobre el inglés—de la evolución de la destreza y las estrategias en distintos contextos, como en otros contextos Rinnert y Kobayashi (2009), Byrnes (2011).

A modo de síntesis, si alguien se pregunta ¿Qué interés puede tener la aplicación de las teorías de Adquisición de Segundas lenguas al desarrollo de la escritura en L2? La respuesta, siguiendo a Manchón, podría ser que la escritura no podía seguir siendo marginal en el estudio de la disciplina puesto que las prácticas letradas (*literacy*) tienen un peso muy importante tanto entre los profesores como entre los alumnos en la mayoría de los contextos educativos donde se produce ese aprendizaje de las lenguas extranjeras objeto de estudio (como ponen de manifiesto estudios como los de Byrnes (2011) y Manchón y Roca de Larios

(2011). Pero también porque se ha acumulado ya suficiente evidencia teórica sobre producción (*output*) de alumnos que revelan el papel que, sobre el desarrollo de su competencia en L2, tienen las prácticas escritas en los distintos contextos de aprendizaje de lenguas extranjeras (Byrnes 2012; Manchón 2011 y 2012); Williams (2008, 2012); o Norris y Ortega (2009) sobre cómo entender el desarrollo de una competencia como la escrita en una lengua adicional, entre otros.

Como Ortega y Carson (2010:48) recuerdan, “Empirical research is a cyclical process that begins and ends with theories and their driven interpretations”. Complementando esta perspectiva, Cumming (2010:20) destaca el papel que debe tener la teoría en la toma de decisiones pedagógicas. La transversalidad de los enfoques y la diversidad de resultados hasta la fecha, para algunos autores (Manchón 2014) ponen blanco sobre negro nuevas perspectivas sobre la escritura en L2 en las que emergen, sobre todo dos puntos importantes:

- (i) el carácter multinivel que tienen las competencias escritoras (en plural en muchos trabajos), así como las realizaciones o materializaciones que lenguaje, textos, géneros, conocimiento o metas escritoras de cada escritor, lo que hace de la tarea de su descripción algo que puede resultar todo menos simple.
- (ii) el abanico de factores y variables que parecen influir en el desarrollo de las estrategias y competencias escritoras en la L2, como un resultado dinámico más que como una foto o explicación fija de todo el proceso subyacente.

### **1.3 El desarrollo de la escritura en L2**

Como hemos visto en el apartado anterior, los estudios que se han ido realizando sobre escritura se han llevado a cabo desde distintas perspectivas. Así, Rinnert y Kobayashi (2009), por ejemplo, estudian el papel de la L1 y el de la experiencia escritora en esa lengua en el aprendizaje de la escritura en L2. Abordan las estrategias de escritura y resaltan el papel de la transferencia de estrategias, del que destacan un aspecto muy interesante: que la transferencia se da en los dos sentidos.

Regarding the issue of transfer, the findings of our studies confirm that novice writers tend to transfer L1 textual features to L2 writing, whereas more advanced writers are more likely to depend upon L2 textual features in the development of L2 writing skills (Ferris & Hedgcock, 1998). While this observation generally holds true, our findings clearly indicate that the transfer of textual features takes place not only from L1 to L2, but rather in both directions even at novice-writer levels. For example, novice EFL writers who had received

only preuniversity writing training tried to construct texts in both Japanese and English by relying on the knowledge they obtained from L1 or L2 writing instruction or their combination, as specifically illustrated in the case of students who used the same rhetorical features in the two languages. (Rinnert y Kobayashi, 2009: 41)

Trabajos de otros investigadores cuyos sujetos son multilingües (como los del estudio de Navés y Celaya 2009; Cumming 2009 o Manchón *et al.* 2009), coinciden con ellos en el papel de la transferencia de estrategias entre lenguas, no sólo de la L1 a la L2, sino también de la L2 a la L3. Bajo esta misma idea, De Angelis y Jessner (2012) afirman que la presencia de uno o más subsistemas de lenguaje influye en cierta manera no sólo en el desarrollo de la segunda lengua (o tercera, podríamos añadir), sino que también afecta al desarrollo del sistema multilingüe en general, pues los distintos subsistemas se interrelacionan entre sí.

Entre los pocos trabajos longitudinales en escritura que podemos tomar como muestra, destacamos el de Schoonen, Sneellings, Stevenson y van Gelderen (2009) que reportan los resultados de su *Nelson Project* en L1 y L2 en la escuela secundaria en Holanda. Es experimentalmente muy completo: pueden disponer información de protocolos de *thinking-aloud*; usan la pulsación de la tecla (*key stroke*), y un estudio más experimental también para mejora de recuperación léxica (memoria). En su explicación basada en los distintos estudios que llevaron a cabo en la investigación, resaltan la interrelación entre los recursos cognitivos y los lingüísticos a la hora de llevar a cabo las tareas de escritura, y los relacionan con modelos distintos en L1 y en LE. Uno de sus resultados más relevantes es que la escritura en LE es más local y está más orientada a la lengua que la escritura en L1. A partir de su investigación y sus resultados, postulan una ‘Inhibition Hypothesis’<sup>5</sup> o hipótesis de la inhibición, según la cual la naturaleza de la demanda o exigencia de procesamiento de lenguaje que implica en la L2/LE el ir a las fuentes y recursos para llevar a cabo la tarea, impide centrarse en los aspectos más formales de la producción textual, el contenido y/o las características textuales en general. Coinciden con Rinnert y Kobayashi (2009), y con Sasaki (2009) al señalar que hay más conocimiento metacognitivo general, de tipología textual y de escritura propiamente dicho asociado a la escritura en L1 y FL/L2 que relación entre competencia en L2 y escritura solamente. Este hallazgo corrobora los resultados de Rinnert

---

<sup>5</sup> Esta línea, que dejan abierta en 2009, es la que se retoma en esta tesis: el conocimiento meta que permite abordar de forma distinta tareas que pertenecen a géneros textuales distintos (como las dos del experimento) y re-utilizar la experiencia previa para mejorar su actuación a lo largo del tiempo (corto) que transcurre entre la primera y la segunda vez que realizan la tarea. Más que hipotetizar sobre “el posible efecto en su producto” que tendría el proceso (grabaciones audio sobre la reflexión estratégica), la comparación de los 2x2 productos permite allegar evidencia empírica de la puesta en práctica de los recursos y estrategias postulados.

y Kobayashi sobre transferencia de estrategias de la L1 a la L2, si bien en esta ocasión con escritores más jóvenes y en un contexto educativo y geográfico distinto.

Otra investigación longitudinal es la de Miyuki Sasaki (2009) que hace el seguimiento de 22 sujetos a lo largo de tres años y medio. En palabras de la autora:

The study is a follow-up of another recent longitudinal study (Sasaki, 2004), and it was also motivated by the findings of five other studies (Hirose & Sasaki, 1994; Sasaki & Hirose, 1996; Sasaki, 2000, 2002, 2007) that I have conducted over the past 15 years. (Sasaki 2009: 49)

De esos seis estudios a los que hace referencia Sasaki, solo uno de ellos se centra en la mejoría o incremento de la calidad de la escritura. La autora se interesó más por el proceso que por el producto, de ahí que en los otros trabajos focalice en el proceso de la escritura, en los cambios en el comportamiento en la escritura en L2 y la motivación y en los estudiantes que están inmersos en la L2 (*stay abroad*); contexto éste típico de LE/FL que, no obstante, no había sido objeto de estudio hasta prácticamente esta década, y que requería de exigencias metodológicas propias para atender a un factor determinante en ese contexto, como la motivación y la importancia: (i) de poder formar parte de una comunidad (más o menos imaginada previamente) de esa L2; (ii) de sumarse efectivamente a las prácticas de esa comunidad. Dörnyei (2005) había señalado anteriormente la importancia de estudiar esa variable en la motivación (específicamente), pero no se había abordado satisfactoriamente hasta que se publicaron trabajos como los de Cumming (2006) o éste de Sasaki (2009).

Sobre este tema (*stay abroad*), los estudiantes que participaron en la investigación de Sasaki (*op. cit.*) reportan en sus entrevistas que la estancia fuera ha tenido un papel tan determinante, que la autora decide llevar a cabo otro estudio *follow-up* (2007) sobre la expresión escrita con siete estudiantes que se van fuera y seis que se quedan en su universidad, en Japón. Curiosamente, aunque los dos grupos mejoran en sus resultados generales, solo el que va fuera (SA o *Stay Abroad*) mejora en su composición escrita. Textualmente, Sasaki dice (2009:52): “The results showed that only the SA group improved their L2 writing ability, although both groups improved their general L2 proficiency.” A partir de ahí, se vuelve cognitivista porque le parece enorme el papel que el entorno y la interacción con éste desempeñan en los aprendices (2009: p.52):

These results confirmed the importance of examining the participants’ cognitive development as situated in their sociocultural environments. At this point, as a researcher, I redirected myself toward a larger theoretical framework than the ‘main-stream’ (Thorne, 2005: 393) exclusively cognitivist research perspective.

Otra línea de investigación interesante es la propuesta por Manchón, Roca Larios y Murphy (2009) sobre el papel de la tarea (tipo de tarea), además de las estrategias individuales en L1 y L2 y el factor tiempo (estudio longitudinal); elementos que conforman los pilares de su estudio con escritores universitarios. La tarea de escritura concebida como resolución de problemas, por un lado, y el tiempo empleado en cada tarea (como indicadores del procesamiento, dificultades, procesos de composición activos, etc.) son los dos polos de este proyecto de investigación de la Universidad de Murcia. Los autores concluyen que demanda más tiempo la escritura en L2<sup>6</sup>, y que, como proceso, también parece más centrado en la lengua en sí (el repertorio lingüístico), sus fuentes y recursos; postulan que el escritor de L2/FL es un usuario lingüístico ‘multicompetente’ y concluyen que la competencia escrita es una actividad ‘multilingüe’ o un hecho multilingüe o bilingüe.

Celaya y Navés (2009) investigan una población de aprendices de inglés LE a lo largo de varios cursos escolares (de primaria y secundaria, aunque se centran en los menores) y describen la evolución de la competencia escritora y el uso de la L1 (transferencia léxica) en las tareas de escritura por parte de aprendices de más baja competencia que los estudiados habitualmente. Sus conclusiones, sintetizadas en la máxima “cuanto mayores, mejores” (“*the older, the better*”), ponen de manifiesto que los aprendices que empiezan más tarde su formación en inglés escriben mejor en inglés L2. Y también que los multilingües recurren a su competencia multilingüe (y a las distintas lenguas) durante el proceso de composición o escritura (coincidiendo con los estudios de Rinnert y Kobayashi (2009); y de Sasaki (2009)).

Trabajos más recientes y que siguen la teoría de los sistemas dinámicos es el de Verspoor y Smiskova (2012:18). En su investigación sobre el desarrollo de la escritura en L2 focalizan en lo que ellas llaman “formulaic sequences, or ‘chunks’”. Es decir, combinación de palabras, colocaciones<sup>7</sup> o lo que en nuestro trabajo llamamos también ‘agrupamientos’. Para las autoras hay dos maneras posibles de aprender dichas combinaciones o agrupamientos de palabras, (i) como un todo o (ii) por partes, y va a depender de diferentes factores, como la extensión del agrupamiento, de la habilidad del

---

<sup>6</sup> Manchón, R.M. (2009): “the internal structure of the process of converting ideas into language (which appears to be more time- and attention-consuming in the L2 condition), the purported recursive nature of composing (which was found to be proficiency-dependent), or the strategic behavior engaged in while writing (which in FL writing involves conspicuous recourse to the various resources in the L2 writer’s linguistic repertoire).”

<sup>7</sup> En nuestro trabajo, aunque también analizamos los agrupamientos que se producen de forma más habitual en cada una de las muestras analizadas, no es el fin último de la investigación.

estudiante para recordar las cadenas de palabras o del nivel de lengua. Según las autoras, hay estructuras que los estudiantes aprenden como un todo antes de conocer el significado aislado de cada una de las palabras, por ejemplo: “How are you”, “fine, thank you, and you?” (op.cit.: 20).

En la teoría de sistemas dinámicos, como subrayan las autoras, la idea básica es que cada individuo es diferente y, por tanto, su desarrollo de la L2 también va a seguir una evolución distinta, pues sus experiencias vitales también lo son y esto influye en el aprendizaje. Siguiendo esta misma idea, De Angelis y Jessner (2012) afirman que hay factores (i) cognitivos y afectivos, (ii) sociales o (iii) sociolingüísticos que van a influir en el desarrollo del aprendizaje.

En la investigación que llevan a cabo Verspoor y Smiskova (*op. cit.*) comparan dos grupos que reciben un tratamiento diferenciado. Mientras uno sigue la enseñanza habitual, el otro sigue una enseñanza en la que se hace hincapié en el *input* que se ofrece. El segundo grupo recibe mayor cantidad de *input* que el primero. Las autoras toman como grupo control a sujetos nativos de inglés. Para evaluar la trayectoria individual y así valorar la teoría de sistemas dinámicos, seleccionaron a un estudiante de cada uno de los grupos para analizar con más detalle el uso de agrupamientos de palabras (*chunks*) en sus escritos. Los resultados que arrojó la investigación es que el grupo que recibió mayor cantidad de *input* no mostró diferencias significativas respecto al otro grupo. No obstante, en un estudio posterior, Verspoor, Schmid y Xu (2012) confirman que el análisis de los agrupamientos de palabras es una forma sólida de medir la competencia porque deja ver de qué manera mejora el rendimiento del estudiante. En esta investigación, el número de participantes aumentó considerablemente, participando un total de 489 estudiantes. Aun así, las autoras indican que:

Despite the fact that our corpus was carefully controlled for L1, age, aptitude, and tasks, there is a great deal of variation among learners. Still most variables do show change, but at different rates. Measures such as the lexical frequency bands changed very little across proficiency levels and showed significant differences only between levels 1 and 5; (*op. cit.*: 257).

De Angelis y Jessner (*op. cit.*), por su lado, focalizan su estudio en medir la competencia a través de las medidas de fluidez, complejidad gramatical y complejidad léxica, pero puesta la mirada en la interacción entre los distintos subsistemas de lenguaje que poseen los estudiantes. Para valorar la fluidez, las autoras toman como medidas el número

total de palabras y de T-Units<sup>8</sup>. Dicen textualmente: “Fluency in this case refers to the amount of text students were able to write within the limit they were given.” (De Angelis y Jessner, 2012: 53). Los resultados que aportan los datos es que existe una gran relación entre los distintos subsistemas de lenguaje que poseen los sujetos de la investigación.

Un estudio reciente de Mancilla *et al.* (2015) compara la producción escrita de estudiantes universitarios nativos y no nativos de inglés durante un periodo de cinco años, entre 2009 y 2013. Basaron su comparativa en diez medidas, entre otras, sobre la complejidad sintáctica, dirigidas a valorar la longitud de las frases, la cantidad de subordinación y coordinación, y el grado de complejidad de las frases. Los resultados, en líneas generales, mostraron diferencias significativas en cuanto a una mayor presencia de subordinación en el grupo de nativos y mayor coordinación en el de no nativos. Por el contrario, no encontraron diferencias significativas entre el grupo de nativos y el grupo de no nativos con competencia más alta, y sí pequeñas diferencias entre el grupo de nativos y no nativos con competencia más baja en relación a la subordinación.

#### **1.4 El desarrollo de la escritura en español L2/LE**

En la década de los ochenta se empieza a vislumbrar un interés creciente por la enseñanza del ELE en todos sus ámbitos y eso da lugar a la creación de congresos<sup>9</sup> dedicados exclusivamente a esa disciplina. Ya entrado en el siglo XXI aparece una obra de referencia para la enseñanza del español como lengua extranjera, *Vademécum para la formación de profesores*, obra dirigida por Sánchez Lobato y Santos Gargallo (2004) en la que se hace un compendio de los principales temas que atañen al ámbito del español. Recientemente se ha publicado otro libro de referencia, *The handbook of Spanish Second Language Acquisition*, editado por Kimberly L. Geeslin (2014). En la introducción se remarca este auge:

Presently, a quick search of research publications on second language Spanish would certainly confirm an explosion of interest and high-quality research. In fact, a survey of the contributing authors to this volume shows that interest in second language Spanish is a pan-national pursuit, extending well beyond Europe and the Americas to institutions from Asia and Australia as well. (Geeslin 2014: 1)

Centrándonos en el tema de la competencia escrita, vemos que aunque las investigaciones sobre el desarrollo de la escritura en español L2/LE no son tan abundantes

---

<sup>8</sup> En nuestra investigación analizamos la fluidez sólo a partir del total de Tokens. Esto es así porque en las producciones escritas de los sujetos NN era difícil establecer qué era un T-Unit por la ausencia de puntuación en la mayoría de los textos.

<sup>9</sup> Los congresos organizados por la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE) desde 1989 hasta la actualidad o los congresos de Expolingua que celebra el primero en 1987.

como las que encontramos en inglés y que hemos mencionado en el apartado anterior; en las últimas décadas han ido apareciendo trabajos que han centrado su interés y, al mismo tiempo, su preocupación acerca de la competencia escrita. Ejemplo de ello lo encontramos en palabras de Lacámara (1989:1):

Uno de los aspectos que más han llamado nuestra atención en la enseñanza-aprendizaje de una lengua extranjera, *en un medio escolar*, ha sido el nivel de adquisición tan bajo alcanzado por los estudiantes en el aprendizaje escrito, tras varios años de estudio de la lengua.

También Isabel García Parejo (1993: 290) expresa su preocupación sobre la dificultad que presentan los aprendices de español para producir textos escritos de forma coherente:

A pesar de que algunos alumnos habían estudiado hasta seis años de español, quedaban clasificados en el nivel más bajo por su incapacidad de poder expresarse correctamente por escrito. Esta experiencia, junto con el análisis de composiciones de estudiantes españoles adultos con escasa formación académica, nos hizo llegar a la conclusión de que, ya que el tipo de errores entre todos ellos era muy parecido, podríamos clasificar dentro de la categoría de analfabeto funcional a nuestros estudiantes extranjeros: pueden leer, escribir y hablar, pero se encuentran incapacitados para desenvolverse dentro de su mundo, en este caso, la universidad española.

Martín Peris (1993: 184) manifiesta la misma inquietud ante el tratamiento que recibe la competencia escrita: “La situación actual puede describirse como aquejada de una cierta desatención a esta destreza.”

Más reciente es la opinión de van Esch *et al.* (2006: 57), siguiendo la misma línea argumental, cuando afirman que:

A pesar de que los estudiantes de lenguas extranjeras de las universidades holandesas dedican durante su carrera un tiempo considerable a la redacción de ensayos, tanto sobre temas generales como sobre temas específicos, no hay en nuestro país una tradición asentada en la enseñanza y evaluación de este tipo de textos.

Esto lleva a que los primeros trabajos que encontramos sobre la competencia escrita se dirijan hacia la reflexión y la propuesta didáctica. Es decir, analizar qué es lo que sucede dentro del aula y proponer nuevas estrategias de trabajo con el objetivo de optimizar el rendimiento.

Así, en el trabajo de Lacámara (*op. cit.*) se propone un acercamiento a la enseñanza de la escritura dentro del aula a partir de la ‘gramática textual’. Según el autor, esta es la manera de destacar los componentes que conforman el texto escrito para llegar a producir textos coherentes. En esa misma línea, qué se debe hacer para que el estudiante de ELE mejore su

rendimiento en la competencia escrita, va el trabajo de Pastor (1994), “El desarrollo de la competencia escrita del alumno de español”. Como ella misma dice para introducir el tema, el objetivo “es ofrecer una serie de reflexiones acerca de la didáctica de la expresión escrita en el aula de ELE que nos permitan un desarrollo efectivo de la competencia discursiva escrita del alumno” (Pastor 1994: 250). Así, lo que la autora propone es que en el aula no solo se debe ‘escribir’, sino que se deben hacer propuestas sobre “qué escribir, a quien, cómo y para qué” (Pastor 1994: 247). Martín Peris (*op. cit.*:187) señala: “Parece que un buen trabajo de didáctica de la expresión escrita deberá incorporar el tratamiento adecuado de las actitudes y procedimientos de los escritores competentes.”, y para ello propone unas tablas de seguimiento, una para el docente y otra para el alumno, con las que se ayuda a tomar conciencia de qué se está trabajando sobre expresión escrita. En *E de escribir*, Díaz y Martínez (1998) hacen una propuesta didáctica focalizando en el proceso de escritura. Otro trabajo también dedicado a la didáctica y dirigido a docentes es el de Díaz y Amerych (2003). La propuesta que hacen las autoras la resumen en las siguientes palabras: “(...) una muestra teórico-práctica de ejercicios que ilustran y fundamentan, paso a paso, el proceso de generación de un texto escrito en E/LE.” (Díaz y Amerych 2003: 4); y el objetivo que se plantean es “indagar en el proceso de escritura y buscar la manera de hacerlo consciente, para potenciarlo y desarrollarlo en el aprendiz de español como lengua extranjera” (Díaz y Amerych 2003: 7).

Además de este enfoque más didáctico, también hay estudios que ofrecen, como el de Cassany (1999: 47), una panorámica de lo que según el autor “el papel que puede -¡y debe!- desempeñar hoy la composición escrita en E/LE.”. También hay investigaciones que centran la atención en la evaluación de la expresión escrita, como los trabajos de Ortega Ruiz (1994), el de Eguiluz Pacheco y De Vega Santos (1996) o Eguiluz Pacheco & Eguiluz Pacheco (2004).

Otra corriente de investigación es la que se adentra en el aula de ELE y mediante corpus<sup>10</sup> de aprendices (Cruz, 2012) conseguidos a partir de las producciones escritas por los mismos estudiantes se analizan diferentes aspectos de la lengua. Estas distintas perspectivas sobre cómo abordar el desarrollo de la competencia escrita en español L2/LE Polio ya las puso por escrito en 2001, a saber: (i) calidad general del texto, (ii) corrección lingüística,

---

<sup>10</sup> En nuestra tesis partimos también de ‘corpus de aprendices’. Por un lado, con un corpus formado por textos escritos por sujetos no nativos de la lengua española y, por otro lado, con un corpus de textos escritos por sujetos universitarios nativos de la lengua española.

(iii) complejidad sintáctica, (iv) características léxicas, (v) contenido, (vi) coherencia y características discursivas, (vii) mecánica, (viii) fluidez y (ix) revisión.

En los últimos años encontramos estudios que han focalizado en alguna de dichas áreas de estudio. Así, los trabajos de Díaz *et al.* (2001, 2003) se centran, por ejemplo, en analizar la calidad<sup>11</sup> del texto a partir de herramientas informáticas (*Conexor* y *Childes*) y considerando diferentes índices de calidad como la riqueza y variedad léxica, y la riqueza sintáctica. En palabras de las autoras, “El objetivo es cuantificar en la medida de lo posible la distancia entre el original producido por un nativo (*input* del modelo) y las producciones (el *output* más el *intake*) de los alumnos.” (Díaz *et al.* 2003: 221). Los resultados del análisis mediante las herramientas informáticas constatan la existencia de diferencias entre el texto modelo nativo y las producciones de los no nativos. Los no nativos presentan producciones con valores más bajos en los tres índices medidos.

Si en el trabajo de Díaz *et al.* (2003) se hacía un análisis transversal, en el de Martínez Arbelaz (2004) el análisis es longitudinal, pues compara las producciones escritas por los mismos estudiantes al inicio y al final del curso, tras tres meses de instrucción<sup>12</sup>. En este caso, la autora investiga si “los alumnos en situación de inmersión experimentan una mejora significativa en lo que a la corrección gramatical se refiere o no.” (Martínez Arbelaz 2004: 115). Para ello, Martínez Arbelaz identifica veintidós rasgos léxicos y gramaticales con los que los alumnos de habla inglesa tienen dificultades. Además de observar si se produce progreso en esos aspectos, también analiza si hay un aumento en la fluidez escrita. Es decir, se centra más en la corrección lingüística y en las características léxicas que en la calidad del texto. Para la recogida de datos, el género textual que utiliza es la narrativa en pasado<sup>13</sup>. Tras un análisis cuantitativo, la autora concluye que se produce una mejora estadísticamente significativa en la fluidez y también en algunos de los rasgos gramaticales y léxicos estudiados.

Resulta de especial interés por la extensión y los temas que abarcan los trabajos realizados por Van Esch *et al.* (2004), De Haan y Van Esch (2005), Teijeira Rodríguez *et al.*

---

<sup>11</sup> En nuestra tesis también analizamos la calidad del texto, pero utilizando los índices de fluidez, diversidad léxica y densidad léxica. Para extraer los datos también hacemos servir herramientas informáticas descritas en el capítulo de metodología.

<sup>12</sup> Este intervalo de tiempo entre la primera recogida de datos y la segunda coincide con nuestra investigación.

<sup>13</sup> El estímulo era relatar un viaje o una anécdota; mantiene cierta similitud con uno de los textos propuestos en nuestra tesis: relatar el primer día de una clase de idiomas.

(2005) y Van Esch *et al.* (2006) con aprendices holandeses de español. Todos esos trabajos se incluyen en un gran proyecto de investigación de la universidad holandesa Radboud University Nijmegen llevado a cabo entre los años 2002 y 2008 y en el que el corpus está formado por textos argumentativos en inglés LE y en español<sup>14</sup> L1 y LE. En los trabajos de Van Esch *et al.* (2004) y De Haan y Van Esch (2005) analizan el desarrollo de la escritura tanto del inglés como del español LE; el de Tejeira Rodríguez *et al.* (*op. cit.*) focaliza en la estructura, la coherencia y la cohesión y compara el corpus formado por textos escritos de español L1 y LE; y el de Van Esch *et al.* (2006) se centra en analizar si la calidad discursiva de los textos escritos del grupo de no nativos se acerca al del grupo de nativos de español L1. Para el corpus de textos de español L1 cooperaron estudiantes de la Universidad de Murcia. De Haan y Van Esch (*op. cit.*: 108) señalan que la comparativa entre los textos de los sujetos nativos y no nativos se puede realizar ya que “In Spanish schools there is no real training in writing and students are not taught to compose, not even in their mother tongue (Murphy, personal communication). So, in all these respects, they are comparable to our students.”, ya que “The Netherlands have no great tradition in the teaching of formal writing, neither in secondary nor in tertiary education.” (*op. cit.*: 101).

Como vemos, en los distintos trabajos han ido ahondando un poco más en el análisis, pero la manera de aproximarse a los datos ha sido desde una perspectiva cualitativa principalmente. No obstante, en el trabajo de De Haan y Van Esch (*op. cit.*) cuando comparan los resultados del corpus de español L1 y LE realizan un análisis cuantitativo y longitudinal. Por un lado, observan la evolución de los estudiantes no nativos a lo largo de un año académico en el que comprueban que hay un incremento considerable en el número de palabras en los textos del segundo año respecto del primero: “a dramatic increase in the number of words produced by the non-native writers.” (*op. cit.*: 113). Asimismo, también se produce una mejora en la variedad léxica al aumentar la ratio *Types/Tokens*. Por otro lado, al comparar los resultados del grupo de no nativos con los nativos, comprueban que la mejora observada en los datos del segundo año de los no nativos están próximos o incluso superan los resultados del grupo de sujetos nativos como, por ejemplo, en el número total de palabras. Hecho éste que indica, según los autores, que hay nativos que presentan menor

---

<sup>14</sup> Para nuestra tesis es de especial interés por las similitudes que presenta con nuestro trabajo: estudio longitudinal de un grupo de sujetos no nativos de español y la comparativa entre nativos y no nativos, aunque el género textual no es el mismo que en nuestra tesis. También hay que resaltar que focalizan en el producto y no en el proceso de escritura.

fluidez que algunos sujetos no nativos en el segundo año.

El otro trabajo en el que también se hace una investigación longitudinal es el de Van Esch *et al.* (2006). El objetivo del mismo es indagar “si se producen cambios significativos en los ensayos escritos sobre el mismo tema y por los mismos estudiantes de un año a otro, en cuanto a la competencia discursiva” (Van Esch *et al.* 2006: 61). En este caso, al contrario que en el anterior estudio, no hay comparativa con sujetos nativos y el análisis principal es cualitativo; simplemente aportan una tabla con valores cuantitativos sobre el número total de *Tokens*, *Types*, longitud de palabras, etc. para comparar la evolución de la calidad de los textos entre el año uno y el año dos. Para realizar el análisis de la calidad discursiva, los autores señalan que hay diferentes elementos que se pueden valorar, como la corrección gramatical, léxica y ortográfica; el cómputo total de palabras y la ratio de *Types* y *Tokens*; la coherencia y la cohesión, o la adecuación del léxico al tipo de texto. Entre todas esas posibles medidas, los autores eligen las de coherencia y cohesión, y complementan el análisis con los datos cuantitativos ya mencionados. Los datos reportan una mejoría en los textos del año dos: “En los textos del segundo año tienden a elidir el sujeto, lo que los acerca un poco más a los patrones españoles.” (Van Esch *et al.* 2006: 72). Concluyen que se observa una evolución del año uno al dos hacia textos más próximos a cumplir con las características inherentes al género argumentativo.

En el trabajo de Teijeira *et al.* (2005) también se comparan textos de nativos y no nativos, igual que en el de De Haan y Van Esch (2005), pero aquí el análisis no es longitudinal, pues el objetivo no es ver la evolución de la competencia escritora de los sujetos no nativos, sino comparar con nativos para identificar las posibles diferencias respecto a organización, coherencia y cohesión. El análisis que se realiza es cualitativo, como en el trabajo de Van Esch *et al.* (2006). Algunas de las conclusiones a las que llegan es que los sujetos no nativos son capaces de escribir textos coherentes a pesar de estar en un estadio no muy avanzado de sus estudios de español (segundo año). También les llama la atención a los autores la similitud en la manera de organizar la argumentación de uno y otro grupo. Una diferencia que atrae su atención es el hecho de que los no nativos tienden a poner título a sus ensayos, mientras que los nativos no suelen hacerlo.

Los trabajos de Díaz y Lucha (2010; 2012) guardan ciertas similitudes con los que acabamos de presentar del grupo de la Radboud University Nijmegen de Holanda. En estas dos investigaciones se analizan también textos argumentativos y se comparan corpus de

nativos y no nativos. Lo interesante de estos dos trabajos es que tienen un factor añadido, los sujetos no nativos se dividieron en dos grupos, uno recibió enseñanza mediante técnicas de *Focus on Form* (FoF), mientras que el otro grupo siguió la enseñanza habitual del curso. En ambas investigaciones se analiza la calidad del texto, la cohesión textual y las estructuras características del género argumentativo, coincidiendo en esto con el trabajo de Teijeira *et al.* (2005). Tanto en las investigaciones de Díaz y Lucha y en el de Teijeira, el objetivo principal era indagar si había elementos estructurales o de léxico que alejasen los textos de los sujetos no nativos de los nativos. Los datos confirmaron que no había diferencias significativas entre nativos y no nativos sin instrucción con técnicas de FoF, pero sí con los sujetos no nativos que estudiaron con técnicas de FoF: “FoF teaching causes visible difference in performance, especially in the linguistic and genre aspects targeted.” (Díaz y Lucha 2010: 29). Otra conclusión a la que llegan las autoras es que el ser nativo de una lengua no supone un rasgo distintivo en este caso, es el tipo de instrucción que se recibe la que hace que los resultados sean más o menos positivos.

Una investigación reciente, Baerlocher (2013), recurre a herramientas informáticas (WordSmith), igual que en los trabajos de Díaz y Lucha (2010; 2012), Díaz *et al.* (2001; 2003), para analizar los datos de su corpus. En este caso, la autora lleva a cabo un estudio pseudolongitudinal centrado en la corrección lingüística (análisis de errores). El objetivo es comprobar si estudiantes brasileños (futuros profesores de ELE) mejoran su competencia lingüística a lo largo de su carrera universitaria. También se analiza la fluidez, la diversidad léxica y la densidad léxica de los textos. En estos tres aspectos<sup>15</sup> los datos revelaron que había una progresión en función de las horas de instrucción recibidas.

---

<sup>15</sup> Estas variables también las analizamos en nuestra tesis.

## 1.5 La lingüística de corpus en la investigación de segundas lenguas

Lo primero que llama la atención cuando consultamos qué es un ‘corpus’ es la variedad de definiciones que encontramos:

- (i) DRAE<sup>16</sup>: (Del lat. *corpus*; propiamente ‘cuerpo’).  
1. m. Conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación.
- (ii) Sinclair (1991: “(...) a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of language.”
- (iii) López Morales (1994: 95): “Llamamos *corpus* al conjunto de textos, sea cual sea su origen, que constituye el material de análisis de una investigación dada.”
- (iv) Alvar Ezquerro *et al.* (1994: 9-10): “Un corpus es un conjunto homogéneo de documentos lingüísticos de cualquier tipo (orales, escritos, literarios, coloquiales, etc.) que se toman como modelo de un estado o nivel de lengua predeterminada, al cual representan o se pretenden representar.”
- (v) EAGLES<sup>17</sup> (1996: 4): “Corpus: A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language”.
- (vi) Sinclair (2004): “A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.”

Como vemos, las definiciones se han ido perfilando con los años. Comprobamos, por ejemplo, como Sinclair pasa de una definición bastante general en 1991 a una más detallada en 2004. Cruz (2012: 23) resume la idea de Sinclair destacando que se trata de una definición que recoge los fundamentos de lo que es un corpus lingüístico:

es una recopilación de textos (o fragmentos de textos) que actualmente se concibe necesariamente en soporte electrónico; debe ser representativo de la realidad de la lengua que se vaya a estudiar y, para ello, se deben seguir unos criterios objetivos de selección de los textos (para recoger todas las estructuras posibles y no solo las que el investigador quiere encontrar), y es una fuente de datos para investigaciones lingüísticas, es decir, que el corpus no constituye una meta en sí, aunque compilarlo es una labor que requiere una profunda reflexión lingüística.

---

<sup>16</sup> Diccionario de la Real Academia Española, Vigésimo segunda edición. Consulta en línea (www.rae.es)

<sup>17</sup> EAGLES: Expert Advisory Group on Language Engineering Standards.

Una vez clarificado qué es un ‘corpus’, el siguiente paso es saber quién lo utiliza y qué se puede hacer con él. Como señalan diferentes autores (Baker, 1993; Laviosa, 1998; Cruz, 2012) las disciplinas que recurren a los corpus para hacer investigación son de diferentes áreas como la sociolingüística, la teoría literaria, la psicolingüística o la lingüística aplicada. Sin duda, el creciente interés en la lingüística de corpus se ha visto reforzado por la irrupción de programas informáticos que han ayudado al procesamiento de los datos (Ramallo, 1999).

Cruz (*op. cit.*: 35) resume muy acertadamente las ventajas de trabajar con corpus:

(...) permiten disponer de muestras reales de la lengua en uso, poner a prueba y comprobar las teorías lingüísticas o las intuiciones, comprobar la gramaticalidad de las oraciones dentro de su contexto, extraer reglas o descripciones lingüísticas a partir de la observación de la lengua, obtener información cuantitativa y, por lo tanto, verificable, y manejar un gran volumen de información en poco tiempo.

Un ejemplo práctico de estos posibles usos es el que Asención-Delaney y Collentine (2011: 301) explican en su trabajo:

The research on the acquisition of Spanish’s two copulas *ser* and *estar* shows how learners associate lexical and grammatical features with particular discourse types.

(...)

Corpus-based L2 research has helped us to understand how learners map form to function. It also elucidates the role of L2 formulaic segments. It has provided important insights into the interaction of grammar and the lexicon.

A continuación presentamos algunos de los ejemplos de *clusters* que Muñoz Lobo (2012: 133) ofrece con el sustantivo ‘tiempo’: “a tiempo; con tiempo; al tiempo; dedicar tiempo; matar el tiempo”.

Otra investigación que también recurre al análisis de corpus es el de Cantos y Sánchez (2011: 15). Los autores comparan los corpus de *Cumbre* (para la lengua española) y *Lacell* (para la lengua inglesa) con el objetivo de establecer las diferencias o similitudes “sobre la estructura y vertebración del inventario léxico en ambas lenguas”.

Para hacer esos estudios, herramientas informáticas como *WordSmith* o *AntConc* permiten la extracción de información como lista de frecuencias, *clusters* (agrupamientos) o concordancias. Estos datos permiten caracterizar el tipo de texto que estamos analizando. Ejemplo de ello lo encontramos en el trabajo de Asención-Delaney y Collentine (2011: 319): “The analysis uncovered four significant clusters that can be considered distinct discourse types. The communicative functions of these discourse types break down into two main

stylistic variations: narrative (characterized by verbal features) and expository (characterized by nominal features).”

Cruz (*op. cit.*) señala que las concordancias ayudan a ver la palabra que se desea analizar en contexto, permitiendo, así, ver el comportamiento de dicha palabra en una situación comunicativa real. Esto tiene unas implicaciones didácticas claras como (i) enseñanza de léxico; (ii) análisis de errores, de esta manera, por ejemplo, se puede detectar cuáles son los errores más frecuentes por grupos de lengua; (iii) aportar más *input*, animando tanto a estudiantes como a docentes a consultar estos corpus en línea. Cruz (*op. cit.*).

Un tipo de corpus de interés para la enseñanza de segundas lenguas es el llamado ‘corpus de aprendices’ (Cruz, 2012). Es un corpus formado por textos producidos por los sujetos que están aprendiendo una lengua (en el caso que nos atañe, el español). En los ejemplos que hemos presentado, el de Asención-Delaney y Collentine (*op. cit.*) pertenece a este grupo, mientras que los de Cantos y Sánchez (*op. cit.*) y el de Muñoz Lobo (*op. cit.*) son corpus generales. Para el español L2, Cruz (*op. cit.*) señala el corpus CEDEL2 como corpus de referencia, pues recoge más de 750.000 palabras. Es un corpus formado por muestras de producciones escritas en español por aprendices de todos los niveles.

Las ventajas del corpus de aprendices es que son “una fuente empírica de datos para trabajar en la línea del ANÁLISIS DE ERRORES y del estudio de la INTERLENGUA y, a partir de los resultados, diseñar cursos y materiales más adecuados a las necesidades reales de los aprendices.” (Cruz, 2012: 137).

## **1.6 Ser nativo de una lengua: ¿Qué lengua enseñar?**

Una preocupación de los docentes de segundas lenguas es saber qué lengua meta llevar al aula. Para ello, se ha recurrido a las propuestas que los investigadores han hecho sobre qué tipo de competencia comunicativa debe tener un hablante y cómo conseguir que el estudiante de una L2/LE adquiera dicha competencia.

Hymes (1972), por ejemplo, habla de competencia comunicativa para referirse a la existencia de reglas de uso sin las cuales las reglas de la gramática son inútiles. El autor resalta también la importancia de observar y analizar los problemas dentro de una sociedad heterogénea y valora de forma especial el hecho de que el niño cuando adquiere el conocimiento de las oraciones, no sólo lo adquiere como gramaticales, sino también como

apropiadas. Es decir, el niño adquiere el conocimiento necesario para saber cuándo tiene que hablar y cuándo no, de qué puede o no hablar, con quién y de qué forma. Según Hymes (*op. cit.*), esta competencia comunicativa se adquiere gracias a la experiencia social y a las necesidades comunicativas que ello conlleva.

Widdowson (1995) resume las principales opiniones sobre el concepto de competencia comunicativa de la siguiente manera:

Autor	Teoría
Chomsky	La competencia es conocimiento
Hymes	La competencia es conocimiento y habilidad
Bialystok y Sharwood-Smith	Habilidad es conocimiento y control

Tabla 2. Concepto de ‘Competencia comunicativa’ Adaptado de Widdowson (1995)

Para Widdowson (*op. cit.*), la formulación que hacen Bialystok y Sharwood-Smith es la misma que la de Hymes, pero en términos distintos. Y para aclarar conceptos, Widdowson indica que la competencia está formada por dos elementos básicos: (i) conocimiento y (ii) habilidad. Y estos dos elementos los relaciona con los parámetros que propone Hymes (*op. cit.*) en su teoría y los reordena en: (i) competencia gramatical (parámetro de la posibilidad) y (ii) competencia pragmática (parámetros de factibilidad, adecuación y realización). Así, para él, el conocimiento se puede “caracterizar en términos de grado de accesibilidad” (Widdowson, 1995: 87). Partiendo de esta idea, Widdowson presenta lo que él considera la diferencia básica entre el enfoque estructural y el comunicativo. El primero se centra en el análisis, dejando de lado lo que es el *acceso*; mientras que el enfoque comunicativo focaliza en el acceso a expensas del análisis. Este nuevo punto de vista permite, según Widdowson, nuevas formas de acercarse a la enseñanza de segundas lenguas.

Un ejemplo de estas nuevas tendencias lo encontramos en Canale, que en 1983 habla de competencia comunicativa, (i) por un lado como el conocimiento que el sujeto posee de la lengua y de aspectos del uso comunicativo de la misma y, por otro lado, (ii) de la habilidad para utilizar ese conocimiento (Canale, 1983). Canale y Swain ya en 1980 realizan un primer acercamiento a lo que ellos entienden por ‘competencia comunicativa’ y en 1983, Canale da forma a su teoría al incluir dentro del concepto de competencia comunicativa cuatro áreas de conocimiento y habilidad como son: (i) **la competencia gramatical**, que implica el dominio del código lingüístico, tanto verbal como el no verbal; (ii) **la competencia sociolingüística** –adecuación de los enunciados al significado y a la forma-; (iii) **la**

**competencia discursiva**, que consiste en combinar las formas gramaticales con los significados para conseguir un texto escrito o hablado coherente y cohesionado, y (iv) **la competencia estratégica**, que implica el dominio de las estrategias de comunicación verbal y no verbal para compensar deficiencias en alguna de las competencias comunicativas, y/o para facilitar la comunicación.

### ***1.6.1 Enfoque comunicativo: ¿Cómo enseñar la lengua meta?***

Canale (1983) parte de este marco teórico para aplicar en la enseñanza de segundas lenguas el enfoque comunicativo. Según él, el enfoque comunicativo es un enfoque integrador mediante el cual se pretende capacitar y animar al alumno de L2/LE a ser partícipe de situaciones reales y significativas de comunicación. Dentro de este contexto que presenta Canale, se entienden las palabras de Williams y Burden (1999) cuando dicen que:

podemos comenzar a ver en el interaccionismo social un apoyo teórico muy necesario para el enfoque comunicativo de la enseñanza de idiomas, pues en él se afirma que aprendemos un idioma mediante el uso que hacemos de la lengua para interactuar de manera significativa con otras personas. (Williams y Burden, 1999: 48)

Así, teniendo como base teórica el interaccionismo social y el concepto de competencia comunicativa, como indican Williams y Burden (*op. cit.*), ya en 1980, Canale y Swain señalan cinco directrices básicas que debe incluir el enfoque comunicativo:

- (i) Extensión de las áreas de competencias: el objetivo del enfoque comunicativo debe ser facilitar al aprendiente la integración de las cuatro competencias que componen la competencia comunicativa y sin que ninguna de ellas destaque sobre el resto. Estas competencias son: (a) la gramatical, (b) la sociolingüística, (c) la discursiva y (d) la estratégica.
- (ii) Necesidades de comunicación: es fundamental que los contenidos gramaticales, sociolingüísticos, discursivos y estratégicos de la lengua meta se adapten a las necesidades e intereses de los aprendientes.
- (iii) Interacción significativa y realista: el aprendiente debe tener oportunidades para interactuar con hablantes competentes en situaciones reales significativas.
- (iv) Las habilidades del aprendiente en su lengua nativa: es importante que el aprendiente, sobre todo en los primeros estadios de su aprendizaje de la lengua meta, pueda recurrir a las habilidades comunicativas adquiridas en su L1.

- (v) Enfoque a través del currículo: se resalta la importancia, en un programa de enseñanza de segundas lenguas, de proporcionar a los aprendientes los conocimientos, la práctica y la experiencia necesarias para cubrir sus necesidades comunicativas. También señala Canale, la necesidad de enseñar la cultura de la lengua meta.

Siguiendo en esta línea, Van Ek (1986) amplía la clasificación de Canale y Swain añadiendo (i) **la competencia sociocultural**, que refiere a la interrelación existente entre una lengua y unos parámetros de conocimiento propios de una determinada sociedad. Esos parámetros pueden ser diferentes en una y otra lengua, como convenciones sociales o referencias culturales. Y (ii) **la competencia social**, que consiste en controlar las habilidades sociales básicas para poder participar de una comunicación adecuada gracias al dominio de la afectividad o el respeto.

A diferencia de Van Ek (*op. cit.*), que considera la competencia sociocultural como algo independiente, hay autores que la relacionan con la competencia discursiva o la sociolingüística. No obstante, todos le otorgan un puesto relevante en la enseñanza de segundas lenguas (Bachman, 1990; Canale, 1983).

Este nuevo punto de vista hace replantear el enfoque sobre el tema de la enseñanza de lenguas pues queda patente que el conocimiento lingüístico exclusivamente es insuficiente para participar de una comunicación adecuada a los diferentes registros y situaciones a los que se puede enfrentar un hablante no nativo (Cenoz, 1996). Así, autores como Barcia (2002) destacan la importancia de que el aprendiente sea capaz de adquirir la ‘competencia sociocultural’ para, de esta manera, poder llegar a entender de qué manera funciona la sociedad de la lengua meta que está aprendiendo. Es decir, comprender, en definitiva, los implícitos culturales de dicha sociedad.

Bachman (1990) da un paso más y habla de la ‘Habilidad Lingüística Comunicativa’ (HLC). Para Bachman, la HLC está compuesta por tres elementos básicos: (i) **competencia lingüística**, compuesta por elementos de conocimiento específico que se utilizan en la comunicación a través del lenguaje; (ii) **competencia estratégica**, son aquellos componentes necesarios para poder relacionar la competencia de la lengua con el contexto, es decir, con el conocimiento sociocultural o conocimiento del mundo, y (iii) **componentes psicológicos**, son los componentes neurológicos y psicológicos que forman parte ineludible del proceso comunicativo, como la luz o el sonido.

Bachman (*op. cit.*) resume su concepto de competencia lingüística o competencia de la lengua en un esquema en el que pone de manifiesto las relaciones jerárquicas que se establecen entre los distintos elementos que conforman la competencia lingüística (CL).

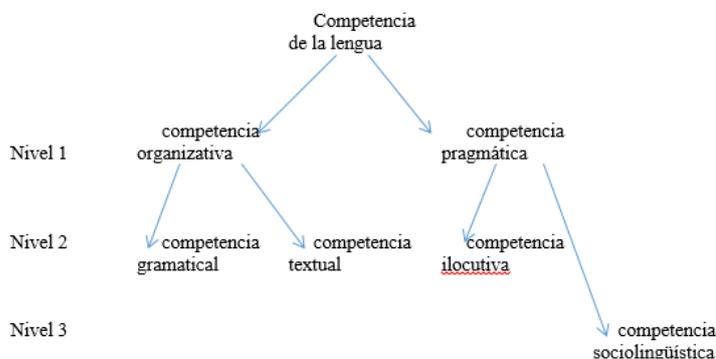


Figura 1. Competencia lingüística. Relaciones jerárquicas Fuente: Adaptado de Bachman (1990)

Sin embargo, Bachman (*op. cit.*) aclara que esos componentes no actúan de forma independiente, sino que están íntimamente relacionados e interactúan entre ellos y junto con el contexto para facilitar el acto comunicativo. Así, en un primer nivel, hace depender de la CL la competencia organizativa (CO), y la competencia pragmática (CP). En el segundo nivel, de la CO depende la competencia gramatical (CG: morfología, sintaxis, vocabulario, etc.) y la competencia textual (CT: cohesión y organización retórica). En ese mismo nivel se encuentra la competencia ilocutiva (CI: son las convenciones pragmáticas que hacen posible la comunicación) que depende de la CP. Y en un tercer nivel coloca la competencia sociolingüística (CS: aquellos componentes que permiten la comunicación dentro de un contexto determinado) que la hace depender directamente de la CP.

En la actualidad, el estudio de la pragmática en el ámbito de la enseñanza de una L2/LE se ve como algo esencial porque eso hace que el aprendiente tome conciencia de la existencia de toda una serie de códigos lingüísticos y culturales propios de cada sociedad (Barcia, 2002; Corros, 2005; Estévez y Fernández, 2006).

Con toda esta base teórica sobre la ASL y su reflejo en el ámbito de la enseñanza de segundas lenguas, desde organismos europeos se animó a los investigadores a crear documentos que promovieran la implementación de nuevos programas didácticos. De esta manera surgen el Marco Común Europeo de Referencia (MCER) en 2001; el Plan Curricular del Instituto Cervantes (PCIC) en 2006 o el Marco de Referencia para los Enfoques Plurales de las Lenguas y de las Culturas (MAREP) en 2007. Todos ellos sirven de referencia a las

instituciones educativas y a los docentes para llevar al aula e implementar los postulados del enfoque comunicativo.

### ***1.6.2 El enfoque transcultural: ¿Qué lengua enseñar?***

La evolución en las investigaciones en torno a la enseñanza de segundas lenguas sigue el camino de resaltar la importancia de un buen conocimiento no sólo de la lengua meta, sino también de la cultura para poder interactuar adecuadamente en cualquier situación comunicativa real. Según Byram y Fleming (2001), un rasgo importante del proceso de aprendizaje de cualquier idioma radica en el conocimiento adecuado de la cultura que rodea la lengua meta. También García García (2004: 3) señala la importancia de seguir un enfoque intercultural dentro del aula:

(...) posibilita el análisis de contenidos culturales, valores, creencias e ideas intrínsecas en el aprendizaje de un idioma, atiende a factores afectivos, cognitivos y situacionales y fomenta la competencia intercultural con la que el individuo puede desarrollar la convivencia y las habilidades culturales.

Sin embargo, hay diferentes maneras de acercarse a la enseñanza de la cultura de una sociedad. Así, Risager (2001) habla, por ejemplo, de cuatro enfoques diferentes para la enseñanza de la L2/LE según la perspectiva desde la que el aprendiente se acerca a la cultura de la lengua meta: (i) enfoque de la cultura extranjera; (ii) enfoque intercultural; (iii) enfoque multicultural, y (iv) enfoque transcultural.

- (i) En el primer enfoque, el de la **cultura extranjera**, la enseñanza se centra en la cultura del país de la lengua meta, sin tener en consideración la cultura del país del aprendiente. El objetivo es el de fomentar en los aprendientes la competencia comunicativa y cultural de manera que se aproxime a la competencia del hablante nativo. Este enfoque, ya desde los años ochenta, ha quedado en un segundo plano, recibiendo bastantes críticas por dejar de lado las relaciones entre las diferentes culturas.
- (ii) El **enfoque intercultural**, en cambio, sí tiene presente en su enseñanza las distintas culturas relacionadas, la de la lengua meta y otras afines, como también la del aprendiente. Se estudia las diferencias o semejanzas entre ellas y esto ayuda al aprendiente a reflexionar acerca de su propia lengua y cultura.

- (iii) El **enfoque multicultural**, da un paso más, y ofrece una concepción más amplia de la cultura del país objeto de estudio. Se presta especial atención a la diversidad étnica y lingüística no sólo del país de la lengua meta, sino también de la variedad existente en el país del aprendiente. El objetivo de este enfoque es que el aprendiente adquiera una competencia intercultural y comunicativa que les permita interactuar con cualquier hablante del país receptor, ya sea su L1 la lengua meta u otra.
- (iv) El **enfoque transcultural** presta atención a los individuos y los grupos en entornos cultural y lingüísticamente complejos, en los que el aprendiente es consciente de que su interlocutor habla la lengua meta, pero con un sustrato de otra lengua. Esto es así, por un lado, gracias a la facilidad que existe actualmente de viajar o trasladarse de un lugar a otro para estudiar, trabajar, hacer turismo, etc. Por otro lado, los movimientos migratorios de los últimos años han favorecido también la proliferación de sociedades multiculturales en la que los hablantes de esa sociedad no tienen como L1 o lengua dominante la lengua meta. En un trabajo más reciente, Kramsch y Whiteside (2007: 911) vuelven a insistir en esa misma idea:

In the last 10 years, homogeneous speech communities made up of monolingual, monoculture nationals have become less and less of a reality. (...) Many of the world's NSs do not speak the standard language, which is in many cases an ex-colonial language, nor do they speak only one language.

### ***1.6.3 La lengua meta a enseñar***

Hace treinta años era incuestionable que el modelo a seguir era el que marcaba el hablante nativo, él era el indicador de la norma. Ser oriundo del país de la lengua meta le proporcionaba autoridad y legitimidad en el uso de dicha lengua. Sin embargo, el nuevo contexto que acabamos de describir ha hecho replantearse a los investigadores, por un lado, el concepto de hablante nativo y, por otro lado, que se empiece a reflexionar sobre cuál debe ser el modelo a imitar o si es igual de válido como modelo de hablante nativo un niño como un adulto, un profesor universitario como un trabajador (Byram 1997, Byram y Fleming 2001, Kramsch 2001).

Kramsch (2001) intenta ofrecer una nueva visión de este concepto a partir de la revisión que realiza de los privilegios que tradicionalmente se le han asignado al concepto

de hablante nativo: (i) autoridad por nacimiento; (ii) derecho adquirido a través de la educación, y (iii) prerrogativa por pertenencia a un grupo social.

Kramersch empieza su revisión negando que ser hablante nativo implique un privilegio de nacimiento, ya que ni nacer en el lugar donde se habla un determinado idioma, ni nacer dentro de una comunidad con una lengua propia, asegura hablar esa lengua. Así, se pregunta si el hablante nativo en lugar de nacer se hace, es decir, si adquiere tal estatus a través de la educación. En este punto, surge la duda de cuánta instrucción, y de qué tipo, es necesaria para considerar que un hablante puede comunicarse socialmente de forma aceptable. De ahí aparece el concepto de ‘hablante nativo maduro y educado’ que intenta dar respuesta a esos interrogantes. Kramersch (*op. cit.*) considera que al definir al hablante nativo como resultado de la educación convierte el privilegio de nacimiento en un derecho educativo. Es decir, no es el privilegio de haber nacido en un lugar donde se habla un idioma determinado lo que confiere el estatus de hablante nativo, sino el haber sido socializado y escolarizado dentro de una sociedad y en un sistema educativo concreto lo que permite hablar de ‘hablante nativo de clase media y normal’.

El tercer punto que apuntaba Kramersch (*op. cit.*) era la prerrogativa por pertenecer a un grupo social. Aquí señala que la distinción entre nativo/no nativo se relaciona con la distinción nosotros/ellos y se vincula con el sentido legal de cultura nacional. Así pues, para ser considerado como hablante nativo no es suficiente con demostrar que se posee una competencia comunicativa adecuada, sino que es necesario que la comunidad de hablantes de la lengua meta considere que mereces ser considerado como tal.

Sin embargo, el dar una definición de hablante nativo no es tan simple. Kramersch (*op. cit.*) menciona el hecho de que cada vez hay una mayor variedad geográfica entre hablantes no nativos, al tiempo que aumenta la diversificación del uso lingüístico entre los propios hablantes nativos, para explicar la dificultad en definir con claridad quién es o no hablante nativo. Pone el ejemplo del idioma inglés y dice que no todos los hablantes nativos ingleses hablan como anglosajones blancos de clase media alta. Por tanto, ¿cuál debe ser el modelo que se debe enseñar? Hasta ahora, dentro de una enseñanza comunicativa, se defendía la búsqueda de la autenticidad, pero ante la duda de qué es lo realmente auténtico, según Kramersch (*op. cit.*), los profesores de idiomas se han planteado si no sería más acertado sustituir la pedagogía de lo auténtico por una pedagogía que abogue por lo que es más adecuado en cada momento. No es posible tener un solo modelo de uso lingüístico como única norma válida. El hablante debe ser capaz de adaptarse y seleccionar las formas

correctas y más adecuadas según el contexto social. Kramersch (*op. cit.*) considera que este tipo de competencia es la que debe poseer un hablante intercultural.

Kramersch y Whiteside (2007) revisan nuevamente el concepto de qué es ‘ser nativo’, junto con los de ‘no nativo’ e ‘interlengua’ como los tres conceptos básicos en ASL y concluyen que en el contexto actual es difícil valorar si estos términos siguen siendo válidos tal y como se entendían años atrás. Insisten en la idea de que ya no es posible hablar de un ‘hablante nativo’ como único modelo de lengua:

Are these concepts still valid in our present world of multilingual migrants, where everyone is more or less native and nonnative, more or less novice and expert, more or less a capable peer? where it seems less important to learn a full linguistic system than to participate as a legitimate social actor in multiple contexts of social life, verbally and nonverbally? where learning and using have become co-extensive?

(...) that native speakers are not only individuals who speak a given language from birth, but people who have learned other languages, dialects, or sociolects along the way, and who may not speak any standard variety at all. (Kramersch y Whiteside, 2007: 917)

De esta nueva perspectiva de entender qué es ser nativo y qué lengua meta enseñar, nace la necesidad y el interés por los corpus formados por muestras lo más amplias posibles de la lengua meta producida tanto por nativos como por no nativos y, así, poder indagar qué patrones rigen una y otra.

A partir de este tipo de investigaciones, el docente podrá acercar la compleja realidad social al aula, trabajando con muestras reales, tanto escritas como orales, y proporcionar al aprendiente el mayor número posible de situaciones cercanas a la realidad de la sociedad de la lengua meta (Kramersch y Whiteside, 2007; Gómez del Estal, 2008).

## **1.7 La enseñanza mediante tareas**

### ***1.7.1 Definición de tarea***

El concepto *tarea* se ha definido desde diferentes perspectivas. Así, por ejemplo Long (1987) se refiere a tarea como todo aquello que una persona puede realizar en su vida cotidiana, en el ámbito de su lugar de trabajo, de ocio, etc. Ello implica un abanico muy amplio de tipo de tareas, que puede ir desde comprar un producto, hacer un examen hasta rellenar un formulario. Una definición más relacionada con el ámbito de la didáctica de

lenguas es la que propone Breen (1987). Este autor concibe la tarea como un conjunto de actividades orientadas a facilitar el aprendizaje de una lengua. Según él, dichas actividades pueden consistir tanto en ejercicios sencillos y breves como en actividades que conlleve en sí la resolución de problemas o la toma de decisiones por parte de un grupo. Nunan (1989; 2006), centrándose también como Breen en el ámbito de la didáctica, introduce el término comunicativo en su definición de tarea. Él habla de tarea comunicativa como parte del trabajo que se desarrolla dentro del aula. Textualmente dice:

(...) a task is a piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form. The task should also have a sense of completeness, being able to stand alone as a communicative act in its own right with a beginning, a middle and an end. (Nunan, 2006: 17)

Aunque señala la primacía de la atención al significado frente a la forma, más adelante aclara que ambos elementos están interrelacionados: “My own definition refers to the deployment of grammatical knowledge to express meaning, highlighting the fact that meaning and form are highly interrelated, and that grammar exists to enable the language user to express different communicative meanings.” (Nunan, 2006: 17)

El Marco Común de Referencia (2001) en su capítulo dos (2.1.5) dice: “La comunicación y el aprendizaje suponen la realización de tareas que no son sólo de carácter lingüístico, aunque conlleven actividades de lengua y requieran de la competencia comunicativa del individuo”. Más adelante, en el punto 7.1 las define como:

Las tareas pueden tener un carácter enormemente variado y pueden comprender actividades de la lengua en mayor o menor medida como, por ejemplo: actividades creativas (pintar, escribir historias), actividades basadas en destrezas (reparar o ensamblar algo), resolución de problemas (rompecabezas, crucigramas), transacciones habituales, interpretación de un papel en una obra de teatro, participación en un debate, presentaciones, planificación de una acción, lectura y respuesta a un mensaje (de correo electrónico), etc.

Ortega (2004: 17), siguiendo en la línea del MCER, dice:

Las tareas pueden ser de muchos tipos, como por ejemplo, narraciones, debates, juegos de construcción o interpretación de gráficos. Se definen por ser actividades que se orientan a la consecución de metas y productos no lingüísticos, y por lo tanto se diferencian, al menos en teoría, de las actividades comunicativas tradicionales.

Martín Peris (2004), igual que Breen (*op. cit.*) y Nunan (*op. cit.*), aporta una definición de cariz didáctico señalando que tarea es cualquier tipo de actividad de uso de la lengua que se realice dentro del aula para promover el aprendizaje de la L2/LE, pero que al mismo

tiempo sean representativas de la vida real. Estas tareas deben cumplir con una serie de requisitos como (i) que estén estructuradas pedagógicamente; (ii) que permitan la intervención activa y modificaciones por parte de los sujetos que participan de las mismas; (iii) que los aprendientes focalicen en el contenido de los mensajes; (iv) que permitan, asimismo, la atención a la forma.

### ***1.7.2 Ventajas de trabajar mediante tareas***

Son muchos los autores que reivindican el uso de la tarea, por un lado, como elemento organizador del trabajo dentro del aula y, por otro lado, de una manera de proporcionar adquisición. Así, Candlin, ya en 1987, hablaba de la conveniencia de trabajar con *tareas* como unidades de estudio y no a partir de elementos lingüísticos como estructuras gramaticales, listas de vocabulario, etc.

Como indicaba recientemente Robinson (2011), y Alba y Zanón (1999) y Foster y Skehan (1999) en las postrimerías del siglo pasado, en los últimos 30 años han sido muchos los estudios que se han realizado en torno a este tema. Foster y Skehan (*op. cit.*) señalan los beneficios que estos trabajos<sup>18</sup> han encontrado en el uso de las tareas: “These studies have indicated that pre-task planning can have beneficial effects upon the nature of task performance, consistently leading to greater fluency and complexity and, less dependably, greater accuracy.” (Foster y Skehan, 1999: 215)

Organismos oficiales como el Instituto Cervantes o el Consejo de Europa han mostrado, también, interés en fomentar la enseñanza mediante tareas. Ya en 1994 el Instituto Cervantes dio a conocer su Plan Curricular, en el cual, además de establecer los objetivos generales del currículo, también se indicaban pautas generales de carácter metodológico y criterios para coordinar las decisiones en materia de evaluación en los diferentes centros Cervantes. Ya en los inicios de este siglo, en el año 2001, el Consejo de Europa dio a conocer el Marco Común Europeo de Referencia (MCER) para las lenguas en el que se describe de forma detallada las características principales de la enseñanza mediante tareas y el papel que juega dentro de la enseñanza de idiomas:

Una tarea puede ser muy sencilla o muy complicada (por ejemplo: estudiar varios diagramas e instrucciones que guarden relación y ensamblar un aparato desconocido y complejo). Una tarea concreta puede suponer un mayor o menor número de pasos o de

---

<sup>18</sup> Foster y Skehan (1999) señalan los trabajos de Crookes (1989) y Foster y Skehan (1996).

subtareas incorporadas. Como consecuencia de ello, los límites de cada tarea pueden ser difíciles de definir.

(...)

Estas tareas de la «vida real», «finales» o «de ensayo» se eligen según las necesidades que tienen los alumnos fuera del aula, ya sea en los ámbitos personal y público, ya sea en relación con necesidades más específicas de carácter profesional o educativo.

(MCER, Marco común europeo para las lenguas, Cap. 7, 2001)

También aconseja cómo se puede afrontar la tarea para facilitar su realización: “El éxito en la realización de la tarea se puede facilitar mediante la activación previa de las competencias del alumno, por ejemplo, en la fase inicial de planteamiento del problema o de definición del objetivo de la tarea” (MCER, 2001)

Nunan (2006: 14) indica que a nivel pedagógico las tareas han ayudado a fortalecer algunos principios y prácticas:

- A needs-based approach to content selection.
- An emphasis on learning to communicate through interaction in the target language.
- The introduction of authentic texts into the learning situation.
- The provision of opportunities for learners to focus, not only on language, but also on the learning process itself.
- An enhancement of the learner’s own personal experiences as important contributing elements to classroom learning.
- The linking of classroom language learning with language use outside the classroom.

En esta misma línea sobre las bondades de la enseñanza mediante tareas, el trabajo de Robinson (2011) presenta un resumen de las principales ventajas de la enseñanza mediante tareas. Según el autor, este tipo de enseñanza (1) proporciona un contexto en el que el aprendiente puede negociar y comprender el significado del *input* presente en la tarea o la lengua que utiliza el compañero que realiza la misma tarea; (2) brinda oportunidades de interiorización (*uptake*) de retroalimentación correctiva, implícita o explícita, proporcionada por un compañero de trabajo o profesor; (3) aporta *input* adaptado o resaltado que puede ser básico para la buena realización de la tarea; (4) ofrece situaciones para identificar, por un lado, la distancia entre el *input* proporcionado y la producción y, por otro lado, posibilita la realización de una reflexión metalingüística; (5) puede requerir la atención en conceptos gramaticales concretos necesarios para una buena expresión en la L2/LE y de esta manera

mejorar la producción; (6) puede focalizar en tareas simples que permiten la automatización y acceso rápido a la interlengua para solucionar la tarea requerida, permitiendo, así, una mejora en la fluidez; (7) brinda la oportunidad de centrar la atención en una secuencia que favorezca la consolidación de esfuerzos anteriores llevados a cabo con éxito; (8) proporciona la posibilidad, en tareas complejas, de que el *output* del aprendiente sea también más complejo y ambicioso. Todo ello, siempre que se dé dentro de un contexto comunicativo, favorece la motivación y, por tanto, el aprendizaje.

### ***1.7.3 Diseño de las tareas***

Igual que hay diferentes interpretaciones de lo que es una tarea, como acabamos de ver, también los autores proponen modelos distintos para llegar a diseñarla.

Candlin (1987) apuntaba ciertos elementos que él consideraba como básicos para que la enseñanza mediante tareas favoreciera el aprendizaje. Según Candlin, cualquier diseñador de tareas o docente tiene que tener presente los siguientes elementos:

1. *Input*: El *input* escrito, oral o visual que el aprendiente recibe y con el que tendrá que trabajar para alcanzar el objetivo de la tarea.
2. *Roles*: El *rol* que el aprendiente adopta durante la tarea, ya sea como receptor o como emisor.
3. *Settings*: El tipo de *agrupación*, dentro o fuera del aula, que va a ayudar al aprendiente a prepararse para comunicar.
4. *Actions*: Las *acciones*, es decir, los procedimientos o pasos que el aprendiente tiene que seguir para realizar la tarea.
5. *Monitoring*: La supervisión del proceso para asegurarse que la tarea se lleva a cabo tal y como se había planeado.
6. *Outcomes*: Los resultados orales, escritos o de comportamiento donde la tarea es el resultado esperado.
7. *Feedback*: Incluye la evaluación total o parcial de la actuación por parte del profesor o de otros compañeros; también incluye retroalimentación correctiva sobre el uso de la lengua.

A partir de estas premisas sobre lo que se debe tener presente a la hora de diseñar tareas, podemos encontrar planteamientos tan simples como el que presenta Wright (1987), con sólo dos componentes básicos en la estructura: (i) la información de entrada, es decir, los materiales, y (ii) una pregunta iniciadora que sirva a los alumnos como guía de lo que deben hacer durante el proceso de aprendizaje. Para Wright ni los objetivos ni los resultados son elementos obligatorios ya que, según él, los resultados, por ejemplo, pueden ser diferentes a los que el docente había previsto durante la preparación de la tarea.

Pero también hay modelos en los que es posible encontrar hasta seis o más elementos básicos dentro de la estructura de la tarea. Shavelson y Stern (1981) proponen una implementación de tarea con seis componentes: (1) contenido, (2) materiales, (3) actividades, (4) metas, (5) alumnado y (6) comunidad social.

Long (1997) propone siete pasos para el diseño de una tarea:

1. Identificar las tareas meta. Estas pueden ser de diferente índole, como por ejemplo comprar un billete de autobús, preguntar direcciones, leer un menú, escribir un informe de laboratorio, etc.
2. Clasificar y agrupar las tareas meta por categorías, como por ejemplo hacer o cambiar reservas.
3. Crear tareas pedagógicas, es decir, el material, las actividades que se van a trabajar en clase. Éstas se deben adaptar a las características propias del grupo al que van dirigidas.
4. Crear un currículo a partir de la secuenciación de las actividades. Este currículo puede ser tanto sintético como analítico.
5. Implementar la tarea con una metodología y pedagogía apropiadas. Long cita el *feedback* y también la atención a la forma como ejemplos de principios pedagógicos que el docente tiene presente en momentos determinados del proceso de aprendizaje.
6. Evaluar con criterios de referencia o exámenes.
7. Evaluar el programa.

Otros ejemplos de posibles estructuras para implementar una tarea los encontramos en Martín Peris (2004) y Nunan (1989). Ambos coinciden en presentar un modelo basado en

sólo tres pasos. Mientras que para Nunan los elementos son: (i) objetivos, (ii) *input* y (iii) actividades; para Martín Peris son: (i) objetivo, (ii) estructura y (iii) secuencia. Según Martín Peris (*op. cit.*) estos tres elementos surgen de dos ejes básicos: (1) la creación de un contexto en el cual los aprendientes pueden poner en práctica la lengua meta; (2) la asignación de una meta de carácter extralingüístico que va a determinar los objetivos de comunicación que los aprendientes deben alcanzar. Como se puede ver, ambas propuestas son muy similares.

Vamos a enumerar a continuación las características de ambas propuestas.

Para Nunan (*op. cit.*):

- (i) los **objetivos** vienen determinados, aunque de forma muy general, según la meta que el docente pretenda alcanzar con la tarea. Nunan pone como ejemplo de tipos de objetivos generales que se pueden pretender conseguir en una tarea los propuestos por el Australian Language Levels Project: (1) comunicativo – establecer y mantener relaciones interpersonales-; (2) sociocultural –llegar a conocer, aunque de forma general, la vida cotidiana del grupo de la lengua meta; (3) aprender a aprender y (4) concienciación de la lengua y la cultura – llegar a tener cierto conocimiento del sistema de la lengua y cómo funciona.

Nunan (*op. cit.*) también diferencia entre objetivos generales y específicos. Así, al referirse a los objetivos a conseguir en la destreza de expresión escrita, por ejemplo, distingue entre los objetivos que pretenden desarrollar las habilidades lingüísticas funcionales básicas –escribir notas, listas de compra o postales- y aquellos que van dirigidos a desarrollar las habilidades de escritura más formal –escribir un ensayo, un informe o una carta de negocios-.

- (ii) el *input* es el material que se suministra al aprendiente y que conforma el punto de partida de la tarea. El origen del *input* puede ser de muy distinta índole: cartas formales o informales, extractos periodísticos, formularios, recetas, gráficos de economía, etc.

- (iii) en las **actividades** se especifica lo que el aprendiente tendrá que hacer con el *input* recibido y que supone el punto de partida de la tarea de aprendizaje.

Nunan clasifica las actividades en tres grandes grupos, según sus características:

- (1) Actividades de *ensayo para la vida real*. Aquí la disyuntiva se establece entre si las actividades deben ser actividades representativas del mundo real o más bien pedagógicas. Nunan (*op. cit.*) señala que los autores no se ponen de acuerdo y es posible encontrar los que defienden que las actividades deben estar lo más cerca posible del mundo real, mientras otros piensan que la función de la metodología es proporcionar al aprendiente una competencia comunicativa, pero para llegar a ello no consideran que la autenticidad de las actividades sea un requisito indispensable.
- (2) Actividades de *uso de las destrezas*. Nunan distingue entre (i) actividades controladas y (ii) de transferencia. Las primeras van dirigidas a que el aprendiente practique formas fonológicas y gramaticales, mientras que en las segundas el alumno debe poner en práctica, a través de la comprensión y la producción de un lenguaje comunicativo, todo aquello que ha adquirido.
- (3) Actividades para adquirir *fluidez y corrección*. Las actividades pueden ir orientadas (i) al desarrollo de la corrección –control de unas reglas- o (ii) al desarrollo de la fluidez –uso de la lengua-. Según Nunan (*op. cit.*), hay autores para los que esta distinción no significa que se trate de conceptos contrarios, sino complementarios. Aunque Nunan reconoce que muchos de los materiales existentes se han diseñado manteniendo la disyuntiva entre fluidez y corrección.

Para Martín Peris (2004):

- (i) el **objetivo** se establece en términos de aprendizaje y forman parte de un currículo. Según el autor, no se debe confundir *objetivo* con *meta*. La meta es de carácter extralingüístico.
- (ii) en la **estructura** se integra el elemento principal, el **producto**, en torno al cual giran los demás componentes de la estructura:
  - 1. Las actividades previas: son las que van a capacitar al aprendiente para la elaboración del producto. Los alumnos, en esta fase, van a trabajar con las estructuras y unidades lingüísticas que les van a ser útiles para la obtención del producto y las van a practicar en actividades de uso de la lengua meta.

2. Las fuentes: es el material a partir del cual se generarán las diferentes actividades. Las fuentes pueden ser un texto, un fragmento musical, un cuadro de pintura, etc.
3. La evaluación: se aplica tanto al proceso como a los componentes de la tarea. También hay una evaluación al final para comprobar que se han alcanzado los objetivos en el aprendizaje. Esta evaluación no sólo se va a limitar a identificar las carencias o problemas en el desarrollo de la tarea, también va a evaluar el interés mostrado por los aprendientes acerca de los diferentes contenidos de aprendizaje, su profundidad y extensión. Esto va a permitir que el proceso de aprendizaje continúe con las actividades derivadas.
4. Las actividades derivadas: tanto por el formato como por el contenido, son iguales a las actividades previas, pero, al contrario que estas, no son para capacitar, sino para reforzar, enriquecer o consolidar el proceso de aprendizaje una vez identificadas las dificultades e intereses de los aprendientes.
5. Los agentes: son las diferentes agrupaciones de alumnos que las actividades exigen en cada momento, es decir, en agentes individuales, parejas, pequeños grupos o grupo clase.
6. Secuencia. Se concreta en tres momentos:
  - (a) El planteamiento (o pre-tarea): consiste en proporcionar al aprendiente todos los datos necesarios para la correcta comprensión y realización de la tarea.
  - (b) La preparación y ejecución: aquí la atención del alumno se centra no sólo en los contenidos comunicativos, sino también en los recursos lingüísticos que tiene que aprender para poder realizar con éxito la tarea.
  - (c) La post-tarea: puede estar formada por diferentes actividades como una evaluación del proceso, una presentación del producto, etc.

Como hemos visto, los elementos que conforman el diseño de una tarea pueden ser variados, se pueden especificar uno a uno o se pueden agrupar en grandes categorías. Sin embargo, todos los modelos incluyen en su planteamiento tres elementos básicos e imprescindibles: (i) definir de una manera clara y objetiva la meta de la tarea, es decir, la

tarea final o producto, (ii) elegir el currículo más adecuado para conseguir esa meta y (iii) elaborar unas actividades.

A Robinson (2011) esta diversidad le hace cuestionarse si el diseño de tareas siguiendo modelos diferentes o incluso una secuenciación distinta de las mismas puede influir en las oportunidades de aprendizaje a las que se puede ver expuesto el aprendiente. Para resolver esta duda, Robinson propone estudiar el impacto de la tarea en un aprendizaje aislado y analizar de qué manera influye en el aprendizaje la realización de una secuencia determinada o la realización de versiones diferentes de una misma tarea y señala la clasificación que algunos autores hacen para referirse a las tareas basadas en el diseño instruccional: (i) tareas de apoyo (*task supported*); (ii) tareas de referencia (*task-referenced*) y (iii) enseñanza basada en tareas (*task-based*). Asimismo, Robinson destaca el hecho de que la tarea se ha concebido como un elemento esencial para el diseño de programas de estudios desde diferentes perspectivas, como por ejemplo más gramatical o más léxica. Esto hace que se tengan que tomar decisiones sobre qué tipo de tarea se lleva al aula y en qué momento o qué orden de secuencia se presenta. Todo ello hace que las recientes investigaciones sobre el uso de las tareas giren en torno a la influencia que las características de las tareas pueden tener en el aprendizaje y el rendimiento o sobre la complejidad de las mismas (Robinson, 2011).

Skehan (1996: 38), ya en la década de los noventa, destacaba la importancia de las teorías de la psicología cognitiva a la hora de diseñar las tareas: “It renews recent developments in cognitive psychology which support a dual-mode perspective for language processing, and then proposes the goals of accuracy, complexity-restructuring, and fluency as most relevant for task-based instruction.”. Según el autor, los alumnos deben tomar conciencia y saber en cada momento cuál es la finalidad, el objetivo de lo que están realizando, es decir, qué es lo que el docente espera que ellos hagan.

## **PARTE II**

### **OBJETIVOS, PREGUNTAS DE INVESTIGACIÓN Y MARCO METODOLÓGICO**

#### **CAPÍTULO 2**

#### **OBJETIVOS, PREGUNTAS DE INVESTIGACIÓN E HIPÓTESIS**

##### **2.1 Objetivo general**

El objetivo general de este trabajo es investigar la fluidez, la diversidad léxica y la densidad léxica en dos tipos de producciones escritas por estudiantes de español como L2 (lengua dos) de nivel B1 (según el MCER) en dos momentos de su proceso de aprendizaje (inicio y final de curso) y compararlas con el estadio último de competencia nativa (grupo de control, grupo nativo).

Este objetivo general se desglosa en los siguientes objetivos específicos.

##### **2.2 Objetivos específicos**

- Caracterizar los dos tipos de expresión escrita que han producido los sujetos participantes de la investigación: (i) inventar una historia, (ii) explicar una experiencia personal.
- Identificar léxico propio de cada una de las producciones escritas en los sujetos no nativos y nativos participantes de la investigación.
- Comprobar si hay mejoría en las producciones escritas de los sujetos no nativos en cuanto a fluidez, diversidad léxica y densidad léxica entre los dos momentos de su proceso de aprendizaje (inicio y final de curso).
- Comprobar si hay diferencia en los resultados de fluidez, diversidad léxica y densidad léxica entre las producciones en las que se trabajó mediante tareas (inventar una historia) y en las que no (explicar una experiencia personal).

- Comparar los resultados de las producciones de los sujetos no nativos con las producciones de los sujetos nativos por género textual y estadios para identificar diferencias y similitudes en la fluidez, diversidad léxica y densidad léxica.

A partir del objetivo general y los objetivos específicos señalados, planteamos las siguientes preguntas de investigación.

### **2.3 Preguntas de investigación**

1. ¿Hay diferencias léxicas o sintácticas que distingan los relatos en pasado en los que se exige más creatividad (inventar una historia) frente a los que exigen cierta introspección (relatar una experiencia personal)?

2. Entre los sujetos no nativos y los sujetos nativos que participaron en la investigación, ¿hay diferencias léxicas o sintácticas entre las producciones de la muestra «F» (inventar una historia) frente a la muestra «1D» (relatar una experiencia personal)?

3. Tras un período de instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, ¿se observa progreso en cuanto a fluidez, diversidad léxica y densidad léxica en la producción escrita de los sujetos no nativos participantes de la investigación entre el M1 (inicio del curso) y el M2 (final del curso)?

4. ¿Hay diferencia en los resultados de fluidez, diversidad léxica y densidad léxica en aquellas producciones que se trabajaron siguiendo una enseñanza mediante tareas (inventar una historia) frente a las que no (explicar una experiencia personal)?

5. En las producciones escritas de los sujetos no nativos y los sujetos nativos, ¿hay diferencia en cuanto a fluidez, diversidad léxica y densidad léxica entre las producciones en las que deben inventar una historia y en las que deben explicar una experiencia personal?

Planteados los objetivos y las preguntas de investigación, a continuación presentamos las hipótesis de las que partimos.

## 2.4 Hipótesis de partida

- Consideramos que entre un relato en el que se exige cierto grado de creatividad, como es inventar una historia, y un relato en el que hay que explicar una experiencia personal, por tanto, es necesario recurrir a recuerdos y hacer un acto de introspección, hay diferencias a nivel de léxico y estructura que afectan a la fluidez, la diversidad léxica y la densidad léxica.
- Pensamos que las producciones escritas de las muestras «F» y «1D» de los sujetos no nativos y nativos van a presentar diferencias en el léxico y en las estructuras sintácticas.
- En un proceso de aprendizaje formal dentro de un programa general de español como L2 siguiendo el enfoque comunicativo y una enseñanza mediante tareas, la competencia de la expresión escrita de los estudiantes va a mejorar en fluidez, diversidad léxica y densidad léxica tras una instrucción de 120 horas.
- Los resultados en fluidez, diversidad léxica y densidad léxica son mejores en las producciones que se han trabajado en el aula siguiendo una enseñanza mediante tareas (inventar una historia) que en las que no (explicar una experiencia personal).
- Los sujetos nativos van a presentar mejores resultados en fluidez, diversidad léxica y densidad léxica en las producciones escritas que los sujetos no nativos.



## CAPÍTULO 3

### MARCO METODOLÓGICO

#### 3.1 La investigación cuantitativa

En el ámbito de la adquisición de segundas lenguas (ASL) la tendencia ha sido la de realizar investigaciones cualitativas en las que ha predominado la perspectiva etnográfica. En este tipo de investigaciones el objetivo es observar qué ocurre en un contexto determinado y con características propias. Así, la investigación cualitativa se caracteriza (1) por realizar descripciones extensas y detalladas; (2) por mostrar una realidad natural y holística; (3) por el número reducido de sujetos observados en la investigación; (4) por mostrar una perspectiva émica; (5) por trabajar mediante un proceso cíclico, y (6) por proponer preguntas de investigación abiertas y generales (Larsen-Freeman y Long, 1991; Duranti, 2000; Mackey y Gass, 2005).

No obstante, en las últimas décadas se ha visto un aumento considerable en el uso de la metodología cuantitativa según un estudio realizado por Lazaraton en 2005. La autora constató en su revisión sobre el uso de dicha metodología de investigación que entre los años 1991 y 2001 el 86% de las investigaciones eran de carácter cuantitativo, mientras que solo un 13% eran cualitativas. Para Ramallo (1999), este aumento en el interés por la investigación cuantitativa se ha visto favorecido gracias al desarrollo de programas informáticos que han permitido el tratamiento de un gran número de datos con una facilidad y, al mismo tiempo, una fiabilidad que hasta hace pocos años era impensable. Este mismo autor señala que ya a principios del siglo XX un miembro de la escuela de Praga, Mathesius, realizó un acercamiento a la investigación cuantitativa al comparar el sistema fonológico de la lengua checa y alemana. Esta nueva manera de aproximarse a los estudios de segundas lenguas abrió la posibilidad a otros investigadores a realizar análisis de, por ejemplo, la extensión media de las palabras de un texto o la frecuencia de uso de un determinado fenómeno lingüístico (Ramallo, 1999). López Morales, ya en 1994, también confirmaba este interés cada vez más notorio hacia las investigaciones cuantitativas en el área de la lingüística. El autor justifica esta tendencia al considerar que la lengua está formada por “unidades discretas, segmentables, que pueden repetirse en el discurso; de ahí que sea posible contar sus apariciones.” (López Morales, 1994).

Ramallo menciona la discrepancia que existe entre seguidores de una y otra metodología de investigación. Según el autor, “lo que subyace a esta oposición es la distinción ontológica entre la cantidad (números) y la calidad (palabras)” (Ramallo, 1999). Los investigadores que defienden uno u otro tipo de metodología critican el método contrario, por ejemplo el cuantitativo, por considerarlo simplista, descontextualizado o distante y no ser capaz de identificar las características propias de los sujetos estudiados. Mientras que los que critican la metodología cualitativa argumentan la falta de validez y fiabilidad en sus investigaciones (Firth y Wagner, 1997; Ramallo, 1999).

Larsen-Freeman y Long (1991) definen la investigación cuantitativa como un experimento diseñado ad hoc con la finalidad de comprobar una hipótesis de partida sobre un fenómeno lingüístico concreto y utilizando unos instrumentos de investigación objetivos y un análisis estadístico adecuado. Siguiendo esta línea, Para Ramallo (1999), la investigación cuantitativa no debe limitarse a contar elementos, sino que debe ir más allá de una simple cuantificación para aportar “interpretaciones sobre el objeto cuantificado”.

Como se ha mencionado más arriba, el desarrollo de la informática ha contribuido a que el análisis estadístico haya pasado a ser un elemento central de los estudios cuantitativos en el campo de ASL ya que facilita no sólo la cuantificación de la ocurrencia de un determinado fenómeno, como la explicación de las relaciones entre variables, sino también la comprobación de la eficacia de la instrucción y generalización de los resultados. Es decir, se pasa de tomar los números de forma aislada a realizar inferencias sobre el tema analizado (Larsen-Freeman y Long, 1991; Ramallo, 1999).

En las investigaciones cuantitativas es importante seguir un determinado protocolo para llevar a buen fin el tratamiento de los datos. No solo hay que tener claro cuál es el objetivo de la investigación, sino que también hay que elegir la prueba estadística más adecuada. Para ello es necesario verificar que los datos obtenidos cumplen con los supuestos teóricos relacionados con cada tipo de prueba estadística (Lazaraton, 2005). Es importante tener en cuenta que cada prueba estadística necesita de unas determinadas condiciones o supuestos teóricos para que los resultados no lleven a error.

Entre las ventajas que presenta la investigación cuantitativa se suele mencionar la fortaleza que muestra en validez externa por su sistematicidad, rigurosidad, objetividad y control que sobre los datos se tiene. Además, también se indica como rasgo relevante el hecho de que la investigación cuantitativa produce resultados fiables, replicables y que se

pueden generalizar, no solo a una población, sino también a otros contextos y situaciones (Sánchez Díaz *et al.*, 2003; Dörnyei, 2007). No obstante, Lazaraton (2005) aconseja no sólo conocer las ventajas y desventajas del paradigma de investigación utilizado, también señala la importancia de circunscribir las conclusiones del estudio a las limitaciones de la metodología empleada. Esto debe ser así ya que resultados parecidos pueden ser producto de diferentes procesos subyacentes y provocar el error en la interpretación de los mismos.

La investigación cuantitativa permite mostrar dos tipos de estadística, una descriptiva y otra inferencial. La primera, como su nombre indica, describe lo que se observa a simple vista. Es decir, ofrece una aproximación básica a los resultados que se caracteriza por ordenar y clasificar los datos. Estos pueden organizarse siguiendo dos tipos de actuaciones: (i) la delimitación de la posición de los elementos analizados en relación con las medidas de tendencia central (media, mediana y modo), y (ii) la delimitación de los parámetros de dispersión, mediante el cálculo de la desviación estándar o de la varianza, por ejemplo (López Morales, 1994). Este tipo de análisis ayuda a comprobar de qué manera los datos se agrupan o se alejan respecto a un valor central (Mackey y Gass, 2005). La estadística inferencial, en cambio, busca, por un lado, medir el grado de relación existente entre las variables estudiadas y, por otro lado, sacar conclusiones y generalizarlas a toda una población partiendo de una muestra representativa de dicha población (Hernández Sampieri *et al.*, 1994; Vallejo, 2008).

El estudio que realiza Lazaraton (2005) sobre el tipo de investigación cuantitativa que se suele llevar a cabo en el ámbito de ASL muestra que la tendencia predominante es la estadística descriptiva. También Ortega y Iberri-Shea (2005) constatan esta tendencia y lo atribuyen al hecho de que suelen ser investigaciones con un número reducido de sujetos participantes, cosa que no permite realizar una estadística inferencial pues esta requiere de una muestra mínima para poder generalizar resultados a toda una población.

Sin embargo, Loewen y Gass (2009) han observado un aumento en el interés por la estadística inferencial en las investigaciones de ASL. Asimismo, también subrayan la rigurosidad en la utilización de los procedimientos estadísticos, así como el interés en la divulgación de los resultados obtenidos. Esta tendencia, favorecida sin duda por la proliferación del desarrollo de la informática (Ramallo, 1999), ha permitido, según Loewen y Gass, que durante la última década haya aumentado el número de trabajos publicados sobre ASL que utilizan métodos estadísticos frente a los que no lo hacen.

Como ya se ha mencionado anteriormente, hay autores que insisten en oponer investigación cuantitativa y cualitativa. No obstante, cada vez más, se intenta ver ambos métodos como complementarios y no contrarios, (Ramallo, 1999; Brown, 2004). También Cantos Gómez (2000), siguiendo esta idea, afirma que el uso combinado de ambos métodos favorece el análisis lingüístico y enriquece los resultados y el análisis final. Otros autores como Lazaraton (2005) y Dörnyei (2007) también abogan por la utilización de este tipo de metodología mixta de investigación ya que ayuda a acercarse al mismo fenómeno desde puntos de vista diferentes que se complementan y, al mismo tiempo, posibilitan una mejor comprensión de los datos.

### **3.2 La investigación pseudolongitudinal**

Los investigadores distinguen entre investigación transversal o longitudinal en función de los momentos de recogida de datos y del tiempo en los que se realiza dicha recogida. Es decir, si la recolección se centra en un momento concreto del proceso de enseñanza-aprendizaje de los aprendientes o si hay un seguimiento de los mismos (Hernández Sampieri. *et al.*, 1997; Gass y Selinker, 2008).

Así, en una investigación transversal el foco se pone en analizar qué es lo que sucede en una o más variantes o la relación existente entre ellas en un momento concreto de un proceso. Para ello se tiende a recolectar información del mayor número de sujetos posible para poder, de esta manera, tener una visión general del fenómeno que se desea estudiar (Hernández Sampieri. *et al.*, 1997; Gass y Selinker, 2008).

En la investigación longitudinal lo que se busca es analizar la evolución de un fenómeno determinado a lo largo de un período de tiempo. Para poder realizar ese estudio, se recopilan datos, normalmente de pocos sujetos, en momentos diferentes de su proceso de enseñanza-aprendizaje. El objetivo es acompañar a los mismos sujetos para observar qué cambios se producen y cómo evoluciona la interlengua de dichos sujetos (Hernández Sampieri. *et al.*, 1997; Han, 2004; Ellis, 2008).

En este trabajo, hemos optado por combinar características de ambos tipos de investigación y llevar a cabo un estudio pseudolongitudinal: vamos a analizar la evolución de un fenómeno a lo largo del tiempo, pero sin realizar el seguimiento de los mismos sujetos. Para ello se recogieron datos en dos momentos diferentes del proceso de enseñanza-

aprendizaje (al inicio y al final de un curso formal de nivel intermedio en una EOI) de grupos y promociones diferentes para contar con el mayor número posible de sujetos y obtener así una muestra amplia que permita realizar un análisis cuantitativo y ofrecer una panorámica de lo que sucede tras un período de instrucción formal (Gass y Selinker, 2008).

Algunos investigadores señalan como ventajas de realizar un estudio pseudolongitudinal: (i) El hecho de no realizar el seguimiento de los mismos sujetos evita la posibilidad de que disminuya el número de participantes de la investigación. Es lo que Mackey y Gass (2005) llaman “*participant mortality*”. (ii) Esta es una forma de asegurarse un número amplio de sujetos, cosa que no es posible cuando se trata de un estudio longitudinal. (iii) El disponer de un gran número de participantes en la investigación permite generalizar los resultados a la población de la muestra (Alba Quiñones, 2009).

### **3.3 Las medidas de calidad general de un texto escrito**

En este apartado vamos a comentar brevemente cuáles son las características principales de los textos escritos en el ámbito de la enseñanza - aprendizaje de segundas lenguas. Asimismo, vamos a presentar las medidas más habituales para valorar la calidad general de un texto escrito: la fluidez, la diversidad léxica y la densidad léxica.

#### ***3.3.1 La producción escrita en el ámbito de las segundas lenguas***

Aunque durante años el interés por la competencia escrita había quedado en un segundo lugar, en la actualidad se ha producido un repunte en el interés de los investigadores, como hemos visto más arriba, en saber qué sucede en el proceso o en el producto final de una producción escrita por estudiantes de segundas lenguas.

Los estudios dentro del campo de Adquisición de Segundas Lenguas han tendido a centrarse más en analizar las destrezas de producción (expresión oral o escrita) que las receptivas (comprensión oral o escrita) (Alba Quiñones, 2009). Esto se debe, según la autora, a que es más fácil compilar y cuantificar datos de una destreza productiva que de una receptiva. En este sentido, Verspoor y Smiskova (2012: 18) señalan respecto a la destreza de expresión escrita:

Moreover, in writing, more than in speaking, the learner can show better what he or she is capable of because writing allows for more reflection and is therefore usually somewhat more complex at both ideational and linguistic levels. An added bonus is that it is easier to collect and assess than spoken data.

Kellerman (2001) aconseja que se recurra a las narrativas para la compilación de datos escritos, ya que, según el autor, supone un instrumento de investigación especialmente rico al aportar datos contextualizados y ello ayuda a enfocar el análisis desde diferentes perspectivas. En esta línea, Agustín Llach (2011) hace también hincapié en el hecho de que la producción escrita se puede estudiar desde varios puntos de vista debido al carácter multidisciplinar que posee. Es decir, desde una perspectiva (i) de proceso o de producto, (ii) como actividad social o cognitiva y (iii) como una actividad focalizada en el contenido o en la forma.

Navés *et al.* (2005) señalan que la elección de la producción escrita como instrumento de investigación no supone un factor especialmente de estrés para los alumnos puesto que estos están familiarizados con este tipo de tarea; la escritura forma parte de su ámbito educativo desde la infancia.

### **3.3.2 La fluidez**

Tradicionalmente el término ‘fluidez’ se ha atribuido a la ‘velocidad’ con la que una persona es capaz de hablar. No obstante, también se utiliza para hacer referencia al número de palabras que se puede escribir en un período de tiempo determinado. Pero también se ha relacionado la ‘fluidez’ con el nivel de coherencia, complejidad o creatividad que presenta un texto, ya sea oral o escrito. Para discernir estas cuestiones, Wolfe-Quintero *et al.* (1998) señala que la solución está en analizar la fluidez de manera independiente a la diversidad y la corrección.

Una de las conclusiones a las que se han llegado los estudios centrados en la investigación del desarrollo de la escritura en el campo de ASL es que los estudiantes según van adquiriendo mayor competencia comunicativa en la lengua meta, también van aumentando la fluidez; es decir, que pueden escribir más palabras en un mismo período de tiempo y dentro de un mismo contexto que en estadios menos avanzados de competencia (Wolfe-Quintero *et al.*, *op. cit.*, Agustín Llach, 2011). Eso significa que cuanto mayor es una producción escrita (más extensa), mayor es el grado de fluidez (Jarvis *et al.*, 2003). Pero Wolfe-Quintero *et al.* (*op. cit.*) subrayan el hecho de que una mayor fluidez léxica no se

relaciona directamente con una mayor sofisticación o exactitud en el uso de vocabulario o estructuras. De ahí que insistan en delimitar lo que es fluidez de corrección y apuntan que hay diferentes maneras de medir la fluidez, contabilizando el número de frases por texto o contabilizando el número total de palabras presentes en el texto<sup>19</sup>.

Autores como Reid (1990) reivindican que aunque no se pueda afirmar que una mayor fluidez léxica, es decir, mayor extensión del texto, esté relacionada directamente con una mejor calidad general, sí se ha observado que hay correspondencia entre extensión del texto y calidad del mismo. Partiendo de esta idea, Agustín Llach (2011) señala que el léxico es un elemento que sin duda ayuda a medir la calidad de un texto, poniendo en relación la posesión de un abanico amplio de vocabulario con la calidad del texto escrito. Con la intención de establecer esa correspondencia, Engber (1995) lleva a cabo una investigación de carácter holístico sobre la calidad del texto en la que analiza, entre otras variables, la variedad léxica o la densidad léxica, llegando a la conclusión de que hay una correlación, es decir, cuánto mejor son los resultados en estas variables relacionadas con el léxico, mayor es la calidad del mismo.

### ***3.3.3 La diversidad léxica***

Hay disparidad de opiniones en cuanto a la terminología a utilizar para referirse al número de palabras distintas que podemos encontrar en un texto. Encontramos autores que utilizan ‘riqueza léxica’ y ‘diversidad léxica’ como conceptos sinónimos para referirse al total de palabras distintas de un texto, mientras que para otros la ‘diversidad léxica’ es un componente más de la ‘riqueza léxica, como la variedad léxica, la sofisticación léxica, la densidad léxica, además de un número reducido de errores (Read, 2000). Recientemente, Pallotti (2015: 125) añade un concepto nuevo, ‘complejidad léxica’ y lo define con las siguientes palabras: “(...) lexical complexity can be operationalized essentially in terms of diversity. A text with a wide variety of lexemes will be said to be more complex than one where the same few words are repeated over and over”. La definición que hace Pallotti de ‘complejidad léxica’ la podemos relacionar con la apuntada por Read (op. cit.) sobre ‘riqueza léxica’.

---

<sup>19</sup> En nuestra investigación hemos seguido esta opción, contabilizar número total de palabras, para medir la fluidez de los textos escritos.

En cuanto a los métodos para medir la ‘diversidad léxica’, también hay diferentes maneras de abordar el tema. Laufer (2003), por ejemplo, considera que la ‘diversidad léxica’ es una combinación de variedad léxica y sofisticación léxica; mientras que Malvern *et al.* (2004) recurren al concepto ‘diversidad léxica’ para referirse al número total de palabras diferentes o únicas identificadas en cada texto (*types*). En esta misma línea está Pallotti (*op. cit.*: 125) que dice: “Lexical diversity at the text level can be gauged basically by looking at type/token ratios, with subsequent refinements proposed to overcome the effects of text length”.

Duran *et al.*, (2004) subrayan, por un lado, los inconvenientes para determinar métodos matemáticos adecuados para la medición de la diversidad léxica, pero, por otro lado, apuntan que estudios empíricos han demostrado la validez del cálculo de la diversidad léxica como indicativo de desarrollo de la producción escrita.

La medida más habitual para calcular la diversidad léxica es la ratio *type/token* (RTT), que consiste en dividir el número total de palabras diferentes (*types*) entre el número total de palabras de un texto (*tokens*) (Linnarud, 1983; Pallotti, 2015). Pero el método de la RTT provoca desajustes en los resultados dependiendo del tamaño del texto. Es decir, cuanto mayor número de *tokens* haya en el texto, más posibilidades hay de que las palabras se repitan, por tanto, eso significa menor número de *types*. Eso produce una merma en los valores de la ratio, que pueden variar entre 0 (cero) y 1 (uno). Esto supone, por tanto, que los textos más extensos, es decir, con mayor número de palabras, tengan un valor más bajo de la ratio de diversidad léxica. Así, en muestras con una gran heterogeneidad de tamaños de textos resulta difícil hacer comparaciones (Vermeer, 2000; Jarvis, 2002; Pallotti, 2015).

Para evitar estos desajustes, se han sugerido diferentes alternativas: (i) contabilizar solamente las 100 primeras palabras del texto (Biber, 1988), pero el no considerar el texto como un todo arrojaba también resultados deficientes; (ii) proponer fórmulas y ecuaciones matemáticas de diferente índole con la finalidad de sortear el efecto del tamaño. Un ejemplo de ello es la fórmula que Guiraud desarrolló en 1960 y que se ha convertido en una de las medidas más ‘populares’. Su propuesta parte de la tradicional RTT, pero añade un factor más, convirtiéndola en Ratio *Type Token Cuadrada* ( $RTTC = \frac{types}{\sqrt{tokens}}$ ). Partiendo de la propuesta de Guiraud, Carroll (1964) planteó la medida de la RTT corregida (RTTC), ( $RTTC = \frac{types}{\sqrt{2tokens}}$ ). Aunque algunos investigadores como Arnaud, (1984) o Malvern *et al.* (2004) mostraron sus objeciones, pues consideraban que ese cambio no corregía el efecto

del tamaño de los textos. Jarvis (2002), en cambio, defiende que la medida RTTC de Carroll funciona bien para textos que tienen entre 1.000 y 10.000 palabras.

El desarrollo de una nueva herramienta informática (vocD) que incorporaba un fórmula matemática especialmente diseñada para llevar a cabo el cálculo de la diversidad léxica (McKee *et al.*, 2000) supuso un estímulo, por su facilidad de uso, para que los investigadores se atrevieran a calcular la diversidad léxica de sus muestras<sup>20</sup>. Esta herramienta, VocD, forma parte de la aplicación informática CHILDES (McWhinney, 2000) que se creó con la finalidad de ayudar en el estudio de la adquisición del lenguaje infantil.

A pesar de la fiabilidad que ha demostrado la medida vocD para el cálculo de la diversidad léxica, Jarvis (2002), tras comparar los resultados de varias medidas de diversidad léxica, concluyó que vocD solo era fiable en textos entre 200 y 400 palabras. Por ello, Jarvis (*op. cit.*) sugiere el uso del índice Uber U<sup>21</sup> (Dugast, 1978, 1979, 1980) para muestras en la que la extensión de los textos sea inferior a las 200 palabras.

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Figura 2. Medida de diversidad léxica Uber U

Esta fórmula presenta la ventaja de realizar un ajuste más exacto entre la RTT y la curva N (número total de palabras).

### 3.3.4 La densidad léxica

Otra medida para evaluar la calidad general de un texto es la densidad léxica (Ávila, 1986; Pallotti, 2015). La propuesta de Ávila (*op. cit.*) consiste en dividir el número de palabras léxicas<sup>22</sup> incluidas en un fragmento de texto determinado por el número total de palabras de ese mismo fragmento. Mientras que Pallotti<sup>23</sup> (*op. cit.*) sugiere otro método

---

<sup>20</sup> Ejemplo de ello están los trabajos de Richards y Malvern, 1999; Malvern y Richards, 2000; Jarvis, 2002; Strömquist *et al.*, 2002; Wright *et al.*, 2003.

<sup>21</sup> Como la extensión de los textos que conforman el corpus de nuestra tesis oscila entre las 50 y las 250 palabras, hemos optado por utilizar el índice Uber U.

<sup>22</sup> Se consideran palabras léxicas los verbos, los sustantivos, los adjetivos y los adverbios.

<sup>23</sup> En nuestra tesis hemos optado por seguir la propuesta de Pallotti (*op. cit.*); ya que los textos que conforman nuestro corpus son cortos, consideramos que es más adecuado tomar como referencia el texto completo y no de forma parcial.

consistente en calcular la proporción de palabras léxicas respecto a las funcionales<sup>24</sup> o respecto al total de palabras del texto.

---

<sup>24</sup> Se consideran palabras funcionales los artículos, las preposiciones, los pronombres y las conjunciones.

## PARTE III

### ESTUDIO Y METODOLOGÍA DE ANÁLISIS

#### CAPÍTULO 4

##### EL ESTUDIO

##### 4.1 El contexto

##### 4.1.1 *Los participantes*

Para realizar la presente investigación se seleccionaron: (i) sujetos Nativos (N) de la lengua española, estudiantes de primero del Grado en Lenguas Aplicadas de la Universidad Pompeu Fabra de Barcelona, y (ii) sujetos no Nativos (NN) de la lengua española, estudiantes de Español como Lengua Extranjera (ELE) de una Escuela Oficial de Idiomas (EOI) de Barcelona.

A continuación detallamos las características de ambos grupos.

##### 4.1.1.1 *Los sujetos Nativos*

Como ya se ha mencionado, los sujetos N eran estudiantes del *Grado en Lenguas Aplicadas* de la Universidad Pompeu Fabra. Dicho grado tiene una duración de cuatro cursos académicos y los sujetos cursaban, en el momento de la recolección de datos, el primer curso del Grado.

El total de sujetos N que participó en el estudio fue de 64 alumnos. La mayoría de ellos era del género femenino, como se puede observar en la tabla 3 y gráfico 1. Sus edades oscilaban entre los 18 y los 21 años, aunque había algún alumno que superaba esta edad, como se puede observar en el gráfico 2.

Sujetos Nativos	N
Hombres	10
Mujeres	54
<b>Total</b>	<b>64</b>

Tabla 3. Género de los sujetos de la muestra en valores absolutos.

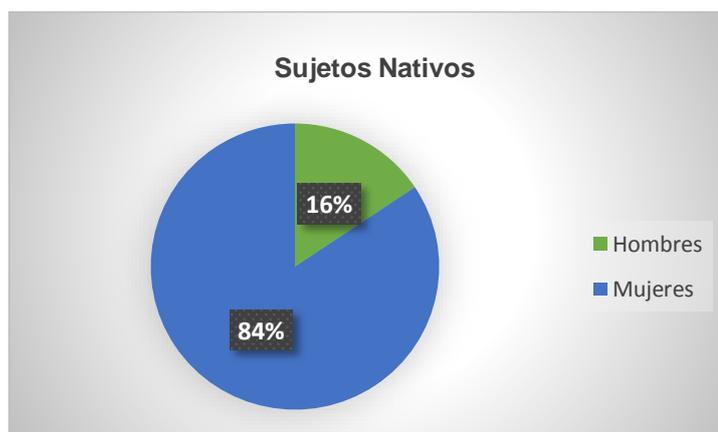


Gráfico 1. Género de los sujetos de la muestra en porcentaje.

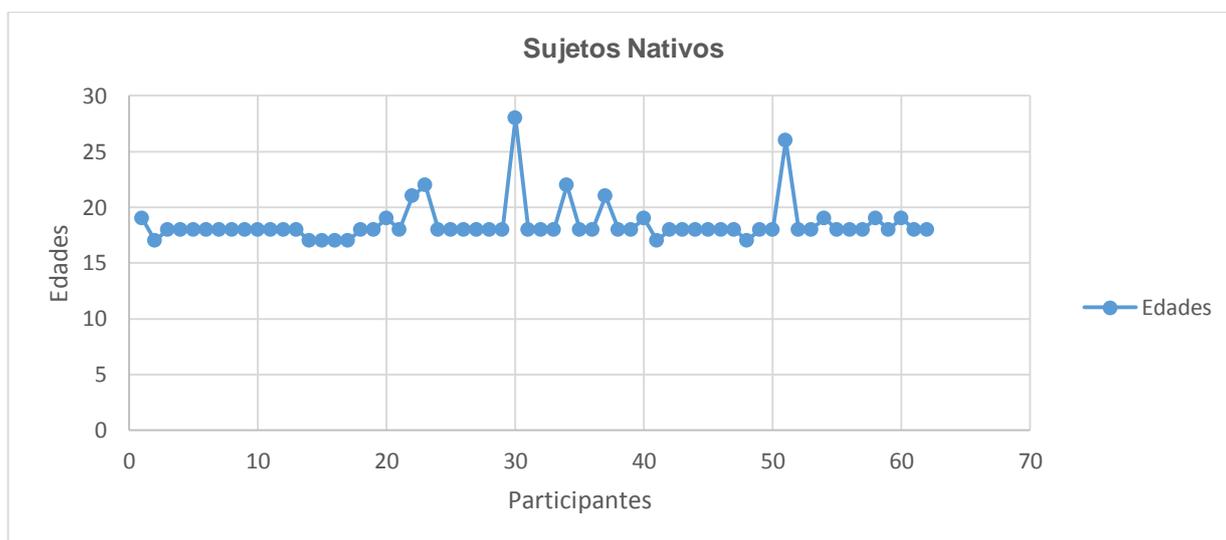


Gráfico 2. Edad de los sujetos Nativos de la muestra.

#### 4.1.1.2 Los sujetos No Nativos

Los sujetos NN que participaron en la investigación cursaban el nivel Intermedio - B1 de Español como Lengua Extranjera, según el Marco Común Europeo de Referencia (MCER), en una EOI de Barcelona.

El total de sujetos de la muestra fue de 133 y, al igual que en la muestra de los sujetos N, la mayoría era de género femenino como se puede ver en la siguiente tabla 4 y en el gráfico 3.

Sujetos No Nativos		N
Hombres		41
Mujeres		92
<b>Total</b>		<b>133</b>

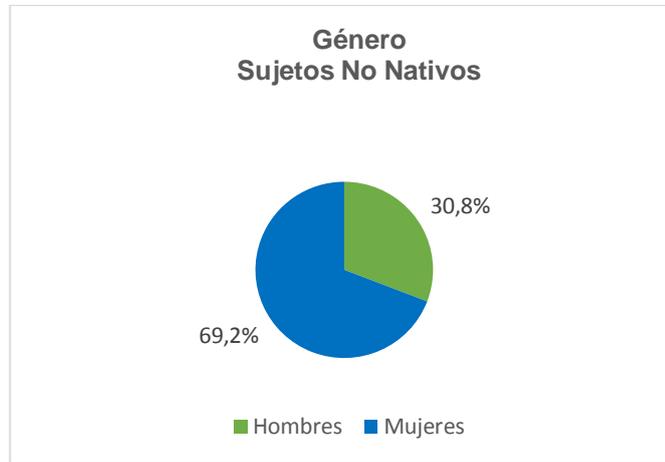


Gráfico 3. Género de los sujetos NN de la muestra en porcentaje.

En cuanto a las edades, la disparidad existente en la muestra de NN es mayor que en la muestra de sujetos Nativos. Se puede observar en el gráfico 4 que las edades oscilan principalmente entre 20 y 30 años. No obstante, podemos encontrar sujetos con edades que superan los 50 años y otros que están por debajo de los 20 años.

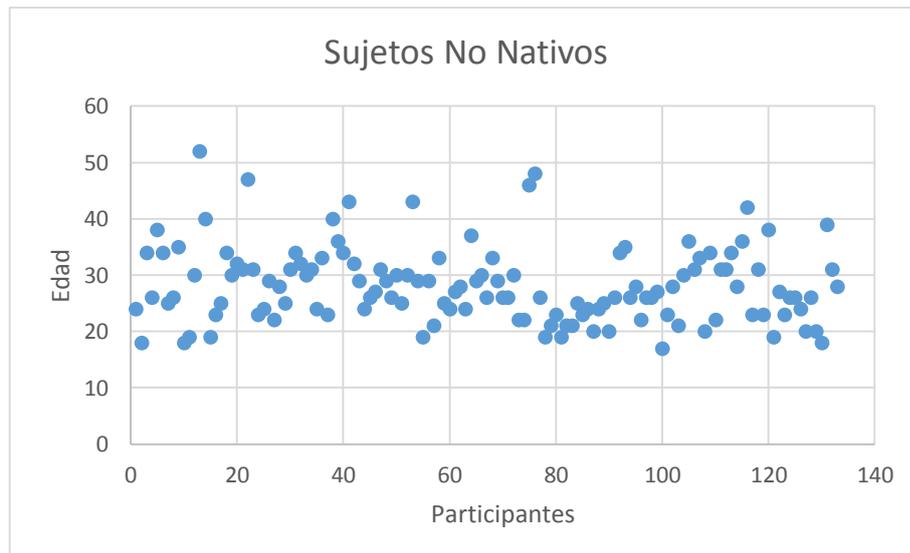


Gráfico 4. Edad de los sujetos NN de la muestra.

En cuanto a las nacionalidades de los sujetos participantes en la muestra, vemos que estamos ante un grupo multicultural y multilingüe formado por un total de 29 nacionalidades diferentes. En el gráfico 5 se puede observar la distribución del total de sujetos por nacionalidad.



Gráfico 5. Distribución de número de sujetos por nacionalidad.

Vemos en el gráfico que la nacionalidad con más sujetos participantes fue la china con un total de 32, seguida de la brasileña con 12. Pero como podemos observar, son más de 10 nacionalidades en las que solo hay uno o dos representantes.

Otra variable que tomamos en consideración fue la relativa a los estudios cursados por los sujetos NN. Se comprobó que la mayoría había cursado estudios superiores. Es decir, del total de 133 sujetos participantes en la investigación, 83 habían estudiado o estaban estudiando alguna carrera universitaria. El resto poseía estudios equivalentes a la educación secundaria.

Estudios	N
Educación secundaria	50
Educación superior	83
<b>Total</b>	<b>133</b>

Tabla 4. Estudios de los sujetos de la muestra.

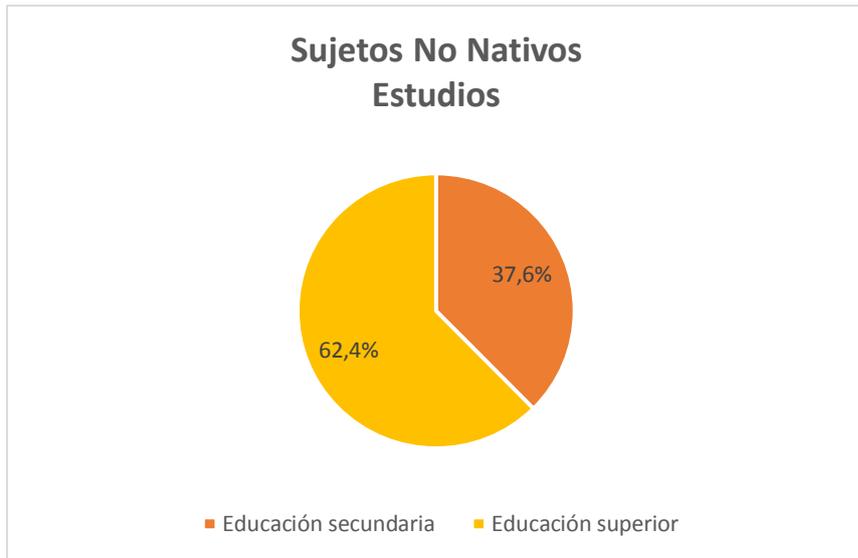


Gráfico 6. Nivel de estudios de los sujetos de la muestra en porcentaje.

Al observar qué tipo de estudios poseían según el género de los participantes, constatamos que era el género femenino el que en mayor número había realizado estudios universitarios.

	Educación secundaria	Educación superior	
hombres	18	23	
mujeres	32	60	

Tabla 5. Nivel de estudios de los sujetos de la muestra según el género.

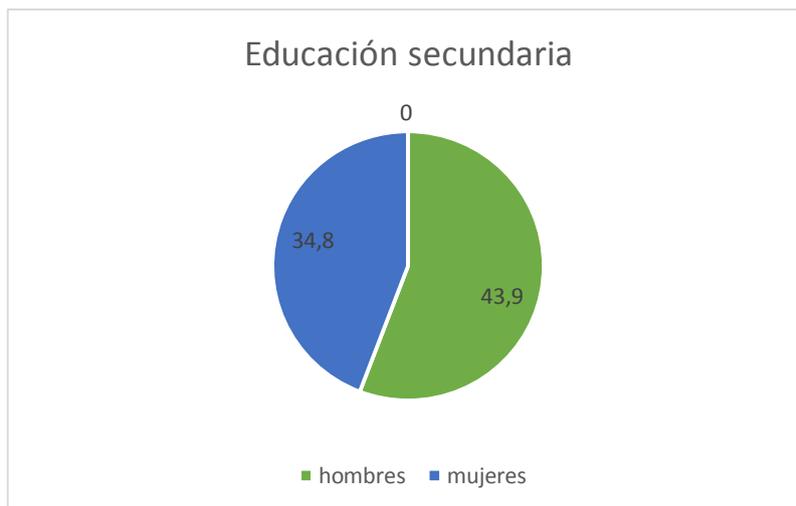


Gráfico 7. Nivel de estudios de los sujetos de la muestra según el género en porcentaje.

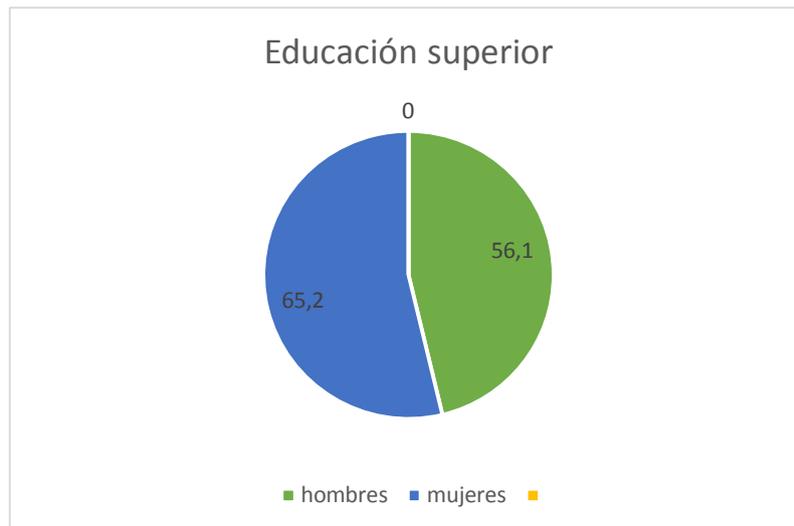


Gráfico 8. Nivel de estudios de los sujetos de la muestra según el género en porcentaje.

Otro dato a tener en cuenta de los sujetos NN es el itinerario de aprendizaje realizado en ELE. De los 133 sujetos participantes en la investigación, 76 habían cursado los niveles A-1 y A-2 en la EOI, mientras que el resto accedió al nivel Intermedio B-1 a través de una prueba de acceso administrada por la misma escuela. Dicha prueba consiste en responder un total de 100 ítems de opción múltiple. Los primeros quince ítems corresponden a una comprensión oral y los 85 restantes están relacionados con el uso de la lengua. Una vez realizada esta primera prueba, los aspirantes a acceder a la EOI realizan una expresión escrita en la que deben relatar sus últimas vacaciones. Para su realización disponen de quince minutos. Según los resultados obtenidos en la prueba de opción múltiple, se realiza una primera clasificación de los sujetos en los cinco niveles que ofrece la EOI. La prueba de expresión escrita se utiliza para matizar con mayor precisión el nivel que se adjudica y, si se cree necesario, se realiza una prueba de expresión oral consistente en una pequeña entrevista informal sobre su estancia en Barcelona y su motivación por realizar el curso.

Acceso a la EOI	N
Alumnos de la EOI	76
Alumnos nuevos	57

Tabla 6. Acceso al nivel Intermedio B-1.

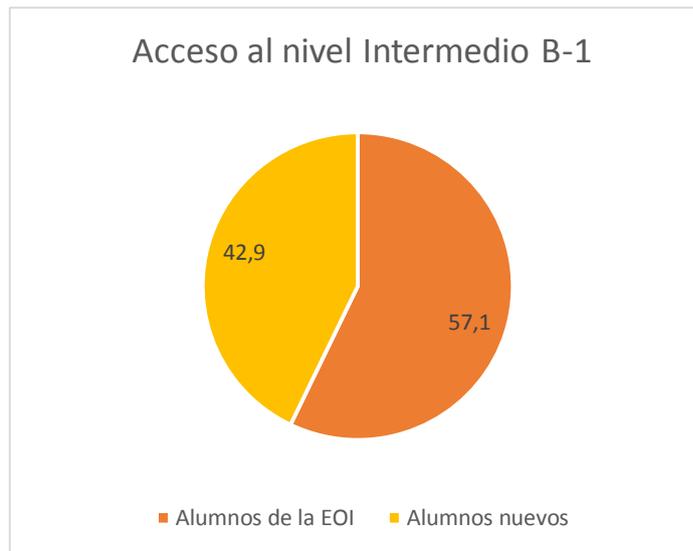


Gráfico 9. Vía de acceso al nivel Intermedio B-1 en porcentaje.

## 4.1.2 Las instituciones

### 4.1.2.1 La Universitat Pompeu Fabra

La Universidad Pompeu Fabra, como institución pública, se rige por la *Ley Orgánica 6/2001, de 21 de diciembre, de Universidades*. En dicha ley, en el capítulo I, artículo 7, se detalla la organización de las universidades: “Las Universidades públicas estarán integradas por Escuelas, Facultades, Departamentos, Institutos Universitarios de Investigación, Escuelas de Doctorado y por aquellos otros centros o estructuras necesarios para el desempeño de sus funciones.”

La Universidad Pompeu Fabra imparte enseñanza en tres ámbitos de conocimiento: i) ciencias sociales y humanas, ii) ciencias de la salud y de la vida y iii) ciencias de la comunicación y las tecnologías de la información. Para impartir dichas enseñanzas, la universidad se organiza en siete facultades, una Escuela Superior Politécnica, ocho departamentos y tres institutos universitarios.

En el artículo 8 de la misma *Ley Orgánica 6/2001*, se explican las funciones de las facultades y las escuelas universitarias: “Las escuelas y facultades son los centros encargados de la organización de las enseñanzas y de los procesos académicos, administrativos y de gestión conducentes a la obtención de títulos de grado.”

Entre las siete facultades de la Universidad Pompeu Fabra se encuentra la *Facultad de Traducción e Interpretación* en la que se llevó a cabo la recolección de datos para la realización de la presente investigación, en concreto en el *Grado en Lenguas Aplicadas*. La duración de este Grado es de cuatro cursos académicos con un total de 240 créditos ECTS.

#### 4.1.2.2 *La Escuela Oficial de Idiomas*

Las Escuelas Oficiales de Idiomas (EEOOI) conforman una vasta red de centros de educación pública de nivel no universitario dedicada a la enseñanza especializada de idiomas modernos. Las EEOOI son instituciones dependientes de las Consejerías de Educación de las distintas Comunidades Autónomas y están enmarcadas dentro de los centros de régimen especial regulados por la *LOE 2/2006 de 3 de mayo*, en cuyos artículos 59, 60, 61 y 62 se establece la normativa básica para las enseñanzas de idiomas.

Así, en su artículo 59.1, la *LOE 2/2006 de 3 de mayo* especifica que “las enseñanzas de idiomas tienen por objeto capacitar al alumnado para el uso adecuado de los diferentes idiomas (...) y se organizan en los niveles siguientes: básico, intermedio y avanzado.”

De igual manera, el acceso a las EEOOI queda regulado en el artículo 59.2 de la misma ley. Dicho artículo resalta que “será requisito imprescindible tener 16 años cumplidos en el año que se comiencen los estudios. Podrán acceder asimismo los mayores de 14 años para seguir las enseñanzas de un idioma distinto al cursado en la enseñanza secundaria obligatoria.”

El *Decret 4/2009 de 13 de enero* establece la ordenación y el currículum de las enseñanzas de idiomas de régimen especial para Cataluña. En el artículo 10 se especifica la organización y la finalidad de dichos estudios:

El nivel intermedio tiene como finalidad principal capacitar al alumnado para el uso del idioma con cierta seguridad y flexibilidad, receptiva y productivamente, tanto de forma hablada como escrita, así como para mediar entre hablantes de diferentes lenguas, en situaciones cotidianas y menos corrientes que requieran comprender y producir textos en una variedad de lengua estándar, con estructuras habituales y un repertorio léxico común no muy idiomático, y que traten sobre temas generales, cotidianos y en los que se tiene un interés personal.

Dentro de los objetivos generales por destrezas se detalla, para la expresión e interacción escrita: “Escribir textos cortos sobre hechos y acontecimientos que traten una

variedad de temas generales, adecuando el léxico y las estructuras al tema y a la situación comunicativa.”

#### *4.1.2.3 El contexto del Español como Lengua Extranjera (ELE) en las EEOOII*

El alumnado de las EEOOII suele ser muy variado, no sólo en cuanto a la edad, como acabamos de ver más arriba, también a los intereses y la motivación, el conocimiento de otras lenguas, el nivel socio-económico y socio-cultural, etc.

A este contexto general que afecta a todos los idiomas que se pueden impartir en las EEOOII, se debe añadir el hecho de que el alumnado de ELE conforma un variado abanico de nacionalidades y culturas que abarca los cinco continentes. Esta heterogeneidad de nacionalidades, lenguas y culturas, lejos de convertirse en un problema dentro del aula, es una ocasión única para fomentar la curiosidad y respeto hacia esas otras culturas con las que van a convivir durante un curso o más, apreciando, así, la diversidad étnica, religiosa, social y lingüística, adquiriendo, de esta manera, valores interculturales que incluyan una visión más amplia y un sentido crítico de su propia cultura y sociedad al compararla con las otras. Este intercambio de experiencias supone un enriquecimiento personal tanto para el alumnado como para el propio docente

## **4.2 Corpus y Recogida de datos**

Para la obtención de los datos, los participantes de la investigación, tanto el grupo de N como el de NN, realizaron dos pruebas de expresión escrita basadas en temas informales: (i) una de ellas consistía en explicar cómo había sido la primera clase de una segunda lengua (1D); (ii) en el otro texto tenían que continuar una historia en pasado a partir de un enunciado dado (F) (Anexo 1). Se eligieron estos dos tipos de textos para seguir el modelo de prueba de expresión escrita que se lleva a cabo en el examen para la obtención del Certificado de Nivel Intermedio (CNI) de las EEOOII, en el que se tiene que escribir un relato en pasado. El texto (ii) sigue exactamente dicho modelo de examen del CNI. En cambio, en el otro se propuso la variante de relato en pasado a partir de una experiencia personal. Se pensó que este tipo de texto resultaría más fácil al no requerir tanta creatividad como inventar una historia.

Para el procesamiento del corpus obtenido, ya que está compuesto por textos manuscritos, se siguieron los siguientes pasos: (i) el primer paso fue transcribirlos todos en formato .txt para, con posterioridad, poder analizarlos con el *software* elegido; (ii) se codificaron todos los textos para, por un lado, mantener el anonimato de los sujetos participantes y, por otro, para identificar a qué grupo, Nativos o No Nativos, y a qué texto y momento de la investigación pertenece cada producción.

Así, podemos identificar las producciones de los sujetos Nativos en dos categorías:

- 1RN\_1D\_05-14-aida
- 1RN\_F\_05-14-aida

La primera parte del código hace referencia a la recogida y al grupo (1RN). Es decir, primera recogida de Nativos. La segunda parte indica el texto que se escribió: (1D) primer día de clase; (F) continuar la frase. La tercera parte del código corresponde a la fecha de realización (05-14 es mayo de 2014) y a la identificación del sujeto con un nombre ficticio. En el corpus de los sujetos Nativos solo se realizó una recogida de datos.

En cuanto a las producciones de los sujetos No Nativos se codificaron siguiendo los mismos criterios que para las producciones de los Nativos. Dado que en el caso de los sujetos NN la recolección de datos se realizó en diferentes grupos y promociones, la codificación resultante fue la siguiente:

- 1RNN\_F\_02-08-ilona
- 1RNN\_F\_07-08-andrea
- 1RNN\_F\_10-08-fabio
- 1RNN\_F\_02-09-peng
- 1RNN\_1D\_02-10-delia
- 1RNN\_1D\_03-10-bryan
- 2RNN\_F\_05-08-anna
- 2RNN\_F\_06-08-eun
- 2RNN\_F\_12-08-jin
- 2RNN\_F\_12-09-henan
- 2RNN\_1D\_05-10a-debora
- 2RNN\_1D\_05-10b-tang

Como en el grupo de NN se realizó la recogida en dos momentos de su proceso de enseñanza-aprendizaje, esta diferencia queda reflejada en la primera parte del código: 1RNN o 2RNN; mientras que la segunda parte del código coincide con la codificación de los textos de los sujetos Nativos: (1D) primer día de clase; (F) continuar la frase. Al igual que para la muestra de N, la tercera parte del código remite a la fecha de recogida de los datos (05-08 es mayo de 2008) y a la identificación del sujeto con un nombre ficticio.

A continuación detallamos cómo se realizó la recogida de la muestra en cada uno de los grupos participantes y qué corpus se obtuvo.

### 4.2.1 *Corpus y Recogida de datos de Sujetos Nativos*

La recogida de la muestra se realizó en una de las aulas de la facultad de Traducción e Interpretación de la Universidad Pompeu Fabra de Barcelona. Los alumnos asistían a la asignatura *Lengua Española del Grado en Lenguas Aplicadas* y la realización de los textos se propuso como un ejercicio de clase para evaluar su corrección lingüística y estilística. Los alumnos dispusieron de 30 minutos para escribir cada texto y se escribieron los dos en la misma sesión.

Se les proporcionó una plantilla para cada uno de los textos en la que tenían que rellenar sus datos personales y en las que constaba también la instrucción para escribir los textos (Anexo 2).

Se tomó la decisión de que escribieran en primer lugar el texto en el que tenían que relatar una experiencia personal ya que se consideró que les resultaría más fácil al tratarse de explicar un *recuerdo*; mientras que el segundo texto requería de más creatividad al tener que escribir una historia a partir de un enunciado determinado. Sin embargo, los alumnos mostraron más dificultad para escribir sobre una experiencia personal que para crear una nueva historia. En sus comentarios decían no recordar su primera clase de una segunda lengua o que no sabían qué explicar.

En ese contexto, el corpus obtenido de los sujetos N fue de un total de 64 producciones escritas tanto del primer texto (1RN\_1D), como del segundo (1RN\_F). No obstante, una vez revisada la muestra obtenida, se descartaron aquellas producciones que superaban las 250 palabras o no alcanzaban las 50 por texto. Así, de las primeras producciones escritas que realizaron (1RN\_1D) se seleccionó un total de 62 y de las segundas (1RN\_F) 55. Por tanto, el total de producciones de la muestra de sujetos N es de 117 con un total de 19407 palabras (*tokens*).

	<b>Producciones</b>	<b>Tokens</b>
1RNN_1D	62	9746
1RNN_F	55	9661
<b>Total</b>	<b>117</b>	<b>19407</b>

Tabla 7. Total de producciones de los sujetos Nativos.

### 4.2.2 *Corpus y Recogida de datos de Sujetos No Nativos*

Los sujetos NN que participaron en la investigación pertenecían a distintas promociones y grupos dentro de la misma EOI. Todos ellos realizaron la prueba escrita en dos momentos de su proceso de enseñanza-aprendizaje: (i) al inicio del mismo, como evaluación de diagnóstico - Momento 1 (M1) - y (ii) al final, como evaluación de aprovechamiento - Momento 2 (M2). Entre ambos momentos transcurrió un período de instrucción de 120h.

Dentro de ese período de 120 horas, se prestó especial atención al relato en pasado. Durante tres semanas, en sesiones diarias de 2h (30 horas en total aproximadamente), se impartió una instrucción siguiendo el enfoque comunicativo a través de una enseñanza mediante tareas (EMT) apoyada en el manual *Rápido, Rápido* (Miquel y Sans, 2002). La enseñanza se centró en la creación de historias o cuentos en pasado a partir de una frase de inicio o a partir de determinados elementos como personajes concretos o situaciones. Así, se focalizó en la producción escrita de forma creativa, siguiendo una EMT, y en relación directa con la prueba que los sujetos tendrían que realizar al finalizar el curso para conseguir el Certificado de Nivel Intermedio.

De la misma manera que a los sujetos N, a los NN también se les proporcionaron las mismas plantillas para que rellenaran sus datos personales y con la instrucción para escribir ambos textos (Anexo 1-2). También dispusieron de 30 minutos, igual que los N, para escribir cada texto. Los NN realizaron las pruebas en momentos diferentes, pero siempre dentro del aula para controlar el tiempo.

En el M1, al inicio del curso, ambas pruebas se plantearon como ejercicios para evaluar sus conocimientos (evaluación de diagnóstico). En la segunda recogida, M2, los ejercicios se programaron como tareas finales y para evaluar el rendimiento del alumnado (evaluación de aprovechamiento).

El total de producciones recopiladas en el M1 fue de 108 y en el M2 fue de 113. Al igual que con el grupo de sujetos N, también se descartaron aquellas producciones escritas que superaban las 250 palabras o no alcanzaban las 50 por texto. Así, contamos con un total de 105 producciones del M1, durante la evaluación de diagnóstico, con un total de 12185 palabras (*tokens*), y de 113 del M2, durante la evaluación de aprovechamiento, con un total de 16715 palabras. Esto significa un total de 218 producciones escritas por sujetos NN con un total de 28900 palabras.

	<b>Producciones</b>	<b><i>Tokens</i></b>
1RNN_1D	34	3097
1RNN_F	71	9088
<b>Total</b>	<b>105</b>	<b>12185</b>

Tabla 8. Total de producciones de los sujetos NN en el M1.

	<b>Producciones</b>	<b><i>Tokens</i></b>
2RNN_1D	31	3571
2RNN_F	82	13144
<b>Total</b>	<b>113</b>	<b>16715</b>

Tabla 9. Total de producciones de los sujetos NN en el M2.

	<b>Total producciones NN</b>	<b>Total <i>Tokens</i></b>
Momento 1	105	12185
Momento 2	113	16715
<b>Total</b>	<b>218</b>	<b>28900</b>

Tabla 10. Total de producciones de los sujetos NN en el M1 y M2.



## CAPÍTULO 5

### METODOLOGÍA DE ANÁLISIS

#### 5.1 Instrumentos para el análisis lingüístico

Los instrumentos utilizados para el tratamiento lingüístico de los datos fueron: (i) el software informático *AntConc* y (ii) el software informático Paramtext.

##### 5.1.1 *AntConc*

El software informático *AntConc* es un software de acceso libre que está compuesto por diferentes herramientas (Concordance-Clusters-Collocates-Wordlist). Estas herramientas y sus instrumentos de análisis permiten, por ejemplo, elaborar listas de palabras monoléxicas o poliléxicas; realizar agrupamientos léxicos o de palabras clave, etc. (<http://www.laurenceanthony.net/software.html>)

A continuación detallamos las herramientas de este software informático utilizadas en la presente investigación.

- *Wordlist*

Una de las herramientas utilizadas ha sido *Wordlist*. Esta herramienta proporciona un listado de palabras de los textos seleccionados ordenado según la frecuencia de uso. Este puede ser de un solo texto o de todos los textos de la muestra. Asimismo, aporta información del número de *types* y *tokens* de los textos.

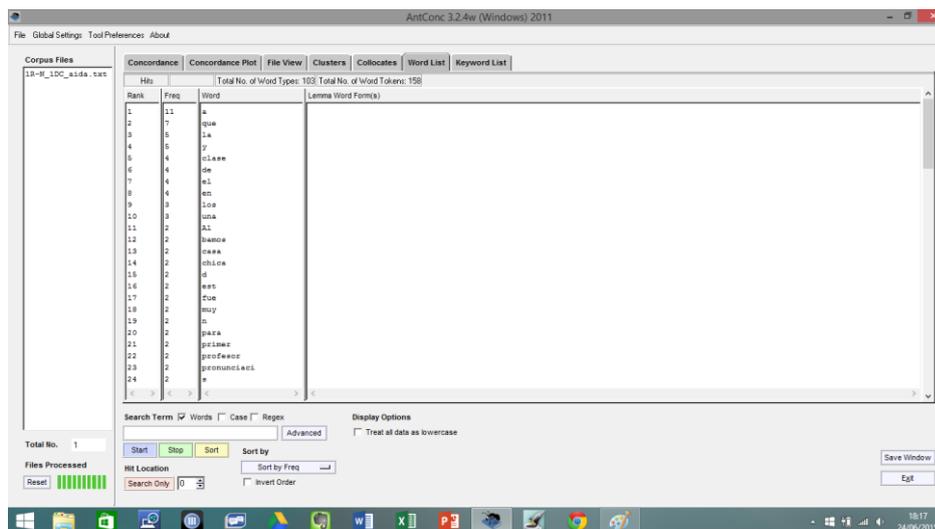


Figura 3. Ejemplo de *output* de la herramienta *Wordlist*.

- *Concordance*

Siguiendo la explicación que de esta herramienta se da en el tutorial, *Concordance* es la aplicación que se utiliza para extraer listados de una palabra concreta y las palabras que la rodean. Esto permite analizar la frecuencia de aparición de patrones lingüísticos dentro de la muestra (Figura 4).

- *Clusters*

Esta aplicación muestra grupos de palabras a partir de un término específico buscado y los ordena por orden alfabético o por frecuencia de aparición (Figura 5). Así, se puede buscar secuencias de palabras para encontrar estructuras fijas.

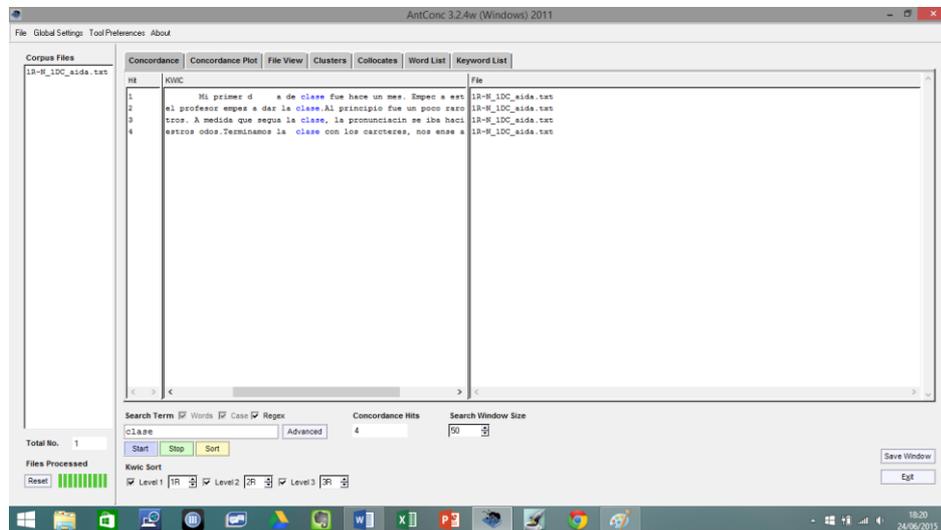


Figura 4. Ejemplo de *output* de la herramienta *Concordance*.

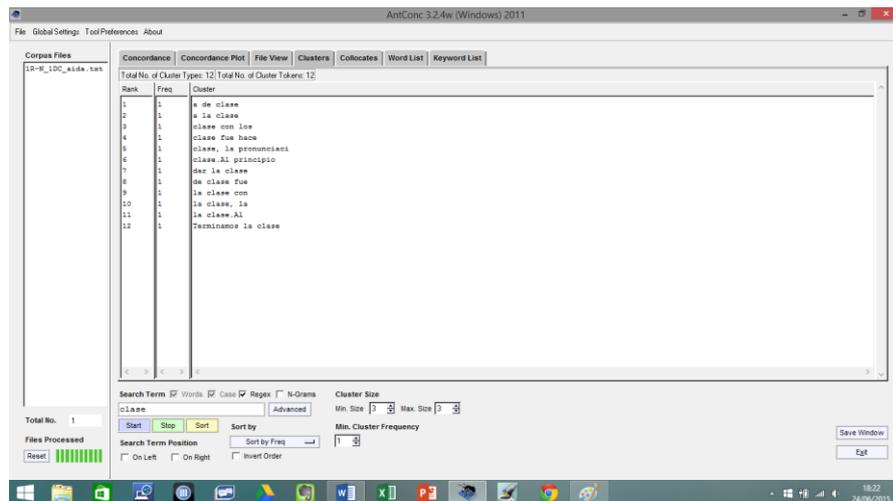


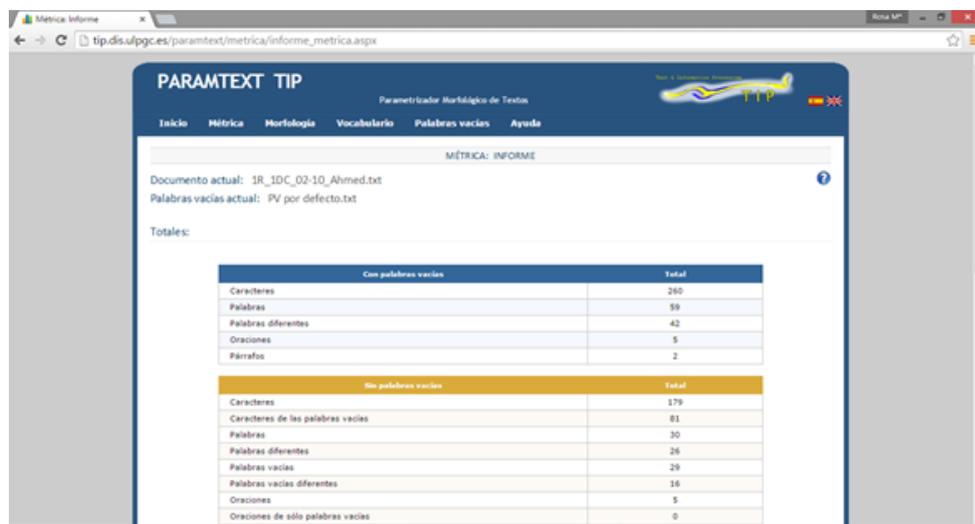
Figura 5. Ejemplo de *output* de la herramienta *Clusters*.

### 5.1.2 Paramtext

El software informático *Paramtext* (Carreras-Riudavets *et al.*, 2011) es un software de acceso libre que está compuesto por diferentes herramientas. En su página principal explica que se trata de una aplicación “destinada a la parametrización morfológica de textos”.

- Métrica

Esta herramienta muestra los resultados del total y los promedios de *types* y *tokens* tanto con palabras funcionales como sin ellas; ofrece también el número total y promedios de oraciones y párrafos. Toda esta información se presenta en tablas Excel.



Documentos actual: 1R\_IDC\_02-10\_Ahmed.txt  
Palabras vacías actual: PV por defecto.txt

Totales:

Con palabras vacías		Total
Caracteres		260
Palabras		59
Palabras diferentes		42
Oraciones		5
Párrafos		2

Sin palabras vacías		Total
Caracteres		179
Caracteres de las palabras vacías		81
Palabras		30
Palabras diferentes		26
Palabras vacías		29
Palabras vacías diferentes		16
Oraciones		5
Oraciones de sólo palabras vacías		0

Figura 6. Ejemplo de *output* de la herramienta Métrica.

- Morfología

En el apartado de morfología, Paramtext, entre otras informaciones, muestra los resultados de la frecuencia de aparición de las distintas categorías morfológicas. (Figura 7)

- Vocabulario

La aplicación de vocabulario proporciona un listado de las palabras del texto con o sin palabras funcionales. Este listado se puede ordenar por orden alfabético, por frecuencia, siguiendo el orden del texto, etc. (Figura 8)

PARAMTEXT TIP  
Parametrizador Morfológico de Textos

Inicio Métrica Morfología Vocabulario Palabras vacías Ayuda

MORFOLOGÍA: INFORME

Documento actual: IR\_1DC\_02-10\_Ahmed.txt  
Palabras vacías actual: PV por defecto.txt

Totales por categoría gramatical:

Utilice los bot de la primera columna para ver las palabras que pertenecen a esa categoría gramatical

Categoría gramatical	Frecuencia
Verbos	13
Sustantivos	26
Adjetivos	12
Adverbios	3
Pronombres	18
Preposiciones	9
Artículos	5
Conjunciones	9
No reconocidas	12
No reconocidas (sin repetidos)	12

Exportar

Figura 7. Ejemplo de *output* de la herramienta Morfología.

PARAMTEXT TIP  
Parametrizador Morfológico de Textos

Inicio Métrica Morfología Vocabulario Palabras vacías Ayuda

VOCABULARIO

Documento actual: IR\_1DC\_02-10\_Ahmed.txt  
Palabras vacías actual: PV por defecto.txt

Con palabras vacías Sin palabras vacías

Situe el cursor del ratón sobre una palabra para ver su información morfológica  
Los corchetes indican la sílaba tónica

Palabra	Sílabas	Frec. (tokens)	Prim. Apellido	Centro de gravedad	Desv. Típ. (C.G.)	Frec. (types)	Palabra invertida
hi	hi	1	1	1	0	829700	hi
primer	pr[i]mer	1	2	2	0	101321	remurp
a	a	1	3	3	0	0	a
de	de	3	4	21	15,14	1891392	ed
clase	[cl]ase	3	5	23	15,72	53166	esalc
fué	fu[e]é	1	8	8	0	0	Éuf
Muy	Muy	1	7	7	0	517061	uM
interesante	an[ter]e[n]te	1	8	8	0	1686	etnereatni
poncas	[p]o[n]cas	2	9	29	20,99	492152	suprop
le	le	1	10	10	0	11063264	el
profesore	pro[fe]so[re]	1	11	11	0	2683	erosefrop
explaste	ex[pl]aste	1	12	12	0	0	atellase

Figura 8. Ejemplo de *output* de la herramienta Vocabulario.

## 5.2 Instrumentos para el análisis estadístico

Como en nuestra investigación comparamos dos poblaciones, sujetos N y sujetos NN, antes de decidir qué tipo de software estadístico era el más apropiado para realizar el análisis de la muestra, tuvimos que seguir una serie de pasos: (i) Determinar qué tipo de variables eran las que formaban nuestro corpus: variables continuas o variables dicotómicas; (ii) Comprobar si la muestra era independiente o relacionada, y (iii) Si la variable era continua, había que comprobar si las muestras eran normales o no.

En nuestro estudio trabajamos con variables continuas. Es decir, variables cuantitativas que toman valores a partir de números, como por ejemplo número de *types* y *tokens*, número de sustantivos, adjetivos o verbos.

En cuanto al tipo de muestra de nuestra investigación, se trata de una muestra independiente. Una muestra es independiente cuando los elementos que se analizan no tienen relación entre sí, es decir, que no son los mismos en una y otra muestra. En el presente estudio las poblaciones que se comparan son, por un lado, sujetos N y sujetos NN, por tanto, son sujetos no relacionados. Y, por otro lado, comparamos también los resultados del grupo de sujetos NN en dos momentos de su proceso de enseñanza-aprendizaje. En este caso hay que tener en cuenta que los sujetos NN que colaboraron en esos dos momentos de recogida de datos pertenecían a grupos y promociones diferentes. No obstante, aunque hubo sujetos que participaron en los dos momentos de recogida de datos, al ser un número reducido, para el análisis general se consideraron como parte de una muestra independiente.

Al tratarse de subgrupos de promociones diferentes, había que comprobar la bondad del ajuste a la normalidad para verificar que se podían considerar dichos subgrupos como uno solo.

A continuación detallamos los instrumentos de análisis que hemos utilizado.

### **5.2.1 *G-Stat 2.0: software estadístico de acceso libre***

El G-Stat 2.0 es un software para análisis estadístico de acceso libre que ha sido revisado por el *Servei d'Estadística Aplicada* de la Universidad Autònoma de Barcelona. Este software dispone de todas las herramientas necesarias para realizar no sólo estadística descriptiva o deductiva, sino también inferencial o inductiva.

La estadística descriptiva permite obtener una visión general de la muestra. Es decir, ofrece un recuento de los datos, los organiza y clasifica a partir de la observación de los mismos. Mientras que la estadística inferencial o inductiva da un paso más y, partiendo de los datos de la descriptiva, realiza comparaciones y/o generalizaciones, por ejemplo, de una muestra o una población. Así, en nuestra investigación, realizamos una primera aproximación al análisis de los datos a través de una estadística descriptiva para profundizar a continuación con una estadística inferencial (Lazaraton, 2005; Ortega y Iberri-Shea, 2005; Martínez, 2005; Laufer y Girsai, 2008; Gallego y Llach, 2009).

En el ámbito de la Adquisición de Segundas Lenguas (ASL), las herramientas más utilizadas en la estadística descriptiva son: (i) medidas de frecuencia, (ii) medidas de tendencia y (iii) medidas de variabilidad o dispersión (Mackey y Gass, 2005).

Como conseguir datos de toda una población es una tarea casi imposible, según indican Mackey y Gass (2005), se suele recolectar una muestra representativa de la misma y, mediante la estadística inferencial o inductiva, se generalizan los resultados a la totalidad de la población investigada. Así, una vez realizado el primer acercamiento a nuestros datos a través de la estadística descriptiva, pasamos a la estadística inferencial para poder comparar los resultados entre los grupos de N y NN, por un lado, y entre el M1 y M2 de los NN, por otro lado, y comprobar, así, si las diferencias observadas eran significativas estadísticamente o no.

Para elegir la herramienta estadística inferencial más adecuada y así evitar caer en errores en los resultados del análisis, primero se tiene que comprobar que la muestra cumple con toda una serie de requisitos (Mackey y Gass, 2005).

El primer supuesto que se debe comprobar es el ajuste a la normalidad<sup>25</sup> de la muestra. Pero según Pedrosa *et al.* (2015) muchos estudios obvian este paso previo al tratamiento de los datos y directamente recurren a pruebas estadísticas paramétricas tales como Anova, correlación de Pearson, prueba t, etc. Para estos autores, esto suele ser así porque las pruebas estadísticas paramétricas han demostrado a lo largo de los años su robustez en el tratamiento de datos o, en ocasiones, simplemente por desidia de los investigadores. Sin embargo, afirman, asimismo, que cuando no se cumplen los supuestos de normalidad ni el de homocedasticidad los expertos en metodología recomiendan el uso de pruebas no paramétricas. En su estudio, Pedrosa *et al.* (2015) analizan las principales pruebas estadísticas para la comprobación del ajuste a la normalidad. Las más conocidas son la de *Shapiro-Wilk* y la de *Kolmogorov-Smirnov*. Tradicionalmente se había considerado que la primera era adecuada para muestras pequeñas (<30), mientras que la segunda era más adecuada para muestras >30. Otros autores, sin embargo, hablan de un tamaño de muestra muy distinto y recomiendan el uso de la prueba de *Shapiro-Wilk* en muestras menores de 2000 y la de *Kolmogorov-Smirnov* para muestras mayores de 2000. (Stevens, 2002)

Los resultados obtenidos por Pedrosa *et al.* (2015) confirman la tendencia extremadamente conservadora de la prueba de *Kolmogorov-Smirnov* y señalan que en la

---

<sup>25</sup> En estadística, la distribución normal representa agrupamientos de comportamientos o puntuaciones en una gráfica que tiene una curva en forma de campana y un solo pico en el centro de la distribución. Esta curva se conoce como campana de Gauss y en el pico de la misma se encuentran las medidas de tendencia central (media, moda y mediana). (Mackey y Gass, 2005; Juan *et. al*, s/f)

mayoría de los casos esta prueba acaba aceptando la hipótesis nula, es decir, el ajuste a la normalidad. En cuanto a la prueba de *Shapiro-Wilk*, los resultados demuestran que su poder de detección es mejor cuando la muestra es igual o superior a 50 y no con muestras menores como habían indicado otros estudios. También concluyen que la prueba de *Shapiro-Wilk* es la más consistente a los cambios de tamaño de la muestra “al contar con un poder de detección razonable y muy cercano al 5% esperado cuando se modifica el tamaño muestral” (Pedrosa *et al.*, 2015: 21). Es por ello que para nuestra investigación decidimos utilizar esta prueba estadística para comprobar la bondad de ajuste a la normalidad.

Otra prueba que se ha de realizar antes de tomar decisiones sobre qué prueba estadística utilizar en la comparación de dos muestras es el cálculo de igualdad de varianzas cuando las muestras presentan ajuste a la normalidad. Para el cálculo se utiliza la prueba F-Snedecor.

#### 5.2.1.1 Fases del análisis estadístico

El análisis estadístico lo vamos a llevar a cabo en tres fases: (i) en la fase uno, vamos a comprobar la bondad de ajuste a la normalidad y el cálculo de igualdad de varianzas; (ii) en la siguiente fase, vamos a realizar una estadística descriptiva, y (iii) en la tercera fase, vamos a llevar a cabo una estadística inferencial de relación y comparación de grupos.

A continuación pasamos a explicar con algo más de detalle en qué consiste cada una de estas fases.

**Fase 1:** Comprobación de la bondad de ajuste a la normalidad y cálculo de la igualdad de varianzas.

El objetivo de esta fase era comprobar la bondad de ajuste a la normalidad para saber si para el tratamiento de los datos había que utilizar una prueba estadística paramétrica o no paramétrica.

Los pasos que seguimos para realizar esta comprobación fueron los siguientes:

(i) Verificar que las distintas muestras de los grupos de NN que participaron en la investigación podían considerarse como un todo y no como diferentes grupos. Para ello se aplicó la prueba estadística de ajuste a la normalidad de *Shapiro-Wilk* a todos los grupos de

NN. Tras comprobar que no se podía rechazar la hipótesis nula ( $H_0$ ) ya que el resultado obtenido, la p-valor, era superior al valor de significancia de 0.05, se verificó que el corpus de los sujetos NN se podía considerar como una sola muestra tanto para el M1 de recogida de datos, al inicio del curso, como en el M2, al finalizar el curso (ver tabla 13). Asimismo, se comprobó también la bondad de ajuste a la normalidad de los dos momentos de recogida de datos del grupo de no nativos para poder tratar estos como una sola muestra y poder comparar, como veremos más adelante, por grupos, nativos frente a no nativos, sin importar las horas de instrucción (ver tabla 14).

(ii) Una vez asumido que los diferentes grupos de NN podían formar parte de una sola muestra, clasificamos las muestras de N y NN en función del tipo de texto para facilitar su posterior comparación:

Clasificación de las muestras por tipo de texto	
NN-M1- M2-F	NN-M1- M2-1D
N-F	N-1D

Tabla 11. Clasificación de las muestras por tipo de texto

Clasificación de las muestras por estadios	
NN-M1-F	NN-M1-1D
NN-M2-F	NN-M2-1D
N-F	N-1D

Tabla 12. Clasificación de las muestras por estadios (nivel de lengua)

(iii) En el tercer paso era necesario comprobar la bondad de ajuste a la normalidad de cada una de las variables<sup>26</sup> seleccionadas para cada muestra. Para ello volvimos a aplicar la prueba estadística de ajuste a la normalidad de *Shapiro-Wilk* y calculamos también la igualdad de varianzas<sup>27</sup>.

(iv) El último paso era decidir qué tipo de prueba estadística podríamos utilizar para el tratamiento de los datos. Ya que los resultados obtenidos con la prueba de *Shapiro-Wilk*

<sup>26</sup> Las variables seleccionadas son: Tokens; Types; oraciones; verbos; sustantivos; adjetivos; adverbios; pronombres; preposiciones; artículos; conjunciones.

<sup>27</sup> Como vamos a comparar, por un lado (a) dos muestras (nativos y no nativos) y, por otro, (b) tres muestras (nivel inicial B1, nivel final B1 y nivel nativo), las pruebas estadísticas que vamos a utilizar son diferentes en un caso y en otro. Así, en el primer caso (comparación de dos muestras), si no se puede rechazar la  $H_0$  de bondad de ajuste a la normalidad, se debe realizar una segunda prueba para comprobar la igualdad de varianzas para poder decidir con más exactitud qué prueba estadística se debe utilizar en estos casos (t-Student de varianzas iguales o t-Student de varianzas diferentes).

mostraron casos en los que debíamos rechazar la  $H_0$ , es decir, que no había ajuste a la normalidad; en otros casos no era posible rechazar la  $H_0$ , es decir, que sí había ajuste a la normalidad. En estas situaciones, si uno de los componentes que se desean comparar no sigue la bondad de ajuste a la normalidad se debe utilizar una prueba no paramétrica. Mientras que si todos los componentes siguen la bondad de ajuste a la normalidad, se utilizará una prueba paramétrica.

Como en nuestras muestras se dan las dos situaciones, las pruebas estadísticas que vamos a utilizar en el tratamiento de los datos son (a) para el tratamiento de aquellas variables que siguen la bondad de ajuste a la normalidad, por un lado la prueba estadística t-Student de varianzas iguales o diferentes, para la comparación de dos muestras, y la prueba estadística ANOVA, para la comparación de tres muestras; (b) para aquellas variables que no siguen la bondad de ajuste a la normalidad, por un lado la prueba estadística de *Mann-Whitney*, para la comparación de dos muestras, y la prueba estadística *Kruskal-Wallis*, para la comparación de tres muestras.

<b>1RNN_F_02-08</b>	<b>2RNN_F_05-08</b>
<b>Para W = 0.9552 y n = 22</b>	Para W = 0.9552 y n = 20
<b>p-valor = 0.3987 &gt; 0.05</b>	<b>p-valor = 0.4529 &gt; 0.05</b>
<b>1RNN_F_07-08</b>	<b>2RNN_F_06-08</b>
<b>Para W = 0.9552 y n = 13</b>	Para W = 0.9552 y n = 23
<b>p-valor = 0.6787 &gt; 0.05</b>	<b>p-valor = 0.3736 &gt; 0.05</b>
<b>1RNN_F_10-08</b>	<b>2RNN_F_12-08</b>
<b>Para W = 0.9552 y n = 24</b>	Para W = 0.9552 y n = 19
<b>p-valor = 0.3497 &gt; 0.05</b>	<b>p-valor = 0.4820 &gt; 0.05</b>
<b>1RNN_F_02-09</b>	<b>2RNN_F_12-09</b>
<b>Para W = 0.9552 y n = 12</b>	Para W = 0.9552 y n = 20
<b>p-valor = 0.7138 &gt; 0.05</b>	<b>p-valor = 0.4529 &gt; 0.05</b>
<b>1RNN_1D_02-10</b>	<b>2RNN_1D_05-10a</b>
<b>Para W = 0.9552 y n = 22</b>	Para W = 0.9552 y n = 22
<b>p-valor = 0.3987 &gt; 0.05</b>	<b>p-valor = 0.3987 &gt; 0.05</b>
<b>1RNN_1D_03-10</b>	<b>2RNN_1D_05-10b</b>
<b>Para W = 0.9552 y n = 12</b>	Para W = 0.9552 y n = 9
<b>p-valor = 0.7138 &gt; 0.05</b>	<b>p-valor = 0.7470 &gt; 0.05</b>

Tabla 13. Resultados de la prueba de *Shapiro-Wilk* para el grupo NN en M1 y M2

<b>NN_F_M1</b>	<b>NN_1D_M1</b>
<b>Para W = 0.9902 y n = 71</b>	Para W = 0.9453 y n = 34
<b>p-valor = 0.8570 &gt; 0.05</b>	<b>p-valor = 0.0886 &gt; 0.05</b>
<b>NN_F_M2</b>	<b>NN_1D_M2</b>
<b>Para W = 0.9858 y n = 82</b>	Para W = 0.9661 y n = 31
<b>p-valor = 0.5069 &gt; 0.05</b>	<b>p-valor = 0.4197 &gt; 0.05</b>

Tabla 14. Resultados de la prueba de *Shapiro-Wilk* para NN

A continuación mostramos ejemplos del *output* que proporciona el software G-Stat 2.0 para el cálculo de la bondad de ajuste a la normalidad y la igualdad de varianzas.

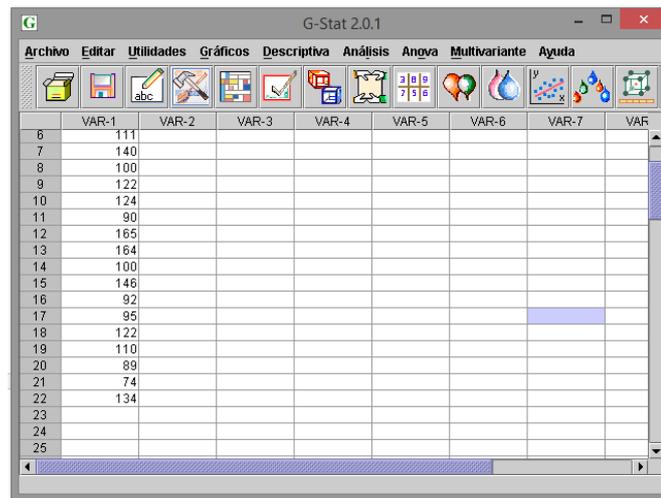


Figura 9. Ejemplo del software G-Stat 2.0.

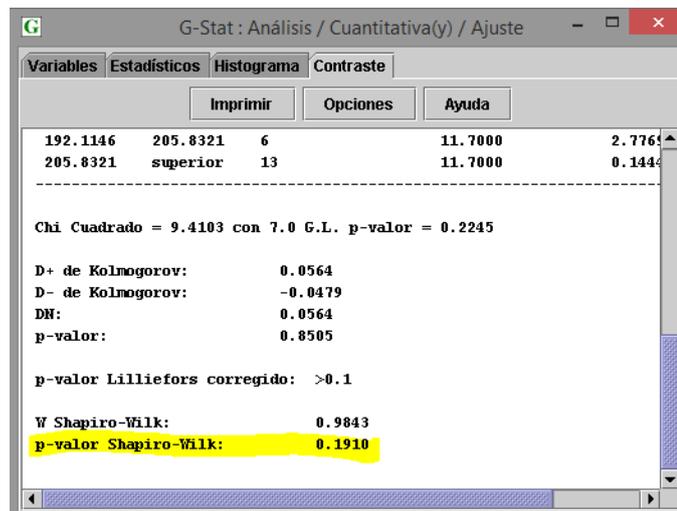
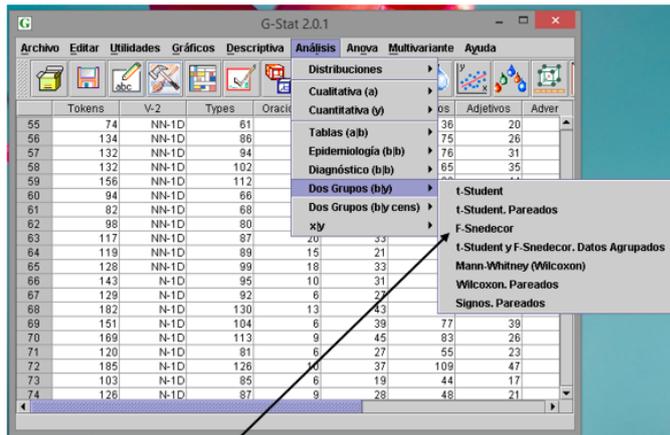
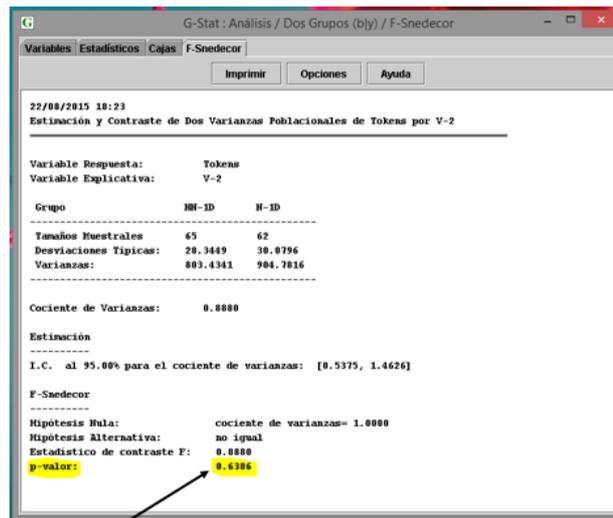


Figura 10. Ejemplo de *output* de cálculo de ajuste a la normalidad



La prueba F-Snedecor calcula la igualdad de varianzas para decidir qué prueba estadística utilizar entre t-Student de varianzas iguales o t-Student de varianzas diferentes.

Figura 11. Ejemplo de *output* de cálculo igualdad de varianzas



El valor de la p-valor indica si las varianzas son iguales o no. En este caso, como el valor es superior al valor de significancia de 0.05, no se puede rechazar la  $H_0$  de igualdad.

Figura 12. Ejemplo de *output* de cálculo igualdad de varianzas

## Fase 2: Estadística descriptiva

El profesor de estadística Jaume Llopis en su blog<sup>28</sup> *La estadística: una orquesta hecha instrumento* señala que hay tres técnicas básicas en estadística para analizar los datos de una muestra y poder ver en ellos lo que la simple mirada del ser humano no puede percibir. Estas técnicas son: (i) estadística descriptiva; (ii) estadística de relación, y (iii) estadística de

<sup>28</sup> <https://estadisticaorquestainstrumento.wordpress.com>

comparación. La primera, como ya hemos indicado más arriba, es una técnica cuyo objetivo es ordenar y clasificar los datos, hacer la descripción de los mismos, sin intención de inferir; mientras que la estadística de relación y comparación son técnicas inferenciales, es decir, son técnicas con las que se intenta generalizar a una población conclusiones extraídas a partir de los datos de una muestra representativa de dicha población.

En la estadística descriptiva se trabaja con «estadísticos», es decir, descriptores que presentan distintas perspectivas de los datos de la muestra. Como explicamos en el apartado 3.1, la información se puede organizar delimitando la posición de los datos en relación con las medidas de tendencia central (media, mediana, etc.), con los parámetros de dispersión (desviación estándar, varianza, etc.) o de forma (coeficiente de asimetría, coeficiente de curtosis, etc.).

En la figura 13 vemos los distintos estadísticos que ofrece el software G-Stat 2.0.

The screenshot shows the G-Stat software interface with the 'Estadísticos' tab selected. The window title is 'G-Stat : Descriptiva / Cuantitativa (y)'. Below the menu bar, there are buttons for 'Imprimir', 'Opciones', and 'Ayuda'. The main content area displays a table of statistics for the variable 'Tokens'.

Estadístico	Tokens
N	117
Media	165.8718
Mediana	166.0000
Moda	167.0000
Media Geométrica	162.9628
Varianza	912.2679
Desviación Típica	31.1812
E.E. de la Media (*)	2.8827
Mínimo	103.0000
Máximo	250.0000
Rango	147.0000
Cuartil Inferior	142.0000
Cuartil Superior	186.0000
Rango Intercuartílico	44.0000
Asimetría	0.2905
Asimetría Estandarizada	1.2830
Curtosis	-0.2065
Curtosis Estandarizada	-0.4559
Coficiente de Variación	18.7984

(\*) Usar con propósito de estimación para el I.C. de la media

Figura 13. Ejemplo de *output* de estadísticos

No obstante, no es imprescindible recurrir a todos los estadísticos para hacer una descripción de la muestra. Cuando la variable no se distribuye siguiendo la bondad al ajuste de la normalidad, Jaume Llopis aconseja aportar el cálculo de la mediana y el rango intercuartílico; mientras que si hay ajuste a la normalidad, aconseja la media y la desviación estándar. Como hemos visto, en nuestro caso se dan las dos situaciones, así, del total de 18 estadísticos posibles, nosotros hemos decidido utilizar diez, pues aportan información

suficiente sobre las medidas de tendencia central y los parámetros de dispersión. Los descriptores utilizados son:

(i) La media: hemos decidido incluirlo por tratarse del estadístico más conocido, aunque como acabamos de decir, el profesor Llopis aconseja utilizar mejor la mediana cuando se trata de muestras que no siguen el ajuste a la normalidad. La media se calcula sumando todos los valores de la muestra y dividiendo por el tamaño total de la muestra.

(ii) La mediana: para calcular la mediana se deben ordenar todos los valores de la muestra de menor a mayor y se separa la lista en dos mitades de tamaño igual. El valor central de esas dos mitades corresponde a la mediana. La mediana presenta la ventaja de ser resistente a los valores extremos.

(iii) La varianza: es un valor de dispersión. Como se ha dicho anteriormente, se aconseja este valor cuando se trata de muestras que no siguen la bondad de ajuste a la normalidad.

(iv) La desviación estándar o típica: es también un valor de dispersión. Según Jaume Llopis, es uno de los descriptores de dispersión más utilizados.

(v) Mínimo: muestra el valor mínimo de la muestra.

(vi) Máximo: muestra el valor máximo de la muestra.

(vii) Cuartil inferior: corresponde con el cuartil 1°.

(viii) Cuartil superior: corresponde con el cuartil 3°. Los cuartiles “son medidas de centralización que dividen los valores ordenados de la variable en cuatro grupos” (Manual G-Stat 2.0).

(ix) Rango intercuartílico: corresponde a la diferencia entre el Tercer cuartil y el Primer cuartil.

(x) Curtosis: es una medida que informa de la forma de la distribución respecto a la distribución normal o campana de Gauss. Es decir, indica si la curva es más aguda o más plana. Cuando la distribución es «normal» presenta «curtosis» cero (se ajusta a la campana de Gauss). Si el coeficiente de curtosis es negativo indica que la distribución tiende a tener menor número de valores extremos que la distribución normal (colas más bajas), es decir, la curva es más plana; mientras que si el coeficiente es positivo tiende a tener mayor número de valores extremos (colas más altas), por tanto, la curva es más aguda. (Manual G-Stat 2.0).

También hemos utilizado el «diagrama de caja» o «Box-Plot» para dar una representación gráfica de la información de los distintos estadísticos. A continuación, mostramos un ejemplo de «diagrama de caja» con una breve explicación de qué representa.

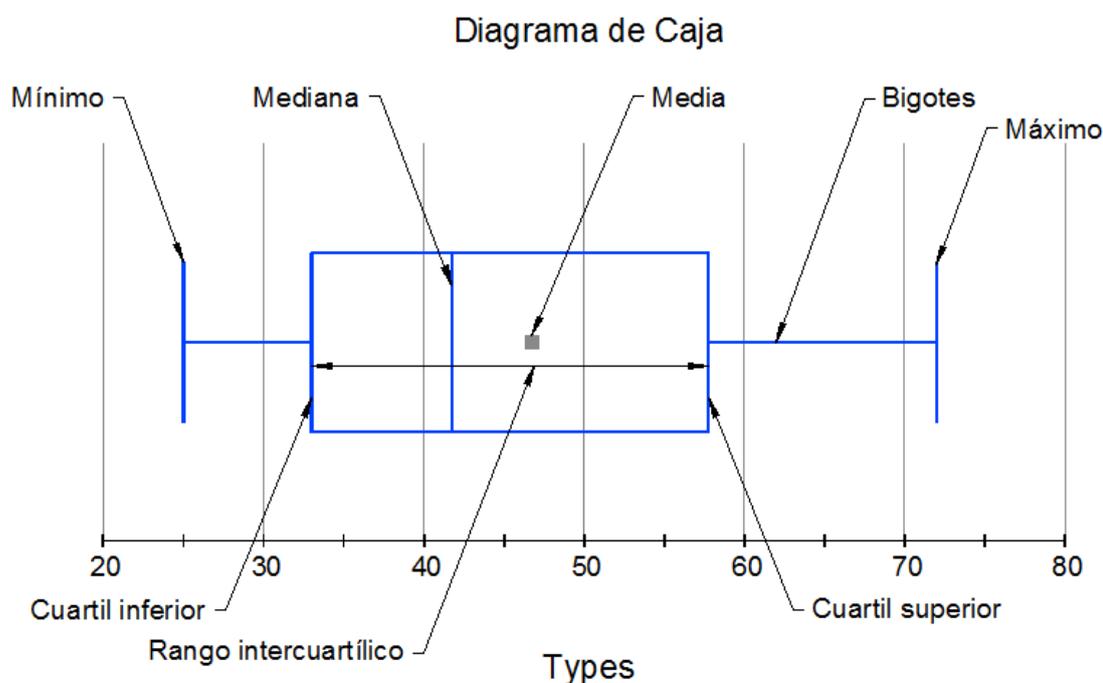


Figura 14: Ejemplo de «diagrama de caja»

### Fase 3: Estadística inferencial

En la estadística inferencial la muestra es el medio para extraer conclusiones y extrapolarlas a una población. Por tanto, va un paso más allá que la estadística descriptiva que se limita, como hemos señalado, simplemente a mostrar lo que se ve.

En esta fase, como ya hemos indicado más arriba, vamos a comparar los datos de las muestras de las que disponemos, (i) comparando por grupo (N frente a NN), (ii) comparando por estadios (NN-nivel inicial B1; NN-nivel final B1; nivel nativo) y (iii) comparando el grupo de NN en el M1 (inicio curso – nivel inicial B1) y M2 (final curso – nivel final B1).

Clasificación de las muestras grupo NN M1-M2	
NN-M1	
NN-M2	

Clasificación de las muestras por grupos	
NN- F	NN- 1D
N-F	N-1D

Clasificación de las muestras por estadios	
NN-M1-F	NN-M1-1D
NN-M2-F	NN-M2-1D
N-F	N-1D

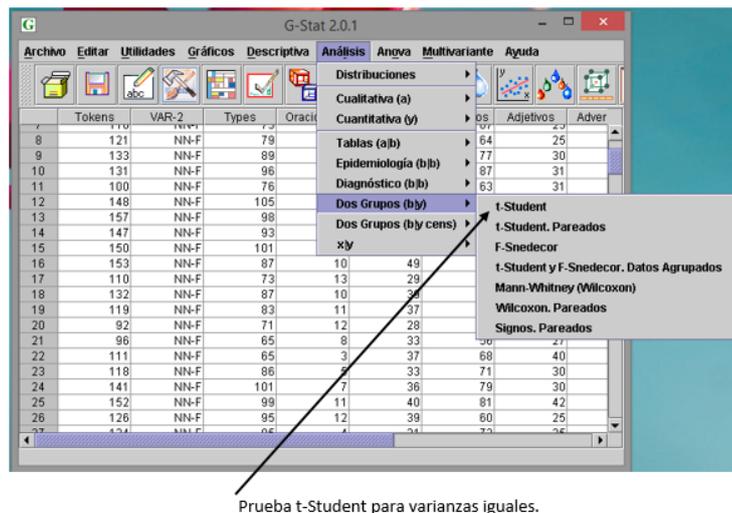
Tabla 15. Clasificación de las muestras

Vamos a comparar entre sí las muestras «F» y las «1D», pero también vamos a buscar si hay diferencias entre un texto y otro, así como ver si hay evolución en el grupo de NN entre el M1 y el M2 de su proceso de enseñanza-aprendizaje y entre estos y el grupo de nativos.

Para ello, vamos a utilizar, como ya hemos indicado, (a) para el tratamiento de aquellas variables que siguen la bondad de ajuste a la normalidad, por un lado la prueba estadística t-Student de varianzas iguales o diferentes, según los resultados, para la comparación de dos muestras, y la prueba estadística ANOVA, para la comparación de tres muestras; (b) para aquellas variables que no siguen la bondad de ajuste a la normalidad, por un lado la prueba estadística de *Mann-Whitney*, para la comparación de dos muestras, y la prueba estadística *Kruskal-Wallis*, para la comparación de tres muestras.

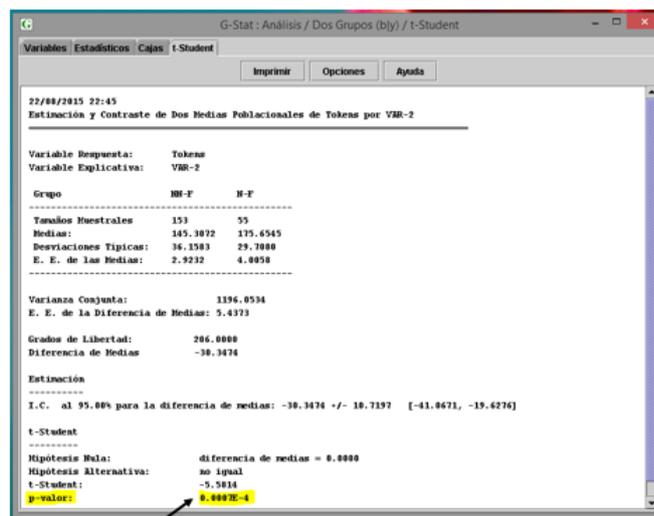
(i) La t-Student para varianzas iguales o diferentes es una prueba estadística paramétrica que se utiliza cuando las muestras siguen la bondad de ajuste a la normalidad. Contrasta que la diferencia de medias sea igual a un valor dado (normalmente cero).

A continuación mostramos dos ejemplos del *output* que se obtiene.



Prueba t-Student para varianzas iguales.

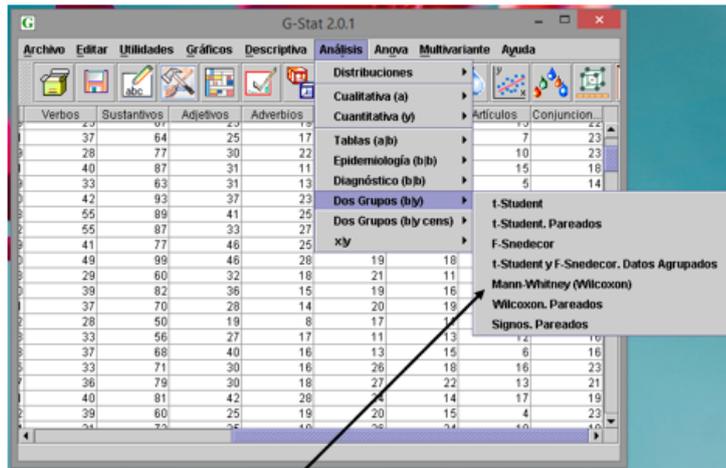
Figura 15. Ejemplo de la prueba t-Student



El p-valor indica que hay diferencia estadísticamente significativa pues el valor es inferior a 0.05

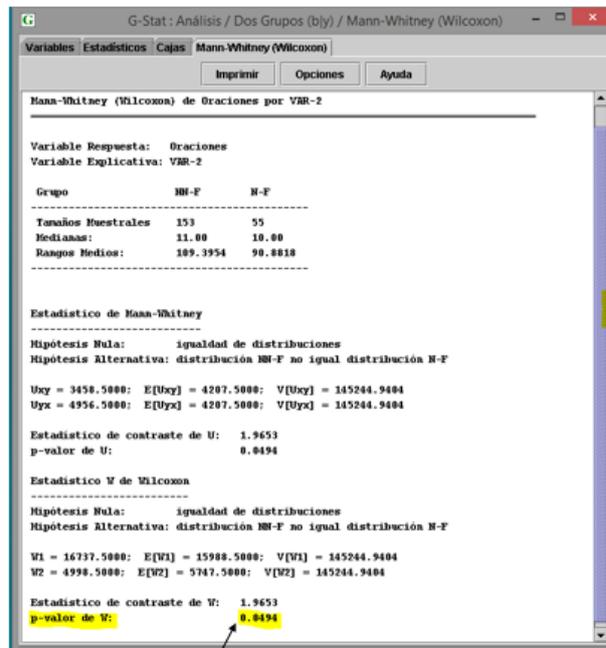
Figura 16. Ejemplo de *output* de la prueba t-Student

- (ii) El test de Mann-Whitney es una prueba no paramétrica que se utiliza para la comparación de dos muestras que no siguen la bondad de ajuste a la normalidad.



El test de Mann Whitney para muestras que no se ajustan a la normalidad

Figura 17. Ejemplo de *output* de la prueba de Mann-Whitney



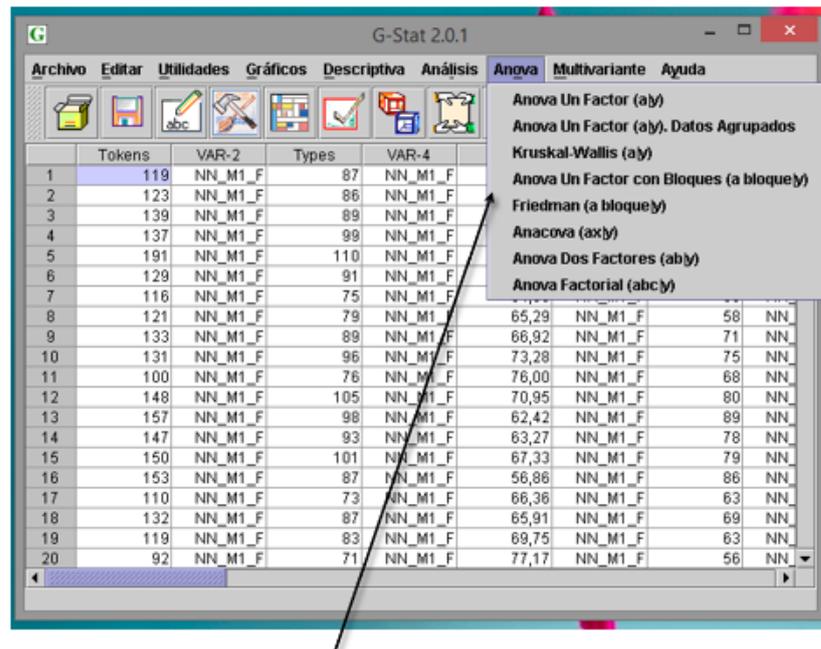
El p-valor indica que hay diferencia estadísticamente significativa

Figura 18. Ejemplo de *output* de la prueba de Mann-Whitney

(iii) La ANOVA es una prueba estadística paramétrica que se utiliza para la comparación de más de dos grupos cuando estos siguen la bondad de ajuste a la normalidad. La ANOVA calcula y compara las medias de los distintos grupos.

La aplicación de G-Stat 2.0 permite también acceder a comparaciones múltiples, pues la ANOVA solo señala que hay diferencias entre los grupos, pero no entre qué grupos

A continuación mostramos dos ejemplos del *output* que se obtiene.



Aquí se muestran las pruebas estadísticas para comparar dos o más grupos.

Figura 19. Ejemplo de la prueba ANOVA

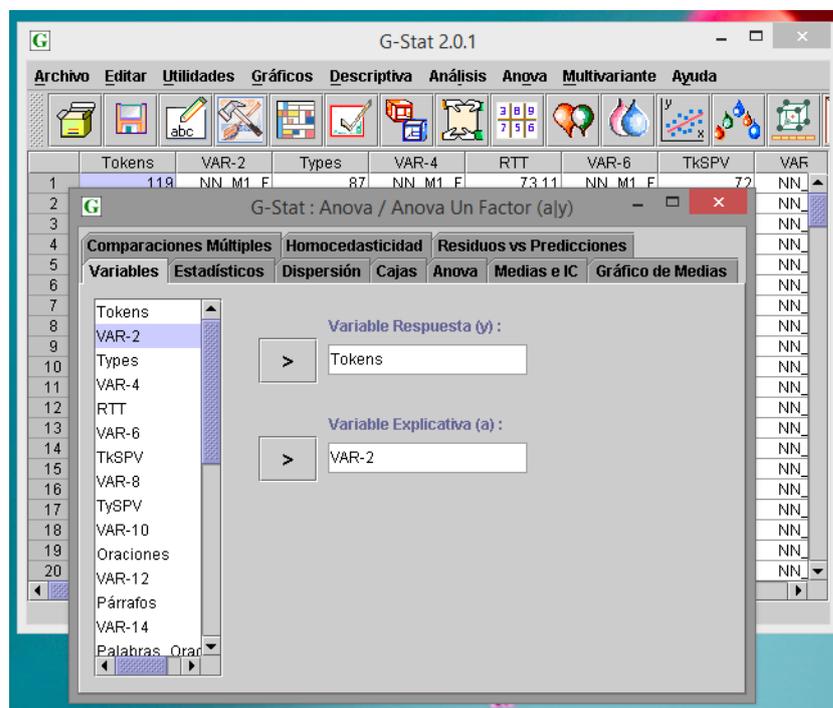
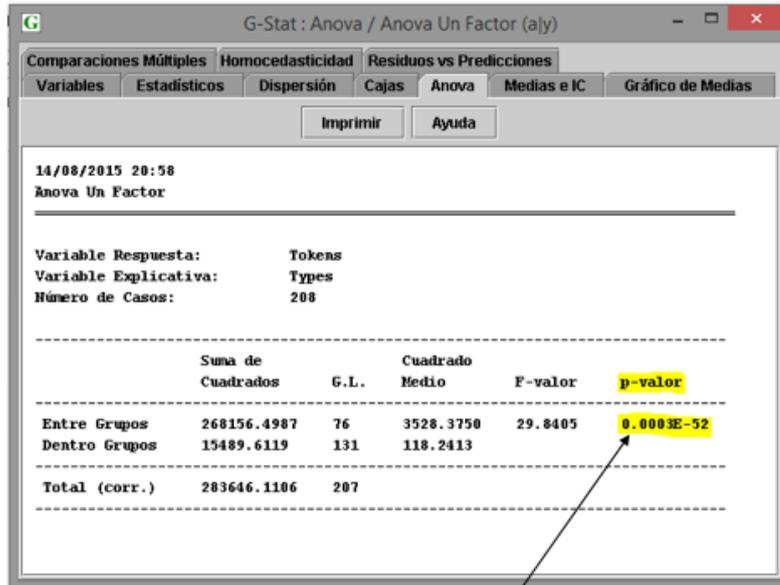
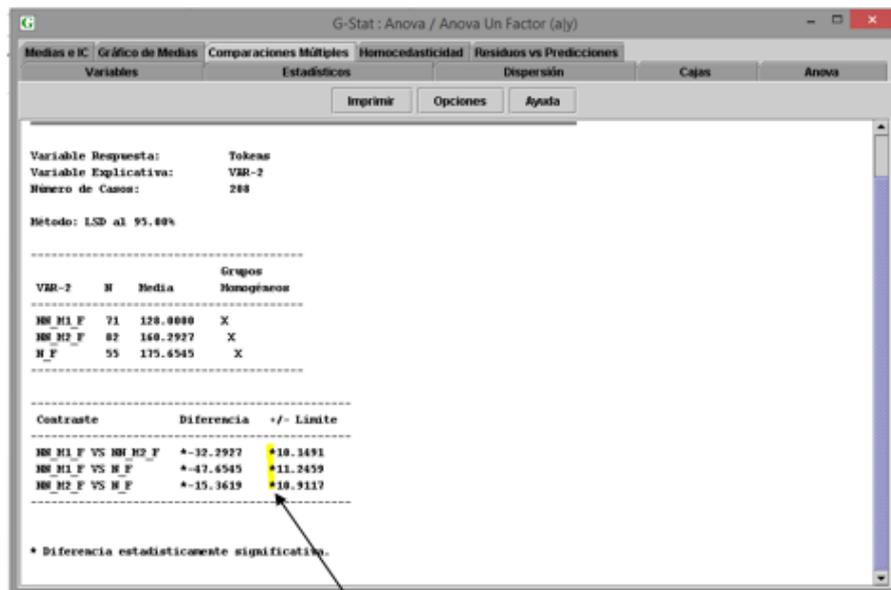


Figura 20. Ejemplo de *output* de la prueba ANOVA



El p-valor es inferior al valor de significación 0.05, por tanto, sabemos que hay diferencia estadísticamente significativa, pero no sabemos entre qué grupos existe esa diferencia.

Figura 21. Ejemplo de *output* de la prueba ANOVA



El asterisco indica entre qué grupos hay diferencia estadísticamente significativa.

Figura 22. Ejemplo de *output* de la prueba ANOVA- Comparaciones múltiples

(iv) La prueba estadística *Kruskal-Wallis* es la prueba equivalente a la ANOVA, pero para muestras que no siguen la bondad a la normalidad. Se trata, por tanto, de una prueba no paramétrica. Esta prueba calcula la igualdad de medianas entre grupos. Las hipótesis de esta prueba son:

$H_0$ : Las  $k$  medianas son todas iguales

$H_1$ : Al menos una de las medianas es diferente

(Manual G-Stat 2.0)

Igual que la ANOVA, la prueba *Kruskal-Wallis* informa solo de si hay o no diferencia, pero no entre qué grupos se da dicha diferencia. Por ello, hay que realizar a posteriori «comparaciones múltiples» que van a indicar entre qué grupos se dan las diferencias si las hay.

A continuación mostramos dos ejemplos del *output* que se obtiene.

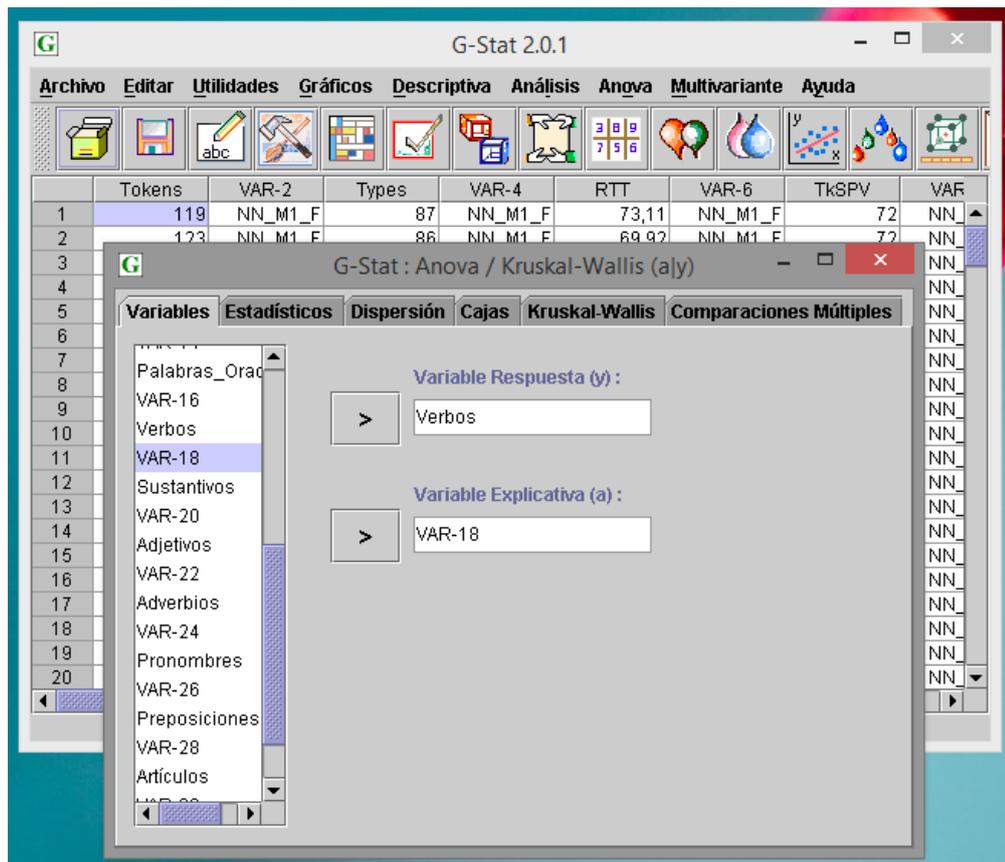
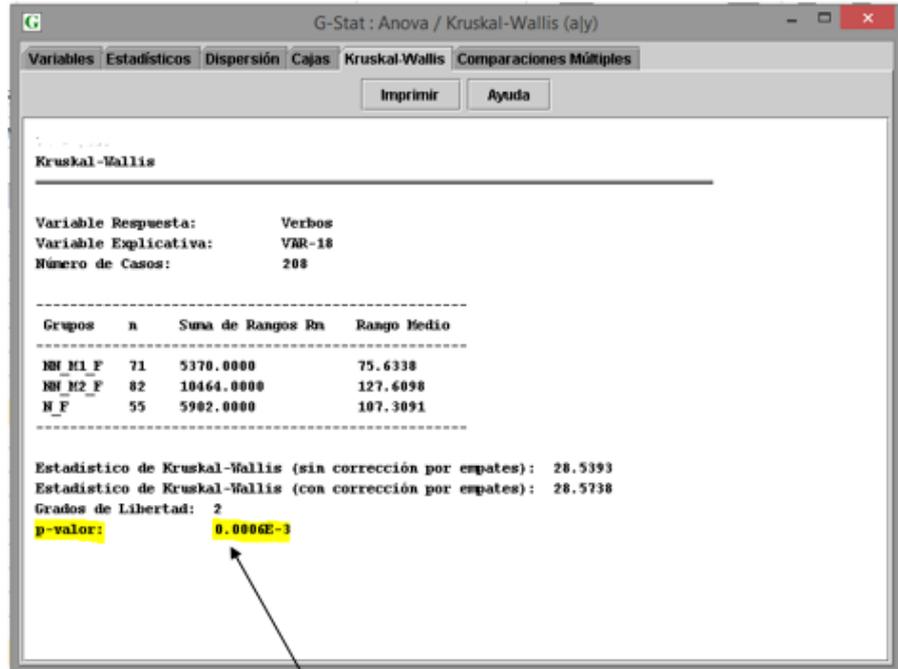
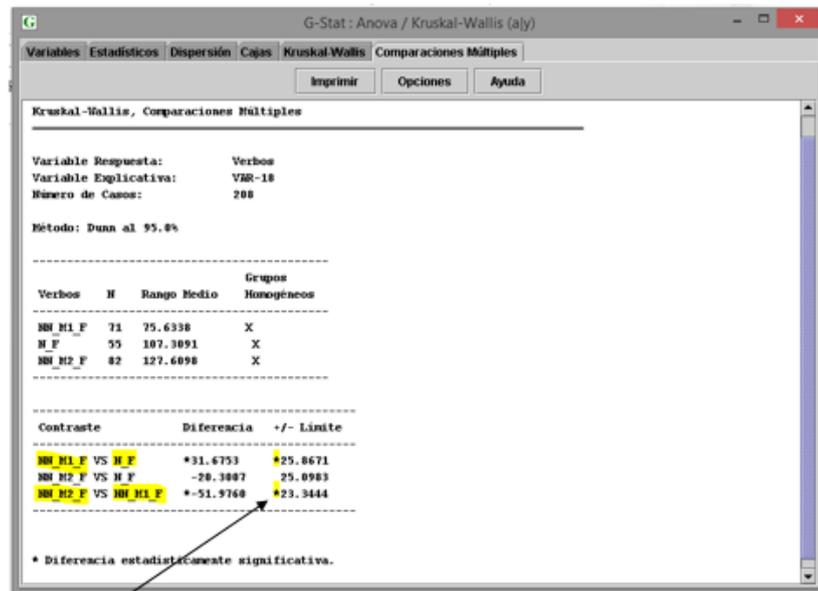


Figura 23. Ejemplo de la prueba de *Kruskal-Wallis*



El p-valor nos indica que hay diferencia entre los grupos.

Figura 24. Ejemplo de *output* de la prueba de *Kruskal-Wallis*



El asterisco indica entre qué grupos hay diferencia estadísticamente significativa.

Figura 25. Ejemplo de *output* de la prueba de *Kruskal-Wallis*- Comparaciones múltiples



## PARTE IV

### RESULTADOS, ANÁLISIS Y DISCUSIÓN

#### CAPÍTULO 6

#### RESULTADOS y ANÁLISIS

##### 6.1 Estadística descriptiva

El primer paso en un análisis estadístico es hacer una estadística descriptiva. Como acabamos de ver en el capítulo anterior, el objetivo principal de ésta es solo «describir». Es decir, mostrar simplemente lo que los datos reflejan para tener una noción de qué es lo que tenemos entre manos.

Así, para poder dar respuesta a las preguntas de investigación y alcanzar los objetivos propuestos, vamos a presentar, en primer lugar, la estadística descriptiva que nos proporcionan los datos recogidos en torno a la tipología textual que los sujetos de la investigación tuvieron que escribir: (i) un relato en pasado a partir de una frase dada, lo cual requería una cierta creatividad por parte del estudiante y (ii) un relato en pasado contando una experiencia personal; aquí el alumno debía recurrir a su memoria, recuperar información. Para esta fase de análisis, solo tenemos en consideración la tipología textual y si se trata de sujetos nativos o no nativos, sin diferenciar en el grupo de no nativos entre los dos momentos de recogida de datos.

Clasificación de las muestras por grupos	
NN-F	NN-1D
N-F	N-1D

Tabla 16. Clasificación de las muestras por grupo y tipo de texto.

A continuación, vamos a presentar la estadística descriptiva diferenciando, en el grupo de no nativos, entre los dos momentos de recogida de datos, al inicio del curso y al finalizar este. Es decir, tras 120 horas de instrucción. Así, aquí vamos a distinguir momentos (M1-M2)

Clasificación de las muestras grupo NN M1-M2
NN-M1
NN-M2

Tabla 17. Clasificación de las muestras por M1-M2 del grupo de NN.

El siguiente paso es presentar la estadística descriptiva por estadios: (i) inicio de nivel B1 (según el MCER); (ii) final de nivel B1 (según el MCER), y (iii) nivel nativo y distinguiendo entre los dos tipos de textos que escribieron los estudiantes.

Clasificación de las muestras por estadios	
NN-M1-F	NN-M1-1D
NN-M2-F	NN-M2-1D
N-F	N-1D

Tabla 18. Clasificación de las muestras por estadios (nivel de lengua) y tipo de texto.

En los casos en los que distinguimos entre los dos tipos de texto que escribieron, la presentación la vamos a hacer en dos fases, primero vamos a mostrar los resultados de las muestras «F» y después los resultados de las muestras «1D». Presentaremos los resultados de las distintas variables seleccionadas para obtener una panorámica general de los dos tipos de texto que tuvieron que producir los sujetos de la investigación.

### 6.1.1 Estadística descriptiva para la muestra «F» por grupos

Como podemos ver en la tabla siguiente, los totales por grupo varían mucho ya que partimos de dos muestras en las que el número total de producciones es muy diferente en un caso y en otro.

	Muestras «F»	
	NN-F (n=153)	N-F (n=55)
<i>Tokens</i>	22.232	9.661
<i>Types</i>	14.749	6.341
Verbos	4.369	1.546
Sustantivos	9.387	3.998
Adjetivos	2.979	1.182
Adverbios	1.337	376
Pronombres	1.703	981
Preposiciones	981	695
Artículos	391	218
Conjunciones	1.085	665

Tabla 19. Totales de la muestra «F»

Así, con números tan diferentes entre el grupo de no nativos y de nativos, vamos a ver a continuación qué dice la estadística descriptiva sobre estos datos; cuáles son los valores de centralización, de dispersión y la forma de su distribución; qué similitudes o diferencias encuentra la estadística en estos valores.

(1) *Tokens*: Número total de palabras.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
<i>Tokens</i>	22.232	9.661

A partir de estos valores absolutos, vamos a ver qué dicen los estadísticos.

En la tabla 20 podemos observar los diferentes descriptores que nos informan de las medidas de tendencia central, de dispersión y de forma.

Para obtener una primera aproximación, los dos descriptores básicos son la media y la desviación típica. Como comprobamos en la tabla 20, las medias de las dos muestras son muy diferentes, siendo la más baja la del grupo de no nativos con un valor media=145.000 frente a la del grupo de nativos que está en media=175.6545. En cuanto a la desviación típica, la más baja corresponde al grupo de nativos ( $s^{29}=29.7080$ ) y el grupo de no nativos presenta una desviación típica mayor ( $s=36.1583$ ). Es decir, que la dispersión de los datos en la muestra de no nativos es mayor. Esto se puede observar en el gráfico 10 de dispersión y también el gráfico 11 de cajas.

**Estadísticos para la variable Tokens**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>145.3072</b>	<b>175.6545</b>
<b>Mediana</b>	<b>139.0000</b>	<b>174.0000</b>
<b>Varianza</b>	<b>1307.4248</b>	<b>882.5636</b>
<b>Desviación Típica</b>	<b>36.1583</b>	<b>29.7080</b>
<b>Mínimo</b>	<b>72.0000</b>	<b>127.0000</b>
<b>Máximo</b>	<b>283.0000</b>	<b>250.0000</b>
<b>Cuartil Inferior</b>	<b>122.0000</b>	<b>152.0000</b>
<b>Cuartil Superior</b>	<b>164.0000</b>	<b>191.0000</b>
<b>Rango Intercuartílico</b>	<b>42.0000</b>	<b>39.0000</b>
<b>Curtosis</b>	<b>1.0932</b>	<b>-0.0715</b>

Tabla 20. Estadísticos para la variable «Tokens» NN - N

<sup>29</sup> La «s» es el símbolo internacional para referirse a la «desviación típica».

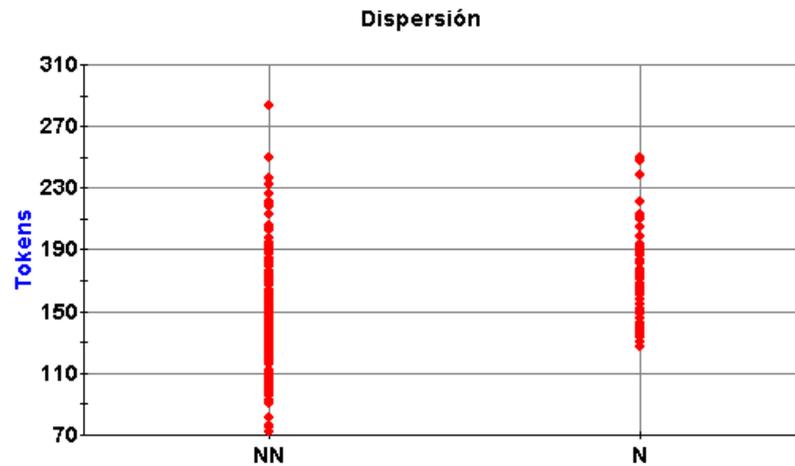


Gráfico 10. Gráfico de dispersión para la variable «Tokens» grupos NN - N

Es precisamente en el gráfico de cajas donde quedan reflejados gráficamente los estadísticos para cada una de las muestras. Como se puede observar, en los dos casos la mediana presenta un valor inferior a la media, aunque en el caso de los nativos esta diferencia es menor. La diferencia entre los valores de la media y la mediana significa que estamos ante casos en los que se dan algunos valores extremos. Esto también se puede observar en el diagrama de cajas.

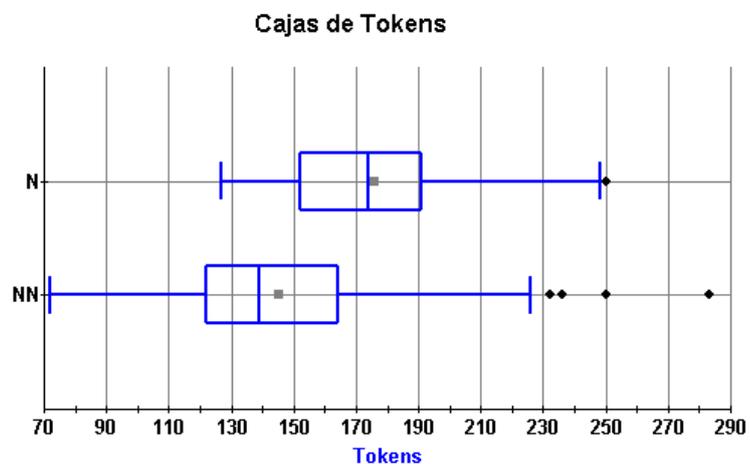


Gráfico 11. Diagrama de Cajas para la variable «Tokens» grupos NN - N

La relación entre media y mediana ayuda a conocer el grado de simetría que presenta la muestra. Si estos valores coinciden o están próximos, podemos decir que la muestra tiende a una distribución simétrica, normal o campana de Gauss. Cuando se alejan, tiende a una distribución asimétrica. Así, vemos que la muestra del grupo de nativos está más próxima a una simetría normal que la del grupo de no nativos.

Al observar el gráfico de cajas, vemos que más allá de los «bigotes» aparecen unos puntos. Esto es así porque los «bigotes» no suelen llegar hasta el extremo, sino hasta las observaciones que se separan de las cajas no más de 1.5 IQR (valor rango intercuartílico). Así, esos puntos lo que indican es que hay datos que quedan fuera del valor 1.5 IQR. Esta dispersión, lógicamente, también se percibe en el gráfico específico de dispersión.

Hay que destacar, asimismo, el valor del estadístico «rango intercuartílico». Este nos informa (i) del tamaño de la caja, pues es el resultado de restar el valor del cuartil superior y el cuartil inferior; (ii) informa de la mayor o menor concentración de los datos respecto a las medidas de centralización (media y mediana). Hay que indicar que dentro de la caja se encuentra el 50% de los valores de la muestra. Así, a un menor tamaño de caja, mayor concentración de los datos en torno a la media y la mediana y cuanto mayor sea la caja, mayor dispersión de dichos datos. Teniendo en cuenta esta información, observamos en la tabla de estadísticos y en el gráfico de cajas que el valor más alto de dicho descriptor es el correspondiente a la muestra del grupo de NN. Es decir, en esta muestra hay mayor dispersión de datos en torno a la media y la mediana que en la muestra del grupo de nativos.

También llama la atención el estadístico curtosis. Vemos que en la muestra del grupo de NN este valor es positivo, mientras que en la muestra del grupo de nativos, es negativo. Recordemos que este valor hace referencia a la forma de la curva. En este caso, la curtosis de la muestra del grupo de NN va a ser más aguda que la de la muestra de N, pues los valores negativos tienden a tener menor número de valores extremos y, por tanto, la forma es más plana. Esto lo podemos observar en los histogramas correspondientes.

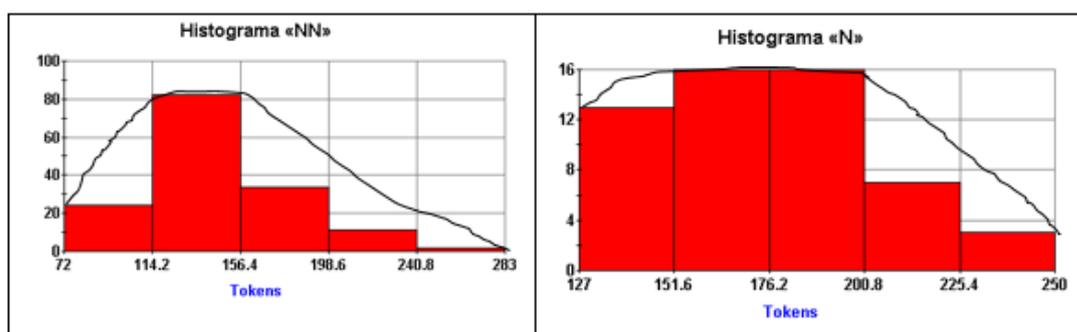


Gráfico 12. Histograma para la variable «Tokens» grupos NN - N

(2) *Types*: Número total de palabras diferentes.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
<i>Types</i>	14.749	6.341

A partir de estos valores absolutos, vamos a ver qué dicen los estadísticos.

Los valores de los descriptores que señalan las medidas de centralización, la media y la mediana, indican que es la muestra del grupo de nativos la que presenta el valor más alto con respecto a la muestra de NN, como podemos ver en la tabla de estadísticos. También observamos que la diferencia entre media y mediana es mínima en la muestra del grupo de nativos; mientras que en la muestra del grupo de no nativos es algo mayor, aunque también pequeña.

Para conocer la dispersión de las muestras, comprobamos los valores del estadístico «desviación típica» y observamos que, al igual que en la variante anterior, el valor más alto corresponde a la muestra del grupo de NN,  $s=19.6128$ , mientras que la muestra del grupo de N presenta un valor de  $s=17.5875$ . Estas pequeñas diferencias las podemos visualizar en el gráfico de dispersión (Gráfico 13).

**Estadísticos para la variable *Types***

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>96.3987</b>	<b>115.2909</b>
<b>Mediana</b>	<b>95.0000</b>	<b>115.0000</b>
<b>Varianza</b>	<b>384.6624</b>	<b>309.3212</b>
<b>Desviación Típica</b>	<b>19.6128</b>	<b>17.5875</b>
<b>Mínimo</b>	<b>51.0000</b>	<b>82.0000</b>
<b>Máximo</b>	<b>153.0000</b>	<b>156.0000</b>
<b>Cuartil Inferior</b>	<b>83.0000</b>	<b>104.0000</b>
<b>Cuartil Superior</b>	<b>109.0000</b>	<b>124.0000</b>
<b>Rango Intercuartílico</b>	<b>26.0000</b>	<b>20.0000</b>
<b>Curtosis</b>	<b>0.1035</b>	<b>-0.1796</b>

Tabla 21. Tabla de estadísticos para la variable «*Types*» NN - N

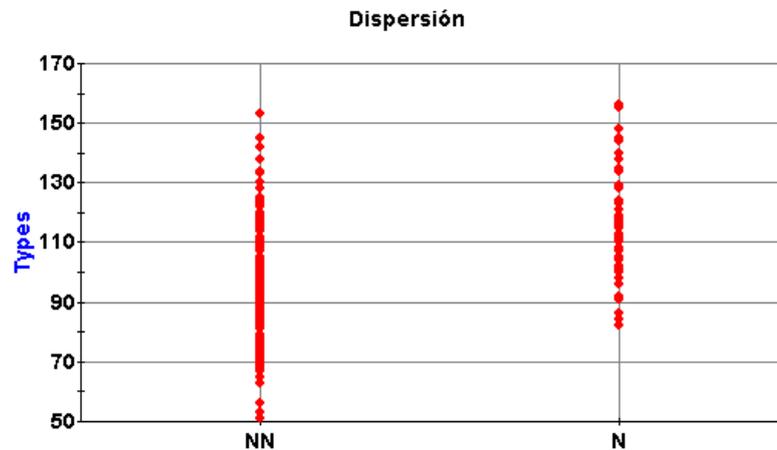


Gráfico 13. Gráfico de dispersión para la variable «Types» NN - N

A continuación mostramos el diagrama de cajas para ver gráficamente cómo quedan representados los estadísticos.

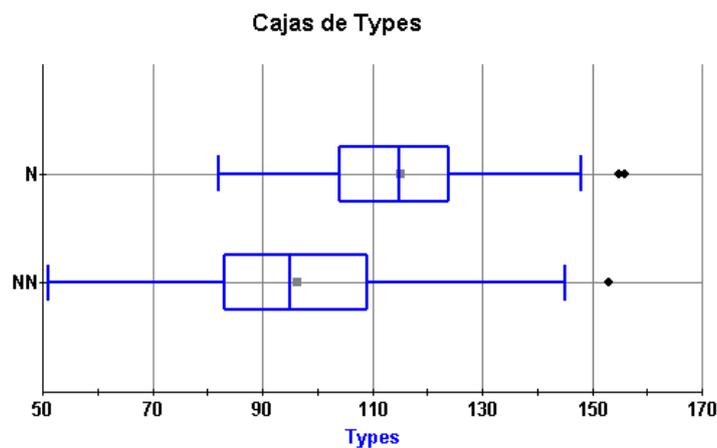


Gráfico 14. Diagrama de cajas para la variable «Types» NN - N

En el diagramas de cajas podemos confirmar la casi coincidencia de la media y la mediana en las dos muestras, como hemos indicado más arriba.

El valor del otro estadístico que informa de la dispersión, el rango intercuartílico, vemos que es mayor en la muestra del grupo de NN (IQR=26.0000) que en la muestra del grupo de N (IQR=20.0000). Así, el 50% de los valores que se encuentran en torno a la media y la mediana presentan mayor dispersión en el grupo de no nativos.

Llama la atención también, al observar el gráfico de cajas, la diferencia que hay entre ambas muestras en el descriptor «mínimo», algo más de treinta puntos; mientras que en el descriptor «máximo» solo hay una diferencia de tres puntos.

En cuanto al valor del estadístico «curtosis», que indica la tendencia de la forma de la muestra, vuelve a ser positivo en la muestra del grupo de NN (curtosis=0.1035) y negativo en la muestra del grupo de N (curtosis= -0.1796). Eso nos indica una tendencia a una curva más aguda o más plana, respectivamente, en relación con la distribución normal o campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

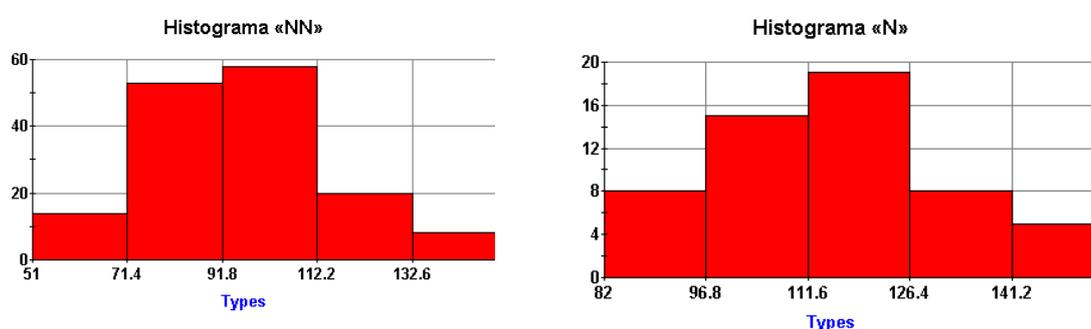


Gráfico 15 Histogramas para la variable «Types» NN - N

(3) Verbos: Número total de verbos. °

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Verbos	4.369	1.546

En la tabla 22 mostramos los estadísticos para comprobar qué diferencias o similitudes encontramos en las medidas de centralización, dispersión y forma.

Respecto a las medidas de centralización, llama la atención, por un lado, la proximidad entre los valores de ambas muestras y, por otro lado, también la cercanía entre los valores de media y mediana entre sí.

En cuanto al estadístico «desviación típica», comprobamos nuevamente que el valor más alto corresponde a la muestra del grupo de NN con un valor de  $s=12.8954$  frente al grupo de N con un valor de  $s=8.2498$ . La mayor dispersión vuelve a presentarla el grupo de no nativos y lo podemos observar en el siguiente gráfico.

**Estadísticos para la variable Verbos**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>41.8431</b>	<b>41.4000</b>
<b>Mediana</b>	<b>40.0000</b>	<b>41.0000</b>
<b>Varianza</b>	<b>166.2910</b>	<b>68.0593</b>
<b>Desviación Típica</b>	<b>12.8954</b>	<b>8.2498</b>
<b>Mínimo</b>	<b>17.0000</b>	<b>28.0000</b>
<b>Máximo</b>	<b>95.0000</b>	<b>65.0000</b>
<b>Cuartil Inferior</b>	<b>33.0000</b>	<b>36.0000</b>
<b>Cuartil Superior</b>	<b>49.0000</b>	<b>44.0000</b>
<b>Rango Intercuartílico</b>	<b>16.0000</b>	<b>8.0000</b>
<b>Curtosis</b>	<b>1.3212</b>	<b>0.7899</b>

Tabla 22. Tabla de estadísticos para la variable «Verbos» NN - N

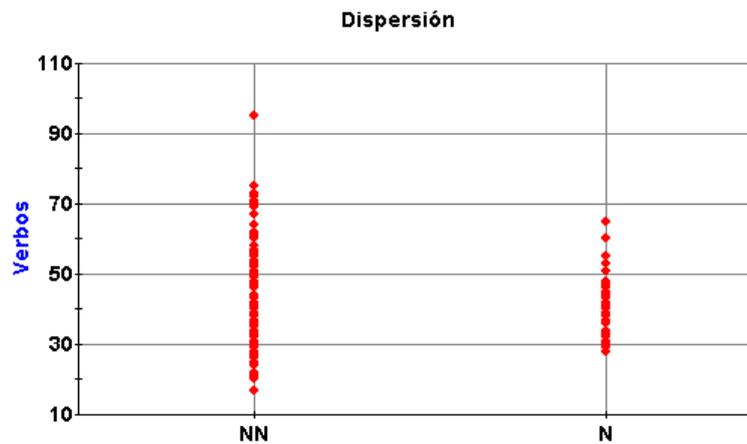


Gráfico 16. Gráfico de dispersión para la variable «Verbos» NN - N

El otro estadístico que informa de la dispersión de las muestras, el «rango intercuartílico», presenta una diferencia de ocho puntos entre ambas muestras. El valor más alto (IQR=16.0000) corresponde al grupo de NN. Esta diferencia la podemos visualizar en el gráfico de cajas que presentamos a continuación.

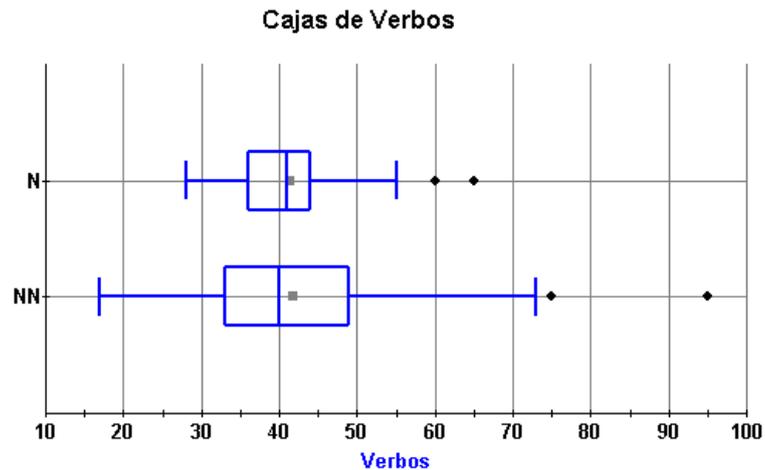


Gráfico 17. Diagrama de cajas para la variable «Verbos» NN - N

Como vemos al visualizar el tamaño de las cajas, la dispersión del 50% de los valores de la muestra es mayor en el grupo de NN.

También podemos observar las diferencias existentes entre media y mediana en cada una de las muestras, como hemos señalado más arriba, además de la proximidad de estos valores entre muestras.

El estadístico «curtosis», que se refiere a la forma, presenta valores positivos en las dos muestras, aunque en la muestra del grupo de N el valor (curtosis=0.7899) está más próximo a una curtosis cero o de campana de Gauss que el valor de la muestra del grupo de NN (curtosis=1.3212). Eso significa que la curva que representa los valores positivos de ambas muestras tiende a una forma más aguda que una curtosis cero. Los histogramas que mostramos a continuación pueden dar una idea de cómo es la curva con estos valores de «curtosis».

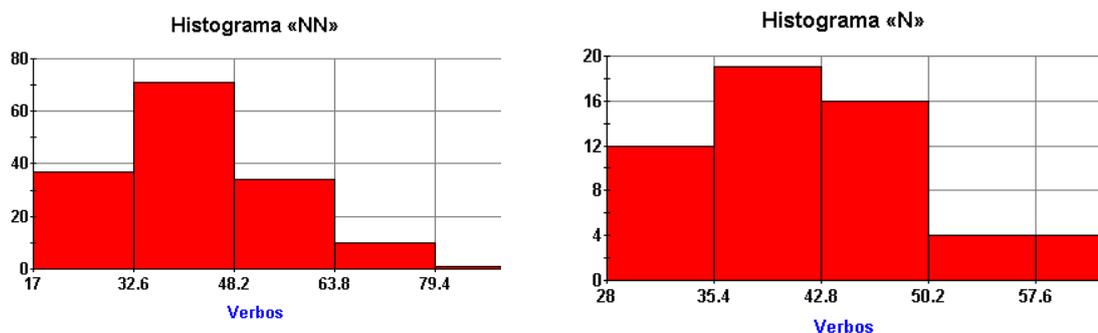


Gráfico 18. Histograma para la variable «Verbos» NN -N

(4) Sustantivos: Número total de sustantivos.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Sustantivos	9.387	3.998

Ante estos totales, vamos a ver qué dicen los estadísticos que encontramos en la siguiente tabla.

**Estadísticos para la variable Sustantivos**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>75.7647</b>	<b>87.3273</b>
<b>Mediana</b>	<b>74.0000</b>	<b>89.0000</b>
<b>Varianza</b>	<b>351.1285</b>	<b>238.7057</b>
<b>Desviación Típica</b>	<b>18.7384</b>	<b>15.4501</b>
<b>Mínimo</b>	<b>33.0000</b>	<b>56.0000</b>
<b>Máximo</b>	<b>132.0000</b>	<b>132.0000</b>
<b>Cuartil Inferior</b>	<b>63.0000</b>	<b>77.0000</b>
<b>Cuartil Superior</b>	<b>87.0000</b>	<b>97.0000</b>
<b>Rango Intercuartílico</b>	<b>24.0000</b>	<b>20.0000</b>
<b>Curtosis</b>	<b>0.3846</b>	<b>0.5846</b>

Tabla 23. Tabla de estadísticos para la variable «Sustantivos» NN - N

Al observar los estadísticos de centralización, la media y la mediana, llama la atención los valores tan diferentes que presenta esta variable respecto a las anteriores. Entre el valor más alto, correspondiente a la muestra del grupo de nativos, y el valor del grupo de no nativos hay una diferencia de casi doce puntos.

Al comparar la distancia entre la media y la mediana dentro de cada muestra, vemos que la diferencia es muy similar en una y otra muestra.

En cuanto a la desviación típica de cada una de las muestras, observamos que el valor más alto ( $s=18.7384$ ) lo encontramos, una vez más, en la muestra del grupo de NN. La muestra del grupo de N presenta un valor de  $s=15.4501$ . Podemos visualizar estos datos en el gráfico de dispersión que mostramos a continuación.

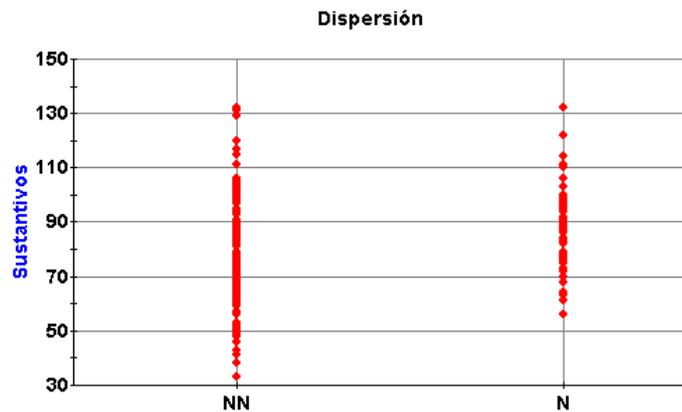


Gráfico 19. Gráfico de dispersión para la variable «Sustantivos» NN - N

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta valores diferentes para cada muestra. Podemos observar en la tabla de estadísticos que la del grupo de no nativos supera en cuatro puntos (IQR=24.0000) el valor de la muestra del grupo de nativos (IQR=20.0000). Esto significa, por tanto, que el tamaño de la caja de la muestra del grupo de NN es mayor y, por consiguiente, el 50% de los valores de dicha muestra presenta una mayor dispersión en comparación con la muestra del grupo de N. En el diagrama de cajas (gráfico 20) podemos ver gráficamente esta información.

Al observar el diagrama de cajas, constatamos la diferencia de tamaño entre la caja de una muestra y la otra. También podemos ver cierta similitud en los valores del estadístico «máximo». Mientras que en los del descriptor «mínimo» hay una distancia llamativa.

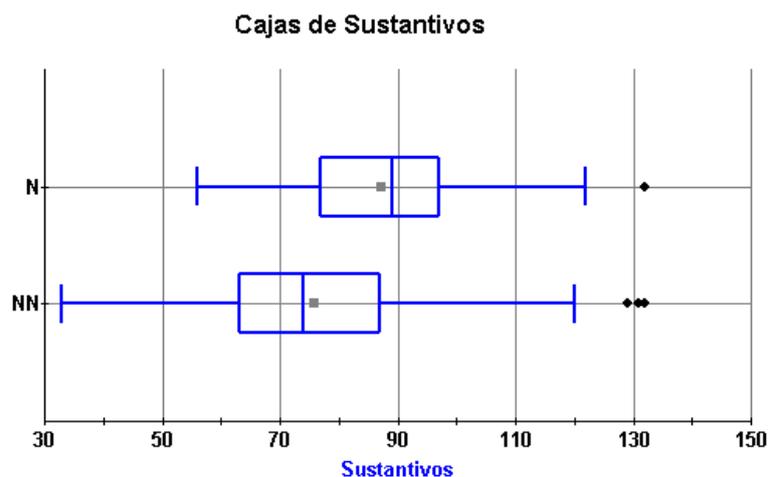


Gráfico 20. Diagrama de cajas para la variable «Sustantivos» NN - N

También podemos visualizar las pequeñas diferencias que hemos señalado más arriba entre la media y la mediana en cada una de las muestras.

En cuanto a la forma de la distribución de las muestras en comparación con una distribución normal o de campana de Gauss, las dos muestras de nuestro estudio presentan una curtosis con valores positivos, es decir, que la curva tiene tendencia a ser más aguda que en una distribución normal con valor cero. No obstante hay que señalar que aun siendo valores positivos, están muy próximos al cero: curtosis= 0.3846, grupo de NN; curtosis=0.5846, grupo N. En los histogramas siguientes podemos apreciar esta tendencia.

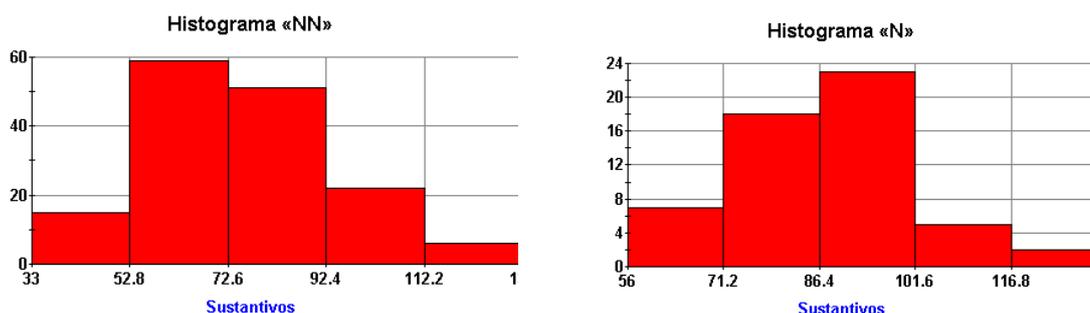


Gráfico 21. Histograma para la variable «Sustantivos» NN - N

(5) Adjetivos: Número total de adjetivos

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Adjetivos	2.979	1.182

En la tabla siguiente encontramos los estadísticos que informan sobre las medidas de centralización, dispersión y forma de la variable «adjetivos» de las dos muestras.

**Estadísticos para la variable Adjetivos**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>33.9477</b>	<b>36.1636</b>
<b>Mediana</b>	<b>33.0000</b>	<b>35.0000</b>
<b>Varianza</b>	<b>114.5894</b>	<b>74.6579</b>
<b>Desviación Típica</b>	<b>10.7046</b>	<b>8.6405</b>
<b>Mínimo</b>	<b>13.0000</b>	<b>20.0000</b>
<b>Máximo</b>	<b>76.0000</b>	<b>72.0000</b>
<b>Cuartil Inferior</b>	<b>27.0000</b>	<b>31.0000</b>
<b>Cuartil Superior</b>	<b>41.0000</b>	<b>41.0000</b>
<b>Rango Intercuartílico</b>	<b>14.0000</b>	<b>10.0000</b>
<b>Curtosis</b>	<b>0.9342</b>	<b>4.5193</b>

Tabla 24. Tabla de estadísticos para la variable «Adjetivos» NN - N

Las medidas de centralización, la media y la mediana, no son muy diferentes en cada una de las muestras. El valor más alto lo encontramos en la muestra del grupo de nativos con una media=36.1636 y una mediana=35.0000. Mientras que la muestra del grupo de NN presenta una media=33.9477 y una mediana=33.0000.

Otro elemento a resaltar es la relación entre la media y la mediana en cada una de las muestras. Podemos observar que la diferencia no es mucha, pero es en la muestra del grupo de N donde encontramos una diferencia algo mayor, algo más de un punto, mientras que en la muestra del grupo de NN no llega al punto de diferencia.

El siguiente estadístico que nos ayuda a entender cómo se distribuyen las muestras es la desviación típica que nos habla de la dispersión de las mismas. Como podemos ver en la tabla de estadísticos, el valor más bajo ( $s=8.6405$ ), y por tanto, menos dispersión, lo encontramos en la muestra del grupo de nativos. La muestra del grupo de NN presenta un valor de  $s=10.7046$ , unos dos puntos más que la otra muestra. Estos datos los podemos visualizar en el gráfico de dispersión.

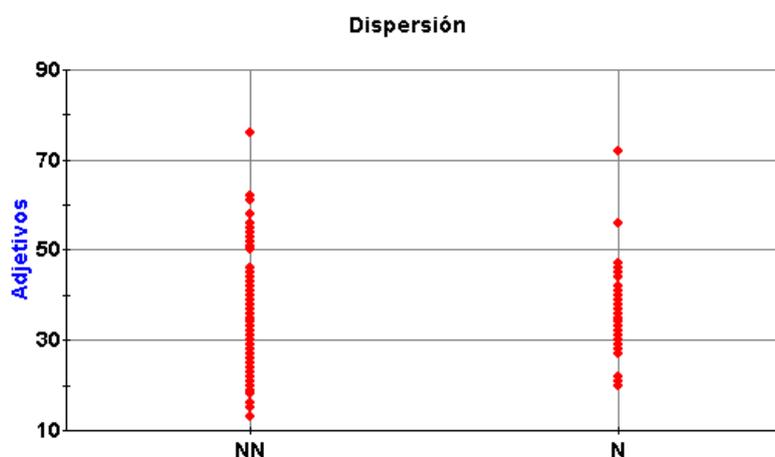


Gráfico 22. Gráfico de dispersión para la variable «Adjetivos» NN - N

También se puede visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras.

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta una diferencia de cuatro puntos entre una muestra y la otra. Mientras que la muestra del grupo de NN registra un valor IQR=14.0000, la muestra del grupo de N presenta un IQR=10.0000. Esto significa, como hemos visto también en las anteriores variables, que la caja de la muestra del grupo de no nativos es mayor y, por tanto, los valores de la muestra,

el 50%, insertos en ella presentan una mayor dispersión. En el diagrama de cajas (gráfico 23) podemos ver gráficamente esta información.

También podemos comprobar en el gráfico de cajas la relación de proximidad entre muestras de las medidas de centralización, como ya hemos señalado más arriba. También es llamativo la cercanía entre la media y la mediana de cada una de las muestras. A pesar de esta proximidad, estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

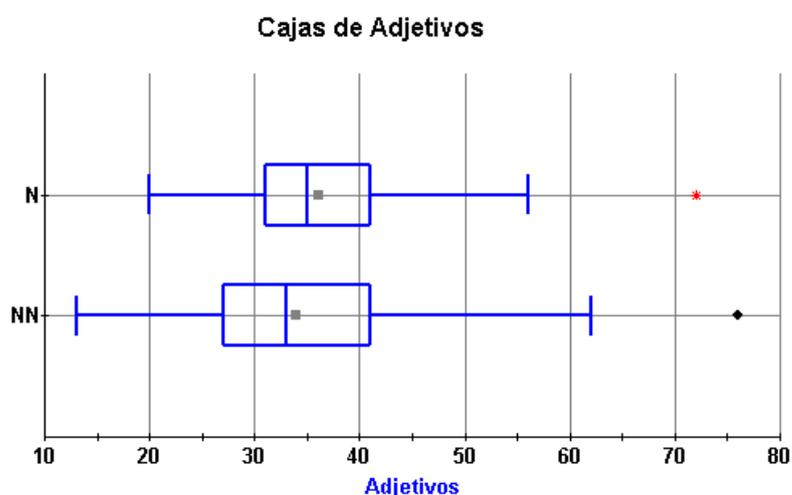


Gráfico 23. Diagrama de Cajas para la variable «Adjetivos» NN - N

Respecto a la forma que adopta cada una de las muestras, vemos en la tabla de estadísticos (Tabla 24) que la curtosis es positiva en los dos casos. Eso nos indica una tendencia a una curva más aguda en relación con la distribución normal o campana de Gauss. Pero vemos que los valores son muy diferentes. Mientras que el valor de la muestra del grupo de NN (curtosis=0.9342) está próximo al valor de curtosis cero (distribución normal y campana de Gauss), la muestra del grupo de N (curtosis=4.5193) se aleja del valor cero. En los histogramas siguientes podemos visualizar esa tendencia.

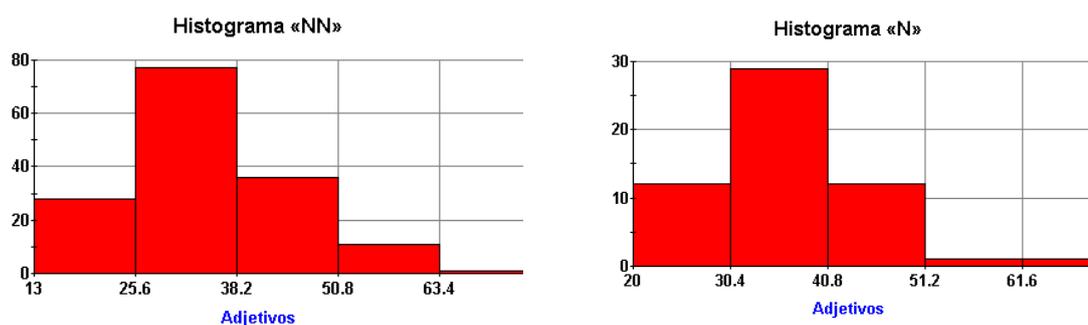


Gráfico 24. Histograma para la variable «Adjetivos» NN - N

(6) Adverbios: Número total de adverbios.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Adverbios	1.337	376

En la tabla siguiente encontramos los estadísticos que informan sobre las medidas de centralización, dispersión y forma de la variable «adverbios» de las dos muestras.

**Estadísticos para la variable Adverbios**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>21.4706</b>	<b>21.4000</b>
<b>Mediana</b>	<b>20.0000</b>	<b>21.0000</b>
<b>Varianza</b>	<b>54.3560</b>	<b>27.0593</b>
<b>Desviación Típica</b>	<b>7.3727</b>	<b>5.2019</b>
<b>Mínimo</b>	<b>7.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>49.0000</b>	<b>36.0000</b>
<b>Cuartil Inferior</b>	<b>16.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>26.0000</b>	<b>24.0000</b>
<b>Rango Intercuartílico</b>	<b>10.0000</b>	<b>6.0000</b>
<b>Curtosis</b>	<b>1.3383</b>	<b>0.3006</b>

Tabla 25. Tabla de estadísticos para la variable «Adverbios» NN - N

Lo primero que llama la atención sobre las medidas de centralización, la media y la mediana, es la proximidad de los valores de cada una de las muestras. La media de ambas muestras casi coincide. Este rasgo lo podemos comprobar en el gráfico de cajas (gráfico 26).

También podemos resaltar la relación entre la media y la mediana en cada una de las muestras. Como se puede observar la diferencia no es mucha en ninguno de los dos casos,

pero es en la muestra del grupo de NN donde encontramos una diferencia algo mayor, algo más de un punto, mientras que en la muestra del grupo de N la diferencia es mínima.

Al observar los estadísticos que informan de la distribución de las muestras, la desviación típica y el rango intercuartílico, vemos que en el caso de la desviación típica el valor más bajo ( $s=5.2019$ ) corresponde a la muestra del grupo de N. Mientras que el grupo de NN ( $s=7.3727$ ) registra algo más de dos puntos respecto a la otra muestra. Por tanto, presenta mayor dispersión de sus valores, como vamos viendo en las variables analizadas hasta el momento. Estos datos los podemos visualizar en el gráfico de dispersión.

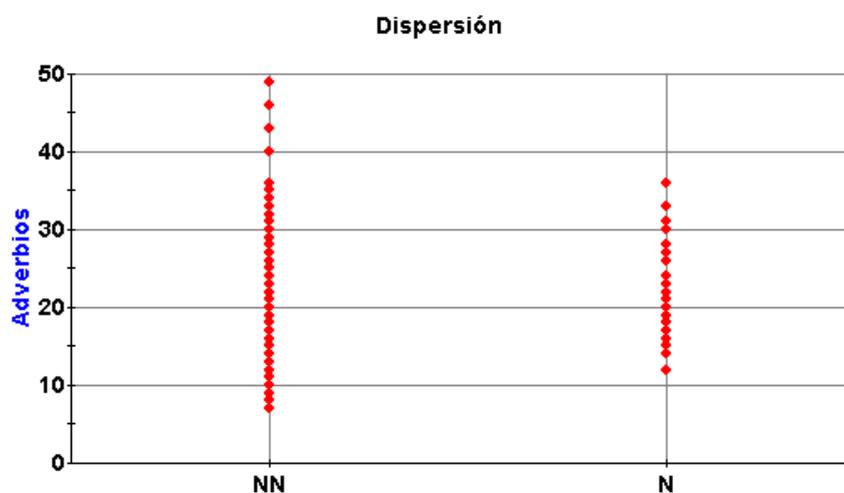


Gráfico 25. Gráfico de dispersión para la variable «Adverbios» NN - N

También podemos visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras.

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta una diferencia de cuatro puntos entre una muestra y la otra. Mientras que la muestra del grupo de NN registra un valor  $IQR=10.0000$ , la muestra del grupo de N presenta un  $IQR=6.0000$ . Esto significa, como hemos visto también en las anteriores variables, que la caja de la muestra del grupo de no nativos es mayor y, por tanto, los valores de la muestra, el 50%, insertos en ella presentan una mayor dispersión. En el diagrama de cajas (gráfico 26) podemos ver gráficamente esta información.

También podemos comprobar en el gráfico de cajas la proximidad entre muestras de las medidas de centralización, como ya hemos señalado más arriba. También es llamativo la cercanía entre la media y la mediana de la muestra del grupo de N, mientras que en la muestra del grupo de NN vemos que hay cierta distancia entre la media y la mediana. Esa pequeña

diferencia indica una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

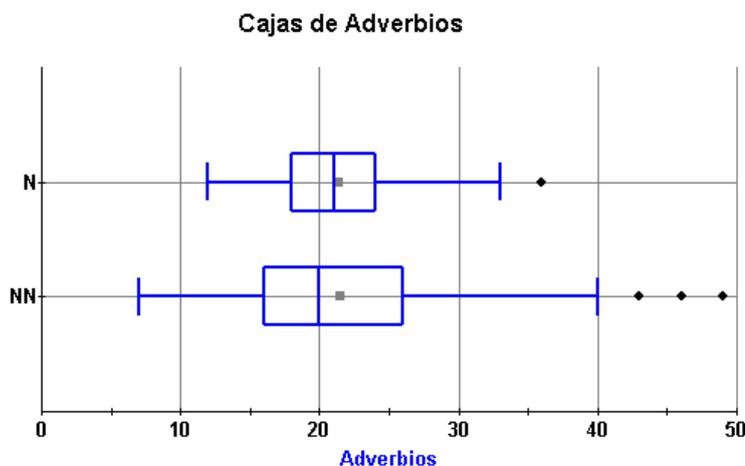


Gráfico 26. Diagrama de Cajas para la variable «Adverbios» NN - N

En cuanto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 26) que la curtosis es positiva en los dos casos. Eso nos indica una tendencia a una curva más aguda en relación con la distribución normal o campana de Gauss. Respecto a la variable anterior, adjetivos, aquí es la muestra del grupo de N (curtosis=0.3006) la que tiene el valor más próximo a una curtosis cero (distribución normal y campana de Gauss). Mientras que el valor de la muestra del grupo de NN (curtosis=1.3383) registra un valor algo más alto. En los histogramas siguientes podemos visualizar esa tendencia.

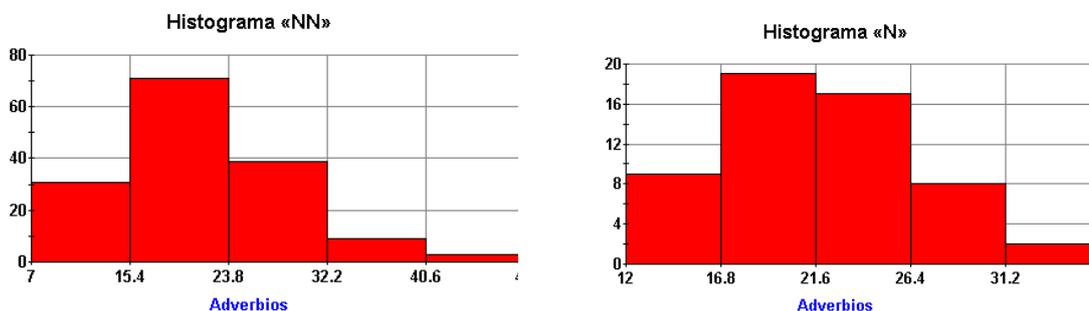


Gráfico 27. Histograma para la variable «Adverbios» NN – N

(7) Pronombres: Número total de pronombres.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Pronombres	1.703	981

En la tabla siguiente encontramos los estadísticos que informan sobre las medidas de centralización, dispersión y forma de la variable «pronombres» de las dos muestras.

**Estadísticos para la variable Pronombres**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>26.0784</b>	<b>33.4182</b>
<b>Mediana</b>	<b>25.0000</b>	<b>33.0000</b>
<b>Varianza</b>	<b>80.9280</b>	<b>83.2478</b>
<b>Desviación Típica</b>	<b>8.9960</b>	<b>9.1240</b>
<b>Mínimo</b>	<b>10.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>64.0000</b>	<b>56.0000</b>
<b>Cuartil Inferior</b>	<b>19.0000</b>	<b>27.0000</b>
<b>Cuartil Superior</b>	<b>31.0000</b>	<b>39.0000</b>
<b>Rango Intercuartílico</b>	<b>12.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>1.5861</b>	<b>0.5550</b>

Tabla 26. Tabla de estadísticos para la variable «Pronombres» NN - N

Al observar los estadísticos de centralización, la media y la mediana, llama la atención los valores tan diferentes que presenta esta variable respecto a la anterior, adverbios, donde eran casi coincidentes. Entre el valor más alto, correspondiente a la muestra del grupo de nativos, y el valor del grupo de no nativos hay una diferencia de más de siete puntos.

En cambio, al comparar la distancia entre la media y la mediana dentro de cada muestra, vemos que la diferencia es mínima en ambas muestras.

En cuanto a la desviación típica de cada una de las muestras, observamos en este caso que los dos valores están muy próximos uno al otro. El valor más alto ( $s=9.1240$ ) lo encontramos en la muestra del grupo de N, pero con poca diferencia respecto a la muestra del grupo de NN ( $s=8.9960$ ). Podemos visualizar estos datos en el gráfico de dispersión que mostramos a continuación.

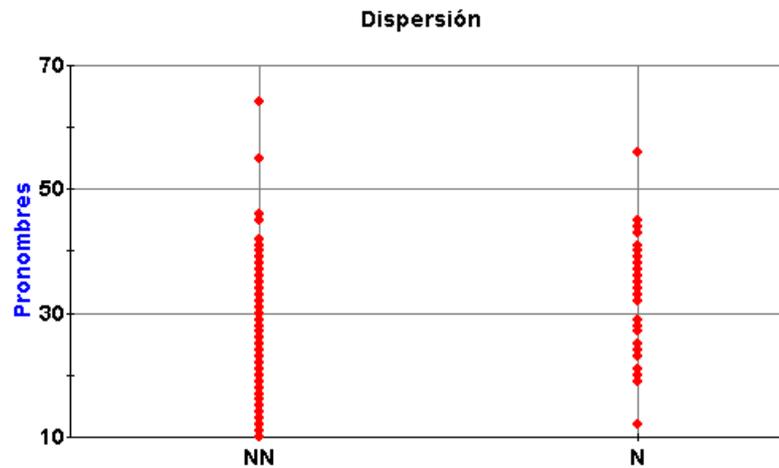


Gráfico 28. Gráfico de dispersión para la variable «Pronombres» NN -N

También se puede visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras.

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta absoluta coincidencia de valores,  $IQR=12.0000$ , en ambas muestras. Esto significa que el tamaño de la caja es el mismo en la muestra del grupo de no nativos y de nativos. Así, el 50%, de los valores inserto en ellas presenta la misma dispersión. En el diagrama de cajas que mostramos a continuación podemos ver gráficamente esta información.

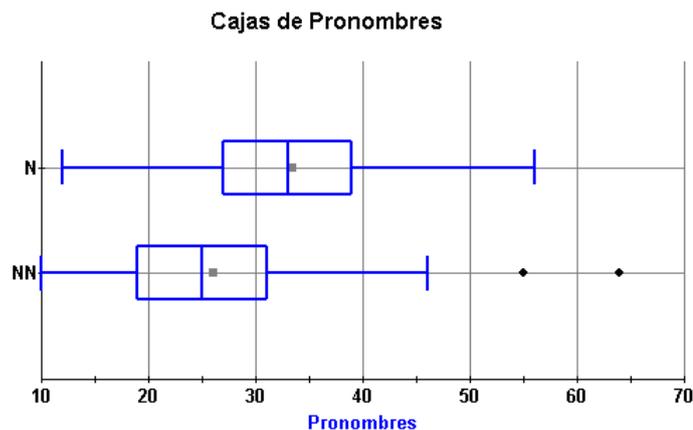


Gráfico 29. Diagrama de Cajas para la variable «Pronombres» NN - N

También podemos comprobar en el gráfico de cajas la cercanía entre los valores de la media y la mediana de cada una de las muestras. A pesar de esta proximidad, estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss. Otro dato a resaltar al observar el diagrama de cajas es la proximidad de los valores del estadístico «mínimo» en ambas muestras.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 26) que la curtosis es positiva en los dos casos. Eso nos indica una tendencia a una curva más aguda en relación con la distribución normal o campana de Gauss. Entre ambos valores hay una diferencia de poco más de un punto, siendo el valor más alto, y por tanto más alejado de una curtosis cero, el de la muestra del grupo de NN (curtosis=1.5861). Mientras que el valor de la muestra del grupo de N es curtosis=0.5550. En los histogramas siguientes podemos visualizar esa tendencia.

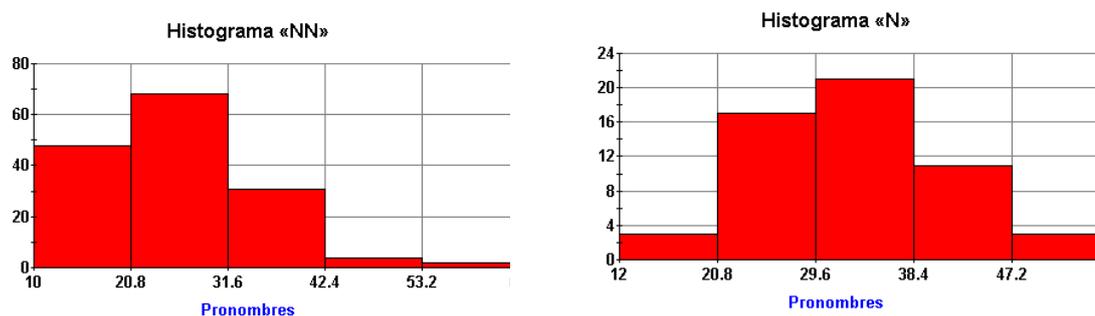


Gráfico 30. Histograma para la variable «Pronombres» NN - N

(8) Preposiciones: Número total de preposiciones.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Preposiciones	981	695

En la tabla de estadísticos (Tabla 27) encontramos los descriptores que informan sobre las medidas de centralización, dispersión y forma de la variable «preposiciones» de las dos muestras.

Las medidas de centralización, la media y la mediana, son bastante diferentes en cada una de las muestras. El valor más alto lo encontramos en la muestra del grupo de nativos con una media=28.7273 y una mediana=29.0000. Mientras que la muestra del grupo de NN presenta una media=20.5294 y una mediana=20.0000.

Otro elemento a resaltar es la relación entre la media y la mediana en cada una de las muestras. Podemos observar que la diferencia es mínima en ambas muestras, ninguna llega al punto, pero la que presenta una distancia algo mayor es la muestra del grupo de NN.

El siguiente estadístico que nos ayuda a entender cómo se distribuyen las muestras es la desviación típica que nos habla de la dispersión de las mismas. Como podemos ver en la tabla de estadísticos, el valor más bajo ( $s=5.8356$ ), y por tanto, menos dispersión, lo

encontramos en la muestra del grupo de nativos. La muestra del grupo de NN presenta un valor de  $s=7.1045$ . Estos datos los podemos visualizar en el gráfico de dispersión (Gráfico 31).

**Estadísticos para la variable Preposiciones**

Grupos	NN	N
N	153	55
Media	20.5294	28.7273
Mediana	20.0000	29.0000
Varianza	50.4745	34.0539
Desviación Típica	7.1045	5.8356
Mínimo	6.0000	18.0000
Máximo	41.0000	40.0000
Cuartil Inferior	16.0000	25.0000
Cuartil Superior	24.0000	33.0000
Rango Intercuartilico	8.0000	8.0000
Curtosis	0.5132	-0.6398

Tabla 27. Tabla de estadísticos para la variable «Preposiciones» NN - N

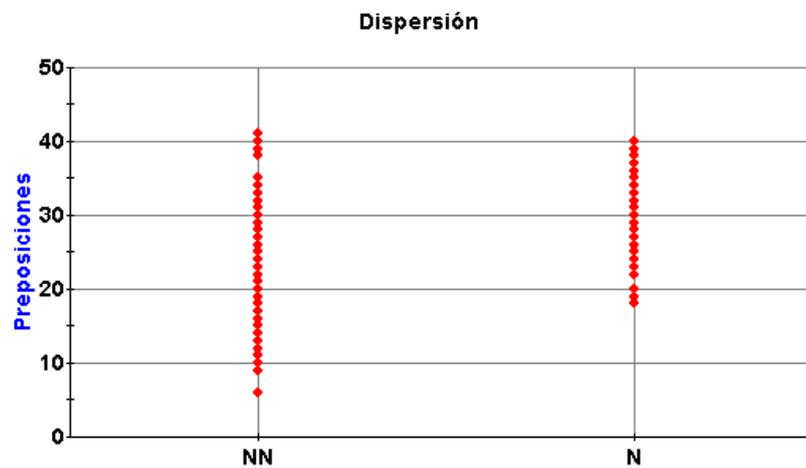


Gráfico 31. Gráfico de dispersión para la variable «Preposiciones» NN - N

También podemos visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras, que en este caso son pocos.

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, igual que en la variante anterior, pronombres, presenta absoluta coincidencia de valores,

IQR=8.0000, en ambas muestras. Esto significa que el tamaño de la caja es el mismo en la muestra del grupo de no nativos y de nativos. Así, el 50%, de los valores inserto en ellas presenta la misma dispersión. En el diagrama de cajas que mostramos a continuación podemos ver gráficamente esta información.

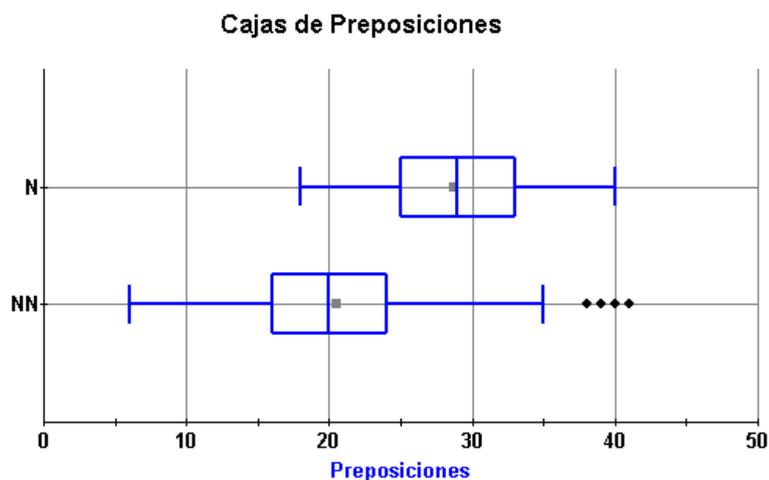


Gráfico 32. Diagrama de Cajas para la variable «Preposiciones» NN - N

También podemos comprobar en el gráfico de cajas la distancia entre muestras de las medidas de centralización, como ya hemos señalado más arriba. Sin embargo, sí hay que resaltar una vez más la cercanía entre la media y la mediana de las dos muestras. Esa pequeña diferencia indica una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 27) que la curtosis es positiva en la muestra del grupo de NN (curtosis=0.5132) y negativa en la muestra del grupo de N (curtosis= -0.6398), aunque en ambos casos el valor está próximo a una curtosis cero (distribución normal o campana de Gauss). No obstante, la curva de la muestra del grupo de NN tiende a una forma más aguda y la de la muestra del grupo de N a una más plana. En los histogramas siguientes podemos visualizar esa tendencia.

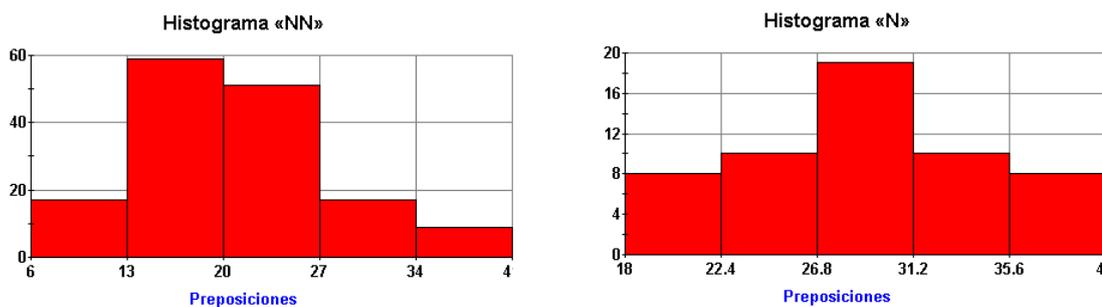


Gráfico 33. Histograma para la variable «Preposiciones» NN - N

(9) Artículos: Número total de artículos.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Artículos	391	218

En la tabla siguiente encontramos los estadísticos que informan sobre las medidas de centralización, dispersión y forma de la variable «artículos» de las dos muestras.

**Estadísticos para la variable Artículos**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>14.0915</b>	<b>18.1455</b>
<b>Mediana</b>	<b>15.0000</b>	<b>18.0000</b>
<b>Varianza</b>	<b>28.0574</b>	<b>18.2377</b>
<b>Desviación Típica</b>	<b>5.2969</b>	<b>4.2706</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>10.0000</b>
<b>Máximo</b>	<b>29.0000</b>	<b>31.0000</b>
<b>Cuartil Inferior</b>	<b>11.0000</b>	<b>16.0000</b>
<b>Cuartil Superior</b>	<b>17.0000</b>	<b>20.0000</b>
<b>Rango Intercuartilico</b>	<b>6.0000</b>	<b>4.0000</b>
<b>Curtosis</b>	<b>0.1070</b>	<b>0.3607</b>

Tabla 28. Tabla de estadísticos para la variable «Artículos» NN - N

Podemos ver que las medidas de centralización, la media y la mediana, son diferentes en cada una de las muestras. El valor más alto lo encontramos, una vez más, en la muestra del grupo de nativos con algo más de cuatro puntos de diferencia respecto a la muestra del grupo de NN.

Otro elemento a resaltar es la relación entre la media y la mediana en cada una de las muestras. Podemos observar que la diferencia es mínima en ambas muestras, ninguna llega al punto, pero la que presenta una distancia algo mayor es la muestra del grupo de NN.

En cuanto a la dispersión de las muestras, llama la atención la proximidad de los valores del estadístico «desviación típica». Como podemos ver en la tabla de estadísticos, el valor más bajo ( $s=4.2706$ ) corresponde a la muestra del grupo de N; mientras que la muestra del grupo de NN registra un  $s=5.2969$ , es decir, mayor dispersión de los datos. Estos datos los podemos visualizar en el gráfico siguiente.

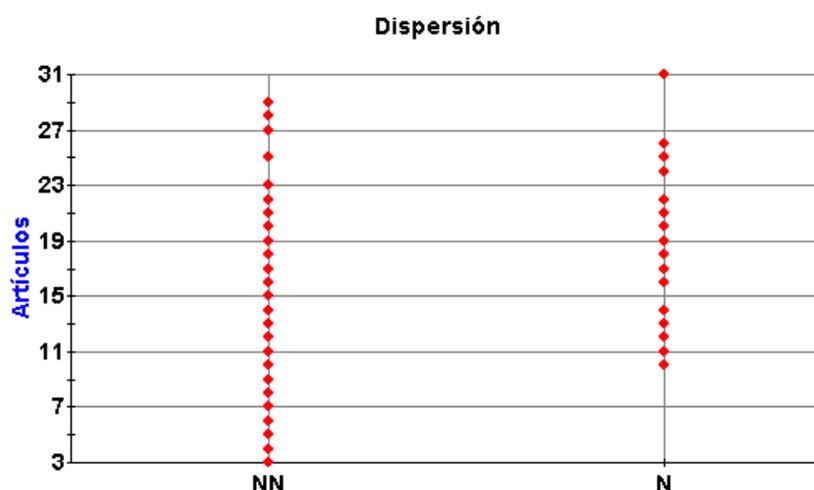


Gráfico 34. Gráfico de dispersión para la variable «Artículos» NN - N

También podemos visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras.

A diferencia de las dos últimas variables, adverbios y pronombres, en las que el valor del rango intercuartílico era el mismo en ambas muestras, aquí vemos que hay valores diferentes. El valor de la muestra del grupo de NN ( $IQR=6.0000$ ) supera en dos puntos al de la muestra del grupo de N ( $4.0000$ ). Eso significa que el 50% de los valores insertos en las cajas tiene una mayor dispersión en el grupo de no nativos. En el diagrama de cajas (Gráfico 35) vemos gráficamente esta información.

Podemos comprobar en el diagrama de cajas el mayor tamaño de la caja de la muestra del grupo de NN, como hemos indicado anteriormente. También llama la atención, al observar el gráfico, la proximidad de los valores del estadístico «máximo». En cambio, vemos que la diferencia en el estadístico «mínimo» es bastante marcada entre las dos

muestras. Asimismo, comprobamos lo señalado más arriba sobre la distancia entre las medidas de centralización de cada una de las muestras.

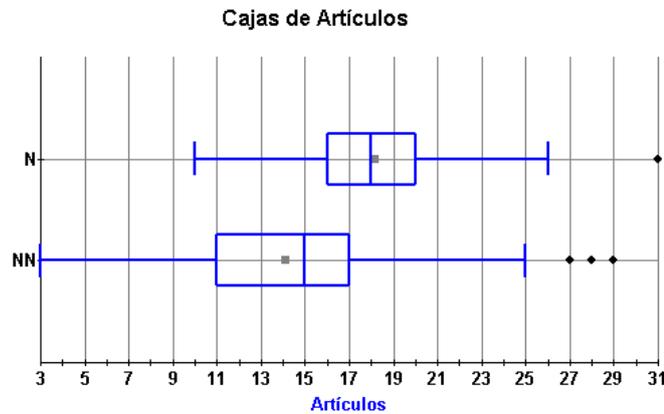


Gráfico 35. Diagrama de Cajas para la variable «Artículos» NN - N

En cuanto a la forma de la distribución de las muestras, vemos en la tabla de estadísticos (Tabla 28) que la curtosis es positiva en las dos muestras y sus valores están muy próximos entre una y otra. El valor de la muestra del grupo de NN (curtosis=0.1070) y la del grupo de N (curtosis= -0.3607). Como se ve, en los dos casos el valor está próximo a una curtosis cero (distribución normal o campana de Gauss). No obstante, el valor positivo del descriptor y al no coincidir con una curtosis cero, la curva tiene tendencia a una curva más aguda en comparación con una distribución normal. En los histogramas siguientes podemos visualizar esa tendencia.

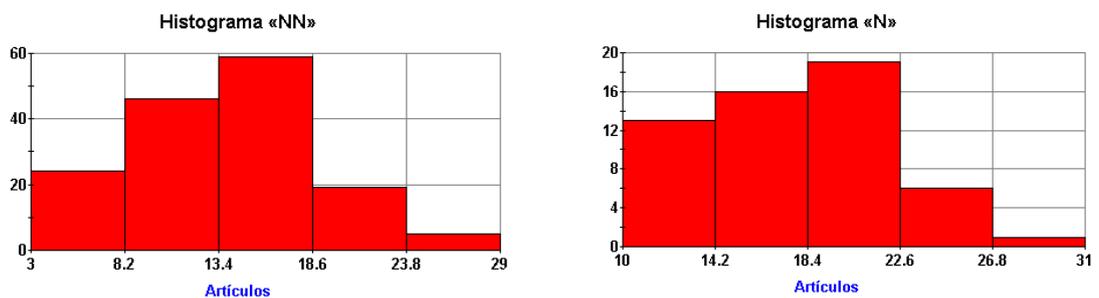


Gráfico 36. Histograma para la variable «Artículos» NN – N

(10) Conjunciones: Número total de conjunciones.

Muestras «F»		
	NN-F (n=153)	N-F (n=55)
Conjunciones	1.085	665

En la tabla siguiente encontramos los estadísticos que informan sobre las medidas de centralización, dispersión y forma de la variable «conjunciones» de las dos muestras.

**Estadísticos para la variable Conjunciones**

Grupos	NN	N
<b>N</b>	<b>153</b>	<b>55</b>
<b>Media</b>	<b>21.6928</b>	<b>27.3091</b>
<b>Mediana</b>	<b>21.0000</b>	<b>27.0000</b>
<b>Varianza</b>	<b>65.4774</b>	<b>43.5138</b>
<b>Desviación Típica</b>	<b>8.0918</b>	<b>6.5965</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>14.0000</b>
<b>Máximo</b>	<b>68.0000</b>	<b>52.0000</b>
<b>Cuartil Inferior</b>	<b>16.0000</b>	<b>23.0000</b>
<b>Cuartil Superior</b>	<b>26.0000</b>	<b>32.0000</b>
<b>Rango Intercuartílico</b>	<b>10.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>6.4785</b>	<b>2.4058</b>

Tabla 29. Tabla de estadísticos para la variable «Conjunciones» NN - N

Cuando observamos los estadísticos de centralización, la media y la mediana, vemos los valores tan diferentes que presenta una y otra muestra. Entre el valor más alto de la media, correspondiente a la muestra del grupo de nativos, y el valor del grupo de no nativos hay una diferencia de más de cinco puntos. Sin embargo, vemos una vez más la proximidad entre los valores de media y mediana dentro de cada muestra.

Respecto a los estadísticos que hacen referencia a la dispersión, desviación típica y rango intercuartílico, comprobamos que hay pequeñas diferencias entre ambas muestras. En cuanto a la desviación típica, observamos que el valor más alto ( $s=8.0918$ ) corresponde a la muestra del grupo de NN, con dos puntos más que el valor de la muestra del grupo de N ( $s=6.5965$ ). Podemos visualizar estos datos en el gráfico de dispersión n°. 37.

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta diferencia entre muestras, solo un punto. El valor más alto corresponde a la muestra del grupo de NN,  $IQR=10.0000$ ; mientras que la muestra del grupo de N registra un valor

IQR=9.0000. Esto significa que el tamaño de la caja de la muestra del grupo de no nativos es algo más grande como podemos comprobar en el gráfico de cajas n°. 38.

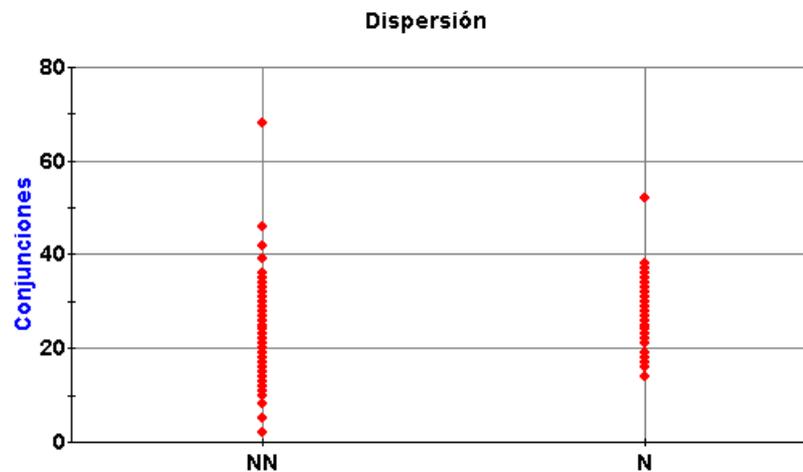


Gráfico 37. Gráfico de dispersión para la variable «Conjunciones» NN - N

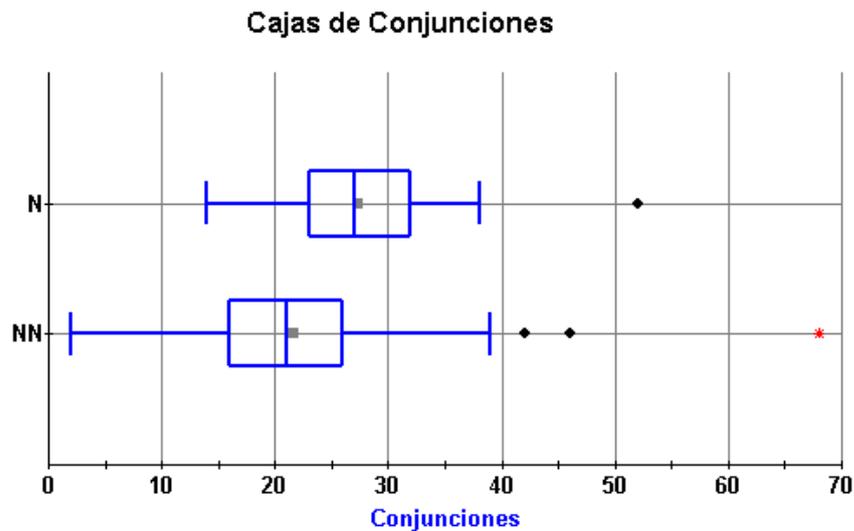


Gráfico 38. Diagrama de Cajas para la variable «Conjunciones» NN - N

También llama la atención, al observar el gráfico, la proximidad de los valores del estadístico «máximo», como en la variable anterior, aunque en este caso, el valor mayor corresponde a la muestra del grupo de NN. También vemos que la diferencia en el estadístico «mínimo» es bastante marcada entre las dos muestras, al igual que en la variable anterior. Asimismo, comprobamos lo señalado más arriba sobre la proximidad de las medidas de media y mediana de cada una de las muestras.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 29) que la curtosis es positiva en ambas muestras, aunque con valores muy diferentes. Mientras que la muestra del grupo de NN presenta un valor de curtosis=6.4785, el valor de la muestra del grupo de N es curtosis=2.4058. Como se puede observar, el valor más cercano a una curtosis cero (distribución normal o campana de Gauss) corresponde a la muestra del grupo de nativos. No obstante, como ambos valores son positivos, la curva tiene tendencia a una curva más aguda en comparación con una distribución normal. En los histogramas siguientes podemos visualizar esa tendencia.

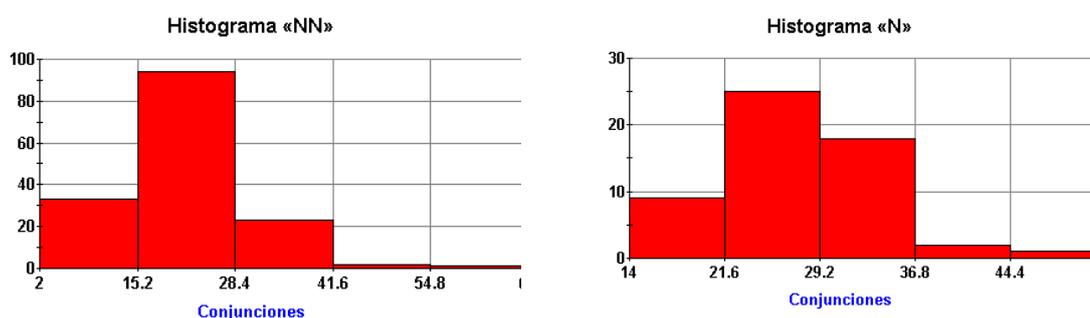


Gráfico 39. Histograma para la variable «Conjunciones» NN – N

### 6.1.2 Estadística descriptiva para la muestra «1D» por grupos

Vemos en la tabla de totales (Tabla 30) de la muestra codificada como «1D» que el número total de producciones en el grupo de NN y N es muy similar, no como en la muestra «F» que acabamos de ver en la que había una gran diferencia entre los dos grupos. Es por ello que aquí los totales de las distintas variables están más cercanos entre sí. Sin embargo, llama la atención que a pesar de la similitud en el número de producciones, los totales del grupo de nativos es superior al del grupo de no nativos.

Es por eso que vamos a ver a continuación qué dice la estadística sobre estos datos, qué semejanzas o diferencias hay en las medidas de centralización, dispersión y en la forma de la distribución de los datos.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
<i>Tokens</i>	6.668	9.746
<i>Types</i>	4.820	6.607
Verbos	1.0582	1.429
Sustantivos	2.853	3.993
Adjetivos	995	1.206
Adverbios	397	507
Pronombres	609	1.149
Preposiciones	353	619
Artículos	151	223
Conjunciones	258	620

Tabla 30. Tabla de totales de la muestra «NN-1D / N-1D»

(1) *Tokens*: Total de palabras.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
<i>Tokens</i>	6.668	9.746

Observamos que la diferencia en el número total de palabras entre la muestra del grupo de nativos y no nativos es de algo más de tres mil palabras a pesar de solo haber una diferencia de tres producciones entre ambas muestras.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable *Tokens* :**

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>102.5846</b>	<b>157.1935</b>
<b>Mediana</b>	<b>100.0000</b>	<b>154.5000</b>
<b>Varianza</b>	<b>803.4341</b>	<b>904.7816</b>
<b>Desviación Típica</b>	<b>28.3449</b>	<b>30.0796</b>
<b>Mínimo</b>	<b>52.0000</b>	<b>103.0000</b>
<b>Máximo</b>	<b>165.0000</b>	<b>230.0000</b>
<b>Cuartil Inferior</b>	<b>84.0000</b>	<b>130.0000</b>
<b>Cuartil Superior</b>	<b>122.0000</b>	<b>182.0000</b>
<b>Rango Intercuartílico</b>	<b>38.0000</b>	<b>52.0000</b>
<b>Curtosis</b>	<b>-0.5691</b>	<b>-0.6184</b>

Tabla 31. Estadísticos para la variable «*Tokens*» NN-1D / N-1D

Al observar los valores de centralización, media y mediana, vemos que las diferencias son llamativas. Por un lado, la media de *tokens* de la muestra del grupo de nativos supera en más de cincuenta y cuatro puntos a la del grupo de NN. Asimismo, el valor de la mediana también es muy superior en la muestra del grupo de N. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana en cada una de las muestras no es muy grande, pero no son coincidentes tampoco. Es decir, que estamos ante muestras que tienden a alejarse de una distribución normal.

En cuanto a los valores que indican el tipo de dispersión que presentan las muestras, desviación típica y rango intercuartílico, comprobamos con el primer valor que el más alto corresponde a la muestra del grupo de N ( $s=30.0796$ ), aunque vemos que la diferencia es pequeña, pues el valor de la muestra del grupo de NN es  $s=28.3449$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas.

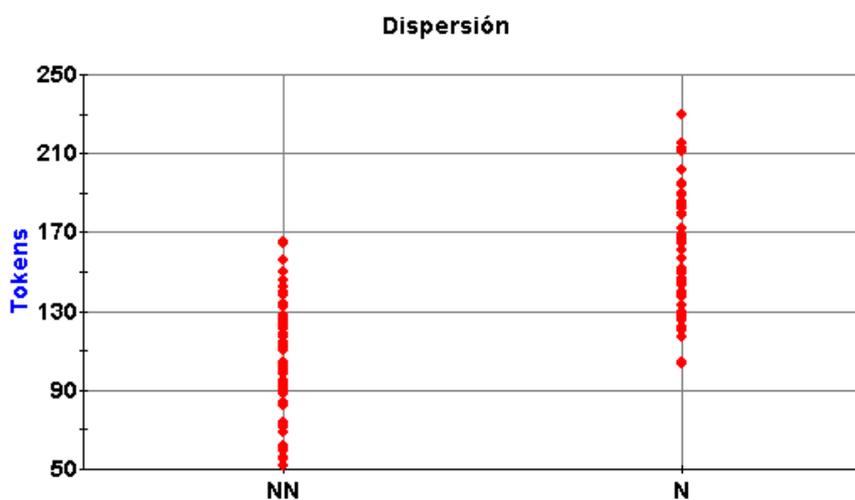


Gráfico 40. Gráfico de dispersión para la variable «Tokens» NN-1D / N-1D

El otro estadístico que hemos indicado que informa sobre la dispersión, el rango intercuartílico, tiene un valor también muy diferente entre una muestra y otra. Observamos que también es la muestra del grupo de N el que presenta un valor más alto ( $IQR=52.0000$ ), frente a un valor de  $IQR=38.0000$  de la muestra del grupo de NN. Así, podemos comprobar en el diagrama de cajas que mostramos a continuación esta diferencia de tamaño entre las dos muestras. Eso significa, recordemos, que el 50% de los valores que se encuentra dentro de la caja presenta una mayor dispersión en la muestra del grupo de nativos.

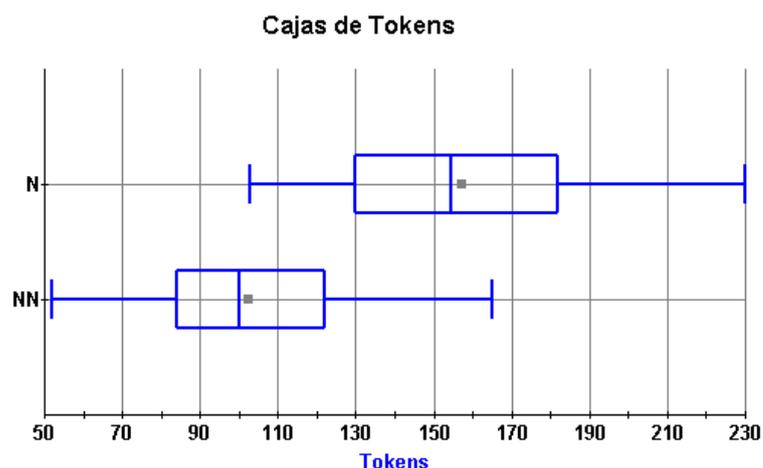


Gráfico 41. Diagrama de Cajas para la variable «Tokens» NN-1D / N-1D

La información que acabamos de dar la podemos visualizar en el diagrama de cajas, pues éste representa gráficamente el valor de los estadísticos. Así, podemos ver claramente la gran diferencia en las medidas de centralización, como hemos indicado más arriba. También es llamativo la diferencia que se da entre los estadísticos «mínimo» y «máximo» de ambas muestras. Otro elemento que hemos señalado anteriormente, la distancia entre media y mediana, vemos en el gráfico que son valores muy próximos, pero no coincidentes.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 31) que la curtosis es negativa en ambas muestras, y vemos, además, que son valores muy próximos entre ellos. El valor de la muestra del grupo de NN presenta un valor de curtosis= -0.5691 y el valor de la muestra del grupo de N es curtosis= -0.6184. Como se puede observar, el valor más cercano a una curtosis cero (distribución normal o campana de Gauss) corresponde a la muestra del grupo de nativos. No obstante, como ambos valores son negativos y muy cercanos entre sí, la curva que presenta cada una de las muestras, aunque tiene tendencia a una forma más plana en comparación con una distribución normal, vemos que no se aleja demasiado de una campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

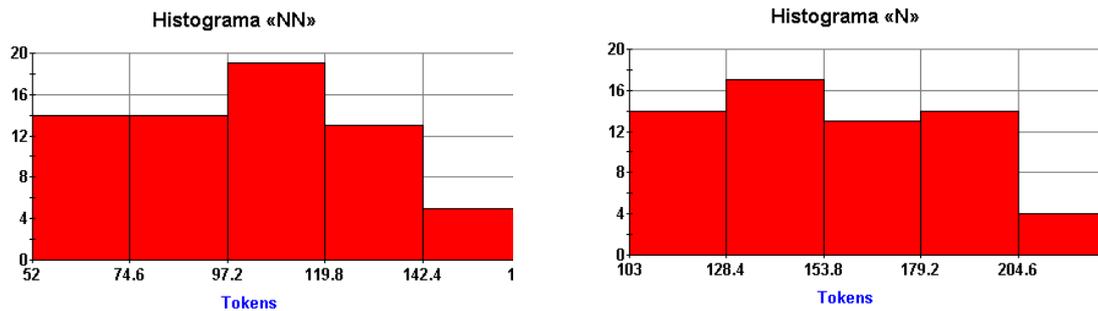


Gráfico 42. Histograma para la variable «Tokens» NN-1D / N-1D

(2) *Types*: Total de palabras diferentes.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
<i>Types</i>	4.820	6.607

En el número total de palabras distintas, vemos que la diferencia entre las dos muestras no llega a dos mil palabras. Sin embargo, sigue siendo una diferencia considerable.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable *Types***

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>74.1538</b>	<b>106.5645</b>
<b>Mediana</b>	<b>74.0000</b>	<b>103.5000</b>
<b>Varianza</b>	<b>320.1322</b>	<b>316.6105</b>
<b>Desviación Típica</b>	<b>17.8922</b>	<b>17.7936</b>
<b>Mínimo</b>	<b>42.0000</b>	<b>76.0000</b>
<b>Máximo</b>	<b>112.0000</b>	<b>148.0000</b>
<b>Cuartil Inferior</b>	<b>63.0000</b>	<b>93.0000</b>
<b>Cuartil Superior</b>	<b>87.0000</b>	<b>118.0000</b>
<b>Rango Intercuartílico</b>	<b>24.0000</b>	<b>25.0000</b>
<b>Curtosis</b>	<b>-0.5615</b>	<b>-0.5335</b>

Tabla 32. Estadísticos para la variable «*Types*» NN-1D / N-1D

Al observar los valores de centralización, media y mediana, vemos, al igual que en la variable anterior, que las diferencias son llamativas entre ambas muestras. Por un lado, la media de palabras distintas de la muestra del grupo de nativos supera en más de treinta puntos

a la del grupo de NN. Asimismo, el valor de la mediana también es muy superior en la muestra del grupo de N. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana en cada una de las muestras es mínima en el caso de la muestra del grupo de NN, mientras que en la muestra del grupo de N hay una pequeña diferencia de tres puntos. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 44).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que en esta variable, *types*, los valores de las dos muestras están muy próximos; para la muestra del grupo de NN  $s=17.8922$  y para la muestra del grupo de N es  $s=17.7936$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas.

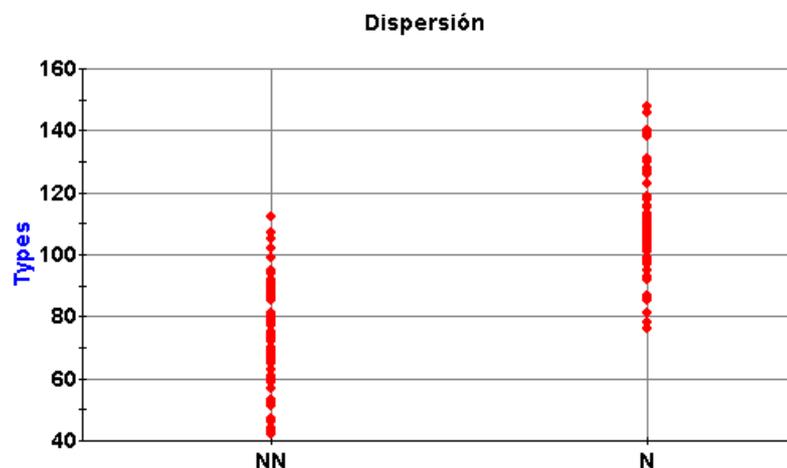


Gráfico 43. Gráfico de dispersión para la variable «Types» NN-1D / N-1D

También en el valor del rango intercuartílico, el otro estadístico que informa de la dispersión, vemos que la diferencia entre las dos muestras es pequeña, solo de un punto. Mientras que la muestra del grupo de NN presenta un  $IQR=24.0000$ , la del grupo de N registra un  $IQR=25.0000$ . Así, la dispersión de los valores dentro de la caja es prácticamente la misma. Recordemos que dentro de la caja se encuentra el 50% de todos los valores de las muestras. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

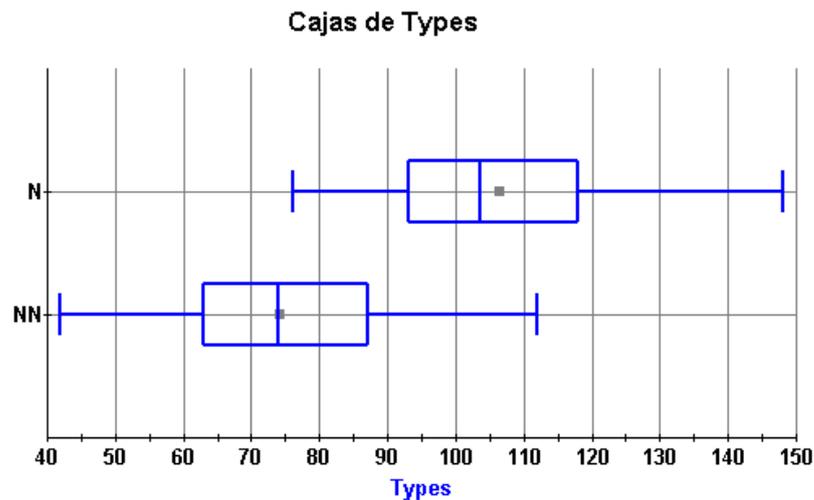


Gráfico 44. Diagrama de Cajas para la variable «Types» NN-1D / N-1D

Constatamos en el gráfico de cajas, en primer lugar, la diferencia que ya hemos señalado sobre los valores de los estadísticos «media» y «mediana». En el diagrama se ve claramente que la distancia existente entre estos descriptores en la muestra del grupo de N es algo mayor que en la muestra del grupo de NN. Esto nos informa que esa muestra se aleja algo más de lo que es una distribución normal. En segundo lugar, también se puede apreciar la diferencia existente entre los valores del estadístico «media» de ambas muestras.

Otro rasgo que llama la atención es la diferencia que se da entre los estadísticos «mínimo» y «máximo» de ambas muestras.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 32) que la curtosis vuelve a ser negativa en ambas muestras, y vemos, además, que son valores prácticamente iguales. El valor de la muestra del grupo de NN presenta un valor de curtosis= -0.5615 y el valor de la muestra del grupo de N es curtosis= -0.5335. Como ambos valores son negativos y muy cercanos entre sí, la curva que presenta cada una de las muestras, aunque tiene tendencia a una forma más plana en comparación con una distribución normal, vemos que no se aleja demasiado de una campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

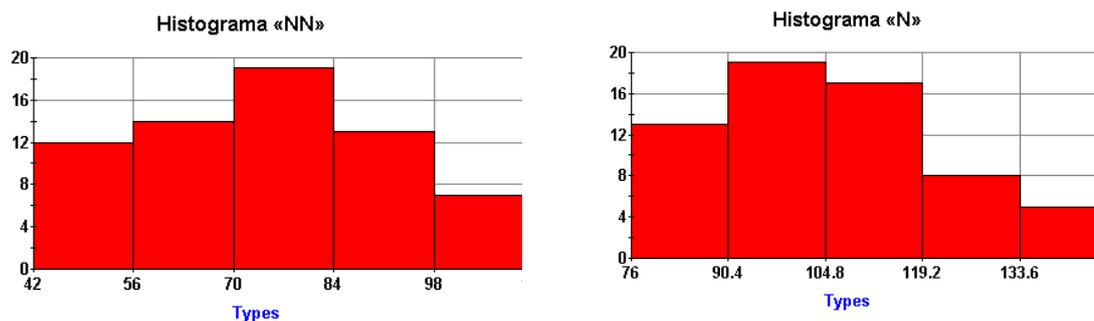


Gráfico 45. Histograma para la variable «Types» NN-1D / N-1D

(3) Verbos: Total de verbos.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
Verbos	1.052	1.429

Vemos en la tabla de totales que el número más alto corresponde a la muestra del grupo de N.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Verbos**

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>22.7846</b>	<b>34.3710</b>
<b>Mediana</b>	<b>23.0000</b>	<b>33.0000</b>
<b>Varianza</b>	<b>76.7966</b>	<b>71.9093</b>
<b>Desviación Típica</b>	<b>8.7634</b>	<b>8.4799</b>
<b>Mínimo</b>	<b>8.0000</b>	<b>17.0000</b>
<b>Máximo</b>	<b>44.0000</b>	<b>51.0000</b>
<b>Cuartil Inferior</b>	<b>15.0000</b>	<b>28.0000</b>
<b>Cuartil Superior</b>	<b>28.0000</b>	<b>40.0000</b>
<b>Rango Intercuartílico</b>	<b>13.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.3414</b>	<b>-0.7659</b>

Tabla 33. Estadísticos para la variable «Verbos» NN-1D / N-1D

Si observamos los valores de centralización, media y mediana, encontramos diferencias de más de once puntos entre medias y de diez entre medianas. Son valores que se tienen que tener en consideración. Asimismo, vemos que la diferencia entre los valores de la media y la mediana entre sí, en cada una de las muestras, es pequeña, pero observamos que la diferencia en la muestra del grupo de N es algo mayor. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 47).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que en esta variable, verbos, los valores de las dos muestras están muy próximos; para la muestra del grupo de NN  $s=8.7634$  y para la muestra del grupo de N es  $s=8.4799$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas.

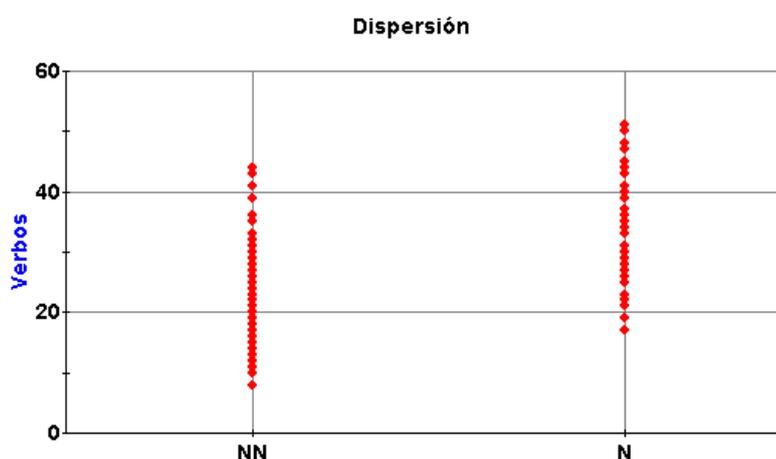


Gráfico 46. Gráfico de dispersión para la variable «Verbos» NN-1D / N-1D

También en el valor del rango intercuartílico, el otro estadístico que informa de la dispersión, vemos que la diferencia entre las dos muestras es pequeña, solo de un punto. Mientras que la muestra del grupo de NN presenta un  $IQR=13.0000$ , la del grupo de N registra un  $IQR=12.0000$ . Así, la dispersión del 50% de los valores dentro de la caja es prácticamente la misma en las dos muestras. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

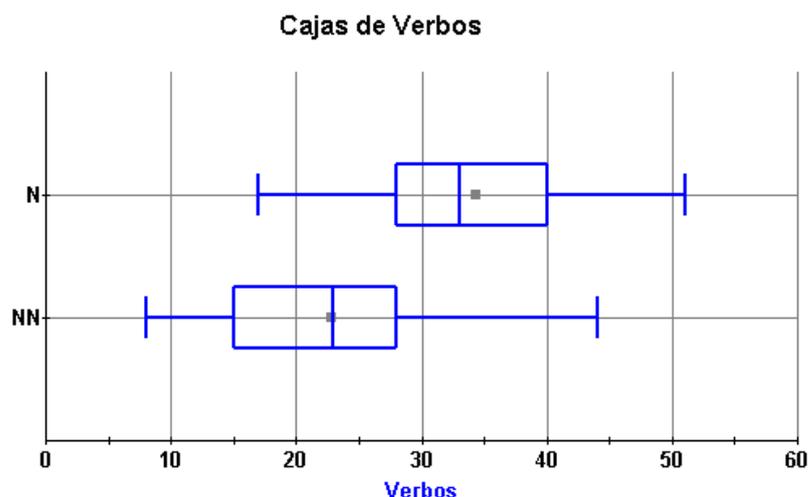


Gráfico 47. Diagrama de Cajas para la variable «Verbos» NN-1D / N-1D

Al visualizar los estadísticos en el diagrama de cajas, constatamos la diferencia que ya hemos señalado sobre los valores de los estadísticos «media» y «mediana». También comprobamos que la distancia entre la media y la mediana en la muestra del grupo de NN es prácticamente nula, mientras que en la muestra del grupo de N hay una pequeña diferencia. Esto nos informa que esta muestra se aleja algo más de lo que es una distribución normal.

Otros estadísticos que hay que resaltar son «mínimo» y «máximo». Aun habiendo diferencias entre las dos muestras, vemos que no son tan marcadas como en las primeras variables analizadas.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 33) que la curtosis es negativa en ambas muestras, y vemos, además, que son valores muy próximos entre sí. El valor de la muestra del grupo de NN presenta un valor de curtosis= -0.3414 y el valor de la muestra del grupo de N es curtosis= -0.7659. Como ambos valores son negativos, la curva que presenta cada una de las muestras, aunque tiene tendencia a una forma más plana en comparación con una distribución normal, vemos que no se aleja demasiado de una campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

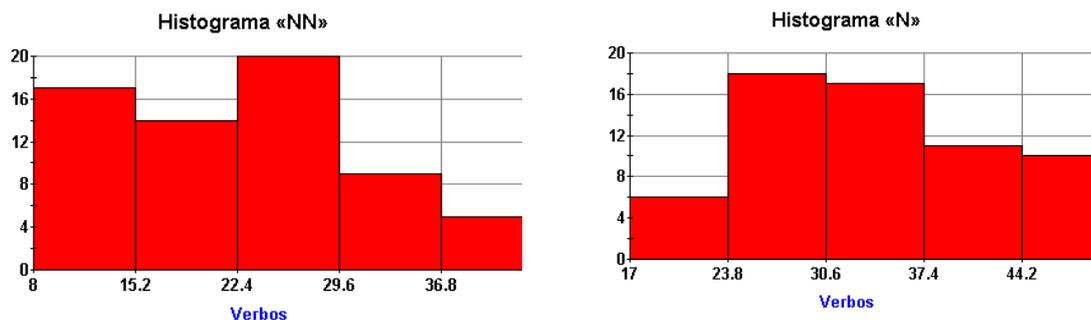


Gráfico 48. Histograma para la variable «Verbos» NN-1D / N-1D

(4) Sustantivos: Total de sustantivos.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
Sustantivos	2.853	3.993

Comprobamos que el número total de sustantivos más alto corresponde a la muestra del grupo de N con una diferencia importante respecto a la muestra del grupo de NN.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Sustantivos**

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>51.7538</b>	<b>76.6452</b>
<b>Mediana</b>	<b>50.0000</b>	<b>77.0000</b>
<b>Varianza</b>	<b>231.8135</b>	<b>300.8556</b>
<b>Desviación Típica</b>	<b>15.2254</b>	<b>17.3452</b>
<b>Mínimo</b>	<b>26.0000</b>	<b>44.0000</b>
<b>Máximo</b>	<b>82.0000</b>	<b>116.0000</b>
<b>Cuartil Inferior</b>	<b>40.0000</b>	<b>65.0000</b>
<b>Cuartil Superior</b>	<b>65.0000</b>	<b>86.0000</b>
<b>Rango Intercuartílico</b>	<b>25.0000</b>	<b>21.0000</b>
<b>Curtosis</b>	<b>-1.0085</b>	<b>-0.6025</b>

Tabla 34. Estadísticos para la variable «Sustantivos» NN-1D / N-1D

Al observar los valores de centralización, media y mediana, vemos que las diferencias entre ambas muestras son importantes. Por un lado, la media de sustantivos de la muestra

del grupo de N supera en casi veinticinco puntos a la del grupo de NN. Asimismo, el valor de la mediana también es muy superior en la muestra del grupo de N, que registra veintisiete puntos más que la del grupo de NN. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana dentro de cada una de las muestras es algo mayor en la muestra del grupo de NN. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 50).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que en esta variable es la muestra del grupo de N la que presenta un valor más alto,  $s=17.3452$ . La muestra del grupo de NN registra un valor  $s=15.2254$ . Así, vemos que, aunque con poco diferencia, el grupo de N presenta mayor dispersión de los datos. En el siguiente gráfico se puede visualizar la diferencia que hay entre ambas muestras.

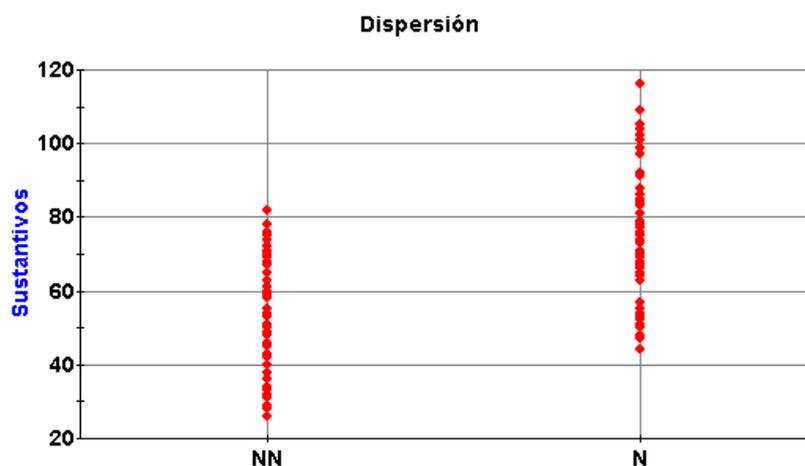


Gráfico 49. Gráfico de dispersión para la variable «Sustantivos» NN-1D / N-1D

El otro estadístico que informa de la dispersión, el rango intercuartílico, presenta el valor más alto en la muestra del grupo de NN ( $IQR=25.0000$ ); mientras que la muestra del grupo de N registra un valor de  $IQR=21.0000$ . Así, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de NN que en la de N, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

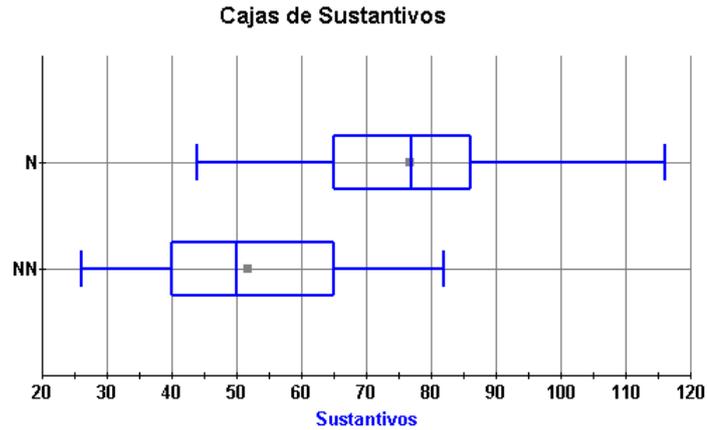


Gráfico 50. Diagrama de Cajas para la variable «Sustantivos» NN-1D / N-1D

Al visualizar los estadísticos en el diagrama de cajas, constatamos la diferencia que ya hemos señalado sobre los valores de los estadísticos «media» y «mediana». También comprobamos que la distancia entre la media y la mediana en la muestra del grupo de N es prácticamente nula, mientras que en la muestra del grupo de NN hay una pequeña diferencia. Esto nos informa que esta muestra se aleja algo más de lo que es una distribución normal.

Otros estadísticos que hay que resaltar son «mínimo» y «máximo». Vemos que entre los descriptores «mínimo» hay una diferencia menor que entre los descriptores «máximo».

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 34) que la curtosis es negativa en ambas muestras. Eso significa que la curva tiende a una forma más plana que una curtosis cero. El valor de la muestra del grupo de NN es  $\text{curtosis} = -1.0085$  y el valor de la muestra del grupo de N es  $\text{curtosis} = -0.6025$ . Los valores nos indican que la forma de las curvas de cada una de las muestras no se aleja demasiado de una campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

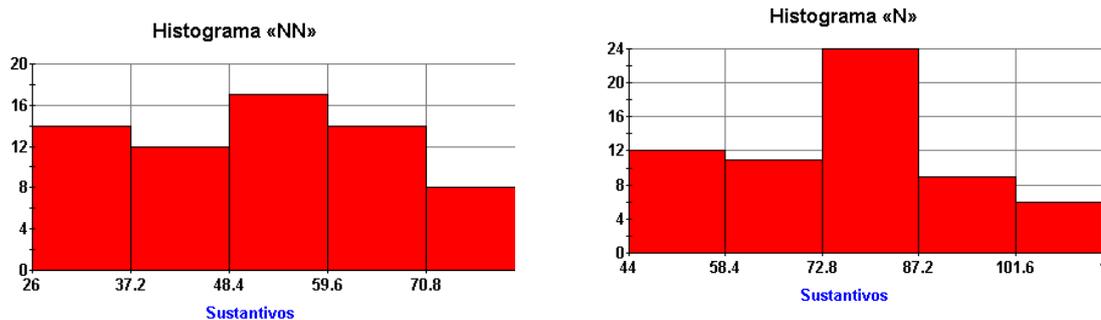


Gráfico 51. Histograma para la variable «Sustantivos» NN-1D / N-1D

(5) Adjetivos: Total de adjetivos.

	Muestras «ID»	
	NN-1D (n=65)	N-1D (n=62)
Adjetivos	995	1.206

Volvemos a constatar que el número más alto de adjetivos lo encontramos en la muestra del grupo de nativos.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Adjetivos**

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>23.0769</b>	<b>31.5968</b>
<b>Mediana</b>	<b>23.0000</b>	<b>30.5000</b>
<b>Varianza</b>	<b>61.4471</b>	<b>76.3429</b>
<b>Desviación Típica</b>	<b>7.8388</b>	<b>8.7374</b>
<b>Mínimo</b>	<b>7.0000</b>	<b>16.0000</b>
<b>Máximo</b>	<b>44.0000</b>	<b>57.0000</b>
<b>Cuartil Inferior</b>	<b>18.0000</b>	<b>26.0000</b>
<b>Cuartil Superior</b>	<b>30.0000</b>	<b>36.0000</b>
<b>Rango Intercuartílico</b>	<b>12.0000</b>	<b>10.0000</b>
<b>Curtosis</b>	<b>-0.3457</b>	<b>0.8133</b>

Tabla 35. Estadísticos para la variable «Adjetivos» NN-1D / N-1D

Al observar los valores de centralización, media y mediana, vemos que las diferencias son importantes entre ambas muestras. Por un lado, la media de la muestra del grupo de N es ocho puntos superior a la del grupo de NN, y la mediana siete puntos. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana en cada una de las muestras es mínima en el caso de la muestra del grupo de NN, mientras que en la muestra del grupo de N hay una pequeña diferencia de un punto. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 53).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que los valores de las dos muestras están muy próximos; para la muestra del grupo de NN  $s=7.8388$  y para la muestra del grupo de N

es  $s=8.7374$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas.

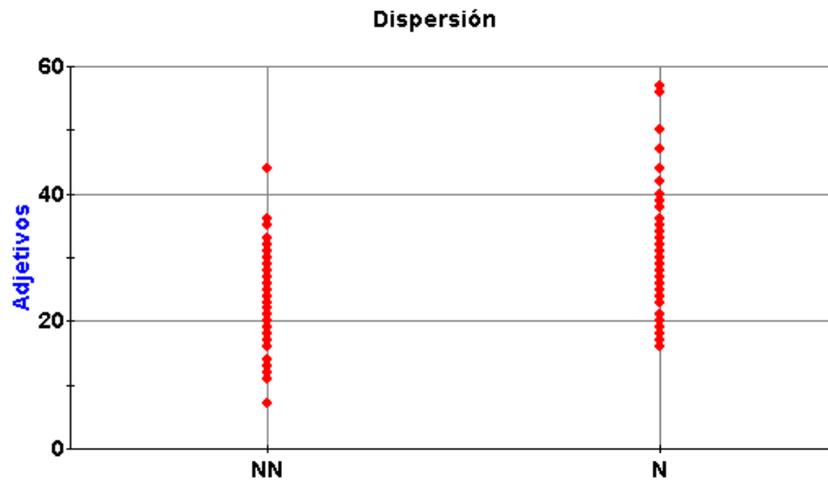


Gráfico 52. Gráfico de dispersión para la variable «Adjetivos» NN-1D / N-1D

También se puede visualizar en el gráfico de dispersión los valores extremos que presenta cada una de las muestras.

Como acabamos de ver, el valor del estadístico «desviación típica» es mayor en la muestra del grupo de N; sin embargo, el otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta un valor más bajo. Es decir, mientras que la muestra del grupo de NN registra un valor  $IQR=12.0000$ , la muestra del grupo de N presenta un  $IQR=10.0000$ . Eso significa que la caja de la muestra del grupo de NN es más grande y, por tanto, que los valores insertos en ella presenta una mayor dispersión que los de la muestra de nativos. En el diagrama de cajas que mostramos a continuación podemos ver gráficamente esta información y los valores del resto de estadísticos.

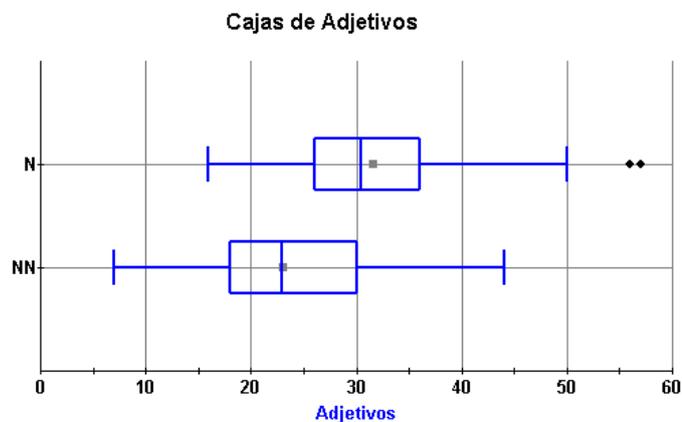


Gráfico 53. Diagrama de Cajas para la variable «Adjetivos» NN-1D / N-1D

También podemos comprobar en el gráfico de cajas la diferencia que existe entre las medidas de centralización, como ya hemos señalado más arriba. Asimismo, podemos visualizar la casi coincidencia entre la media y la mediana de la muestra del grupo de NN, mientras que en la muestra del grupo de N podemos apreciar una pequeña distancia. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

Otros estadísticos que hay que resaltar son «mínimo» y «máximo». Vemos que a diferencia de las variables anteriores analizadas, en ésta las diferencias entre muestras no son tan llamativas.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 35) que la curtosis es negativa en la muestra del grupo de NN (curtosis= -0.3457), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del grupo de N es positivo, curtosis=0.8133, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

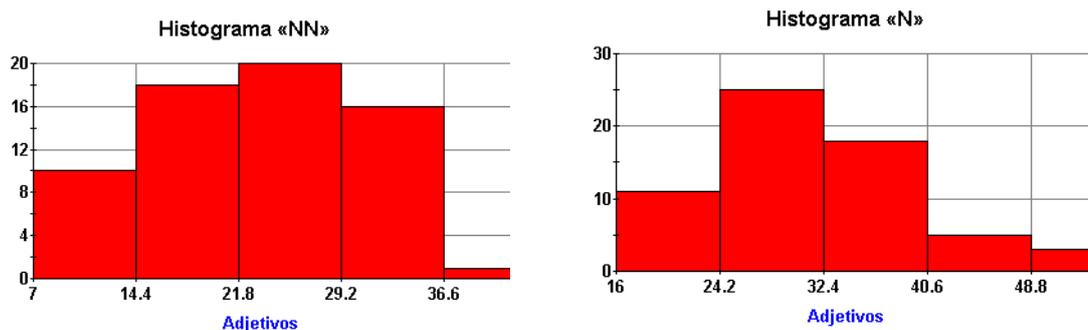


Gráfico 54. Histograma para la variable «Adjetivos» NN-1D / N-1D

(6) Adverbios: Total de adverbios.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
Adverbios	397	507

Volvemos a constatar que el número más alto de adverbios lo encontramos en la muestra del grupo de nativos.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Adverbios**

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>12.9538</b>	<b>20.3548</b>
<b>Mediana</b>	<b>12.0000</b>	<b>19.5000</b>
<b>Varianza</b>	<b>32.6385</b>	<b>51.1835</b>
<b>Desviación Típica</b>	<b>5.7130</b>	<b>7.1543</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>8.0000</b>
<b>Máximo</b>	<b>27.0000</b>	<b>39.0000</b>
<b>Cuartil Inferior</b>	<b>9.0000</b>	<b>15.0000</b>
<b>Cuartil Superior</b>	<b>17.0000</b>	<b>24.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>-0.2954</b>	<b>0.0054</b>

Tabla 36. Estadísticos para la variable «Adverbios» NN-1D / N-1D

Al observar los valores de centralización, media y mediana, volvemos a constatar que las diferencias son importantes entre ambas muestras. Por un lado, la media de la muestra del grupo de N es unos ocho puntos superior a la del grupo de NN, y la mediana algo más de siete puntos. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana en cada una de las muestras es mínima en ambos casos. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 56).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que los valores de las dos muestras están próximos, con una diferencia solo de algo más de un punto; para la muestra del grupo de NN  $s=5.7130$  y para la muestra del grupo de N es  $s=7.1543$ . En el siguiente gráfico se puede

visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas.

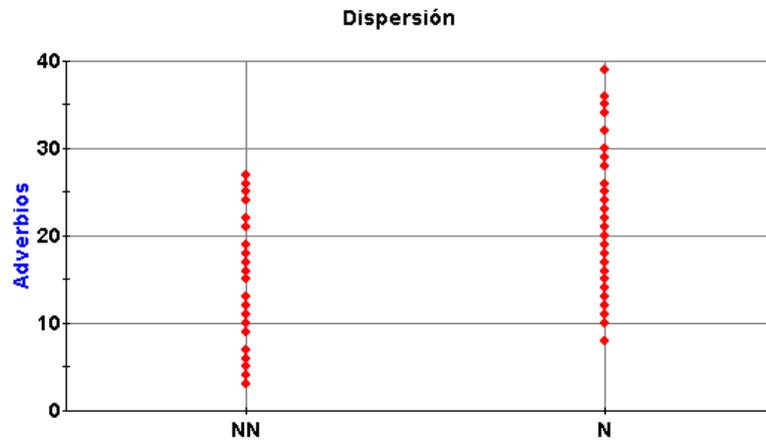


Gráfico 55. Gráfico de dispersión para la variable «Adverbios» NN-1D / N-1D

El otro estadístico que informa de la dispersión, el rango intercuartílico, presenta el valor más alto en la muestra del grupo de N (IQR=9.0000); mientras que la muestra del grupo de NN registra un valor de IQR=8.0000. Así, a diferencia de la variable anterior, vemos que los dos estadísticos que señalan la dispersión de los valores de las muestras coinciden en esta variable en presentar el valor más alto en la misma muestra. Por tanto, y aunque con poca diferencia, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de N que en la de NN, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

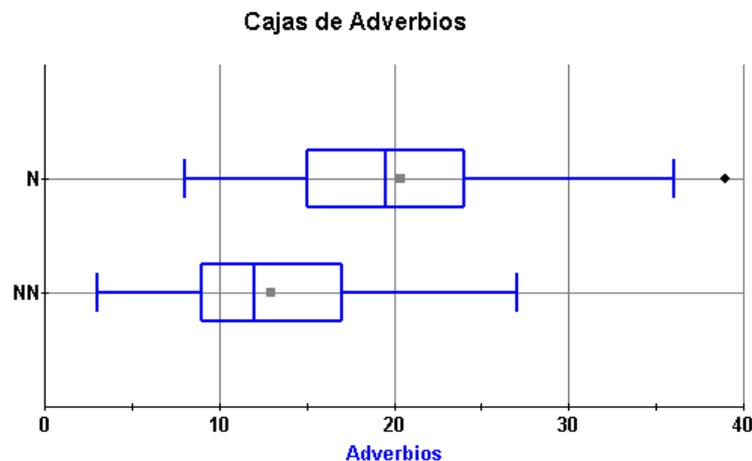


Gráfico 56. Diagrama de Cajas para la variable «Adverbios» NN-1D / N-1D

Lo primero que llama la atención al observar el diagrama de cajas es la distancia que se aprecia entre los descriptores de «media» y «mediana» en cada una de las muestras. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss. También podemos comprobar el mayor tamaño de la caja de la muestra del grupo de N.

En cuanto a los descriptores «mínimo» y «máximo», vemos que la diferencia entre muestras es mayor en el estadístico «máximo».

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 36) que la curtosis es negativa en la muestra del grupo de NN (curtosis= -0.2954), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del grupo de N es positivo, curtosis=0.0054, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

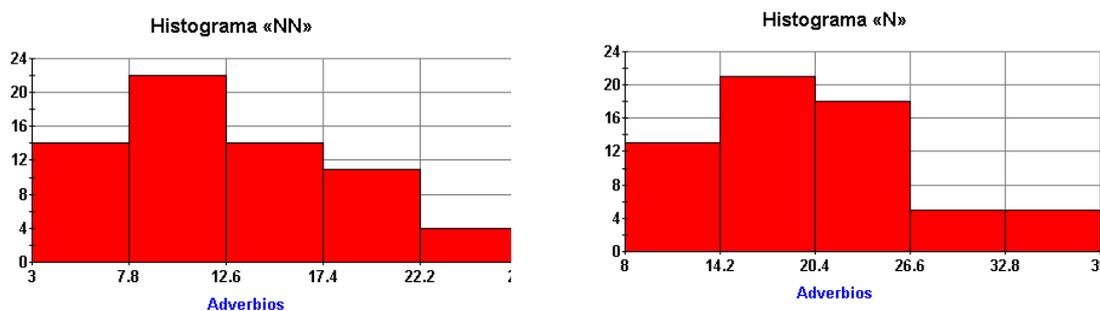


Gráfico 57. Histograma para la variable «Adverbios» NN-ID / N-ID

(7) Pronombres: Total de pronombres.

Muestras «ID»	
NN-ID (n=65)	N-ID (n=62)
Pronombres	609
	1.149

Comprobamos una vez más que el número más alto de pronombres lo encontramos en la muestra del grupo de nativos.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Pronombres**

---

Grupos	NN	N
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>19.0154</b>	<b>31.8710</b>
<b>Mediana</b>	<b>19.0000</b>	<b>30.0000</b>
<b>Varianza</b>	<b>33.5466</b>	<b>72.9011</b>
<b>Desviación Típica</b>	<b>5.7919</b>	<b>8.5382</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>19.0000</b>
<b>Máximo</b>	<b>32.0000</b>	<b>55.0000</b>
<b>Cuartil Inferior</b>	<b>15.0000</b>	<b>25.0000</b>
<b>Cuartil Superior</b>	<b>23.0000</b>	<b>37.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.5102</b>	<b>0.6445</b>

---

Tabla 37. Estadísticos para la variable «Pronombres» NN-1D / N-1D

Cuando observamos los valores de centralización, media y mediana, vemos que las diferencias son importantes también en esta variable. Por un lado, la media de la muestra del grupo de N es unos diez puntos superior a la del grupo de NN, y la mediana once puntos. Por otro lado, vemos que la diferencia entre los valores de la media y la mediana en cada una de las muestras es mínima en el caso de la muestra del grupo de NN, mientras que en la muestra del grupo de N hay una diferencia de casi dos puntos. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 59).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta la muestra del grupo de N,  $s=8.5382$ . Mientras que la muestra del grupo de NN registra un valor  $s=5.7919$ . En el gráfico 58 se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas, así como los valores extremos que presenta la muestra del grupo de NN.

El valor del rango intercuartílico, el otro estadístico que informa de la dispersión, presenta el valor más alto en la muestra del grupo de N ( $IQR=12.0000$ ); mientras que la muestra del grupo de NN registra un valor de  $IQR=8.0000$ . La diferencia es bastante

llamativa. Por tanto, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de N que en la de NN, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

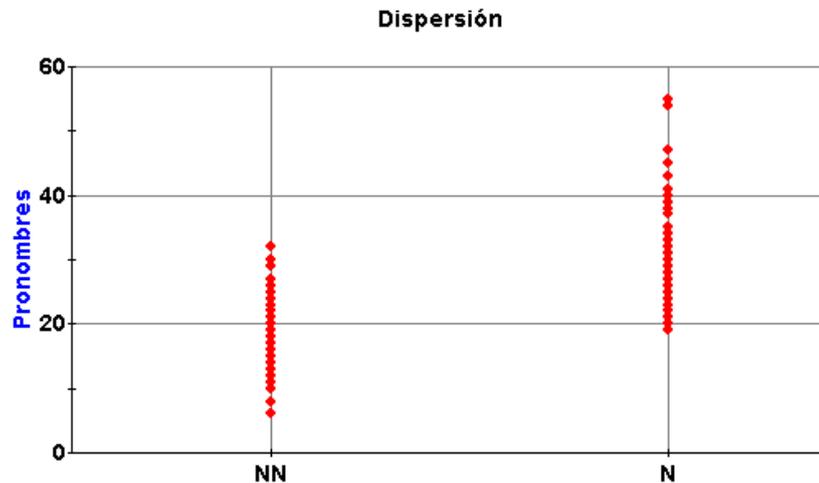


Gráfico 58. Gráfico de dispersión para la variable «Pronombres» NN-1D / N-1D

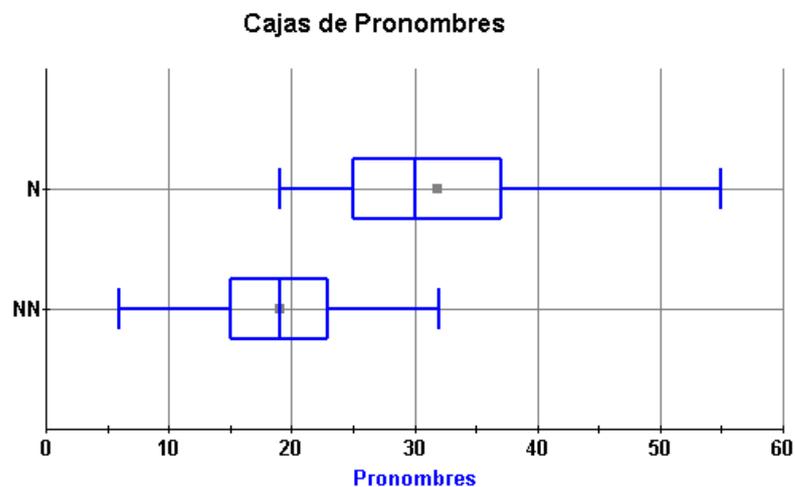


Gráfico 59. Diagrama de Cajas para la variable «Pronombres» NN-1D / N-1D

Lo primero que llama la atención al observar el diagrama de cajas es la gran diferencia en el tamaño de las cajas. Como ya hemos indicado, el mayor tamaño de la caja de la muestra del grupo N hace que los valores incluidos en ella (recordemos que es el 50% del total) presenten una mayor dispersión que los de la muestra del grupo de NN. El otro rasgo llamativo es la práctica coincidencia de media y mediana en la muestra del grupo de NN y

la distancia que hay entre esos mismos descriptores en la muestra del grupo de N. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a los descriptores «mínimo» y «máximo», llama la atención cómo el estadístico «máximo» de la muestra del grupo de N se aleja de la muestra del grupo de NN.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 37) que la curtosis es negativa en la muestra del grupo de NN (curtosis= -0.5102), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del grupo de N es positivo, curtosis=0.6445, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

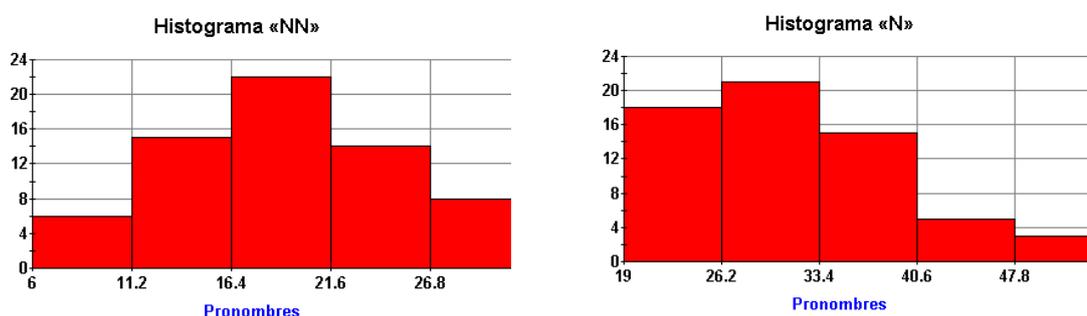


Gráfico 60. Histograma para la variable «Pronombres» NN-1D / N-1D

(8) Preposiciones: Total de preposiciones.

Muestras «1D»		
	NN-1D (n=65)	N-1D (n=62)
Preposiciones	353	619

Comprobamos una vez más que el número más alto de preposiciones lo encontramos en la muestra del grupo de nativos.

En la tabla 38 mostramos los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

Cuando observamos los valores de centralización, media y mediana, vemos que las diferencias son importantes también en esta variable. Por un lado, la media de la muestra del grupo de N es unos nueve puntos superiores a la del grupo de NN, y la mediana más de once puntos. Por otro lado, vemos que la muestra del grupo de N presenta una diferencia mayor entre los descriptores de «media» y «mediana». Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 62).

**Estadísticos para la variable Preposiciones**

---

<b>Grupos</b>	<b>NN</b>	<b>N</b>
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>14.0154</b>	<b>23.2581</b>
<b>Mediana</b>	<b>13.0000</b>	<b>24.5000</b>
<b>Varianza</b>	<b>26.9841</b>	<b>42.2274</b>
<b>Desviación Típica</b>	<b>5.1946</b>	<b>6.4983</b>
<b>Mínimo</b>	<b>4.0000</b>	<b>9.0000</b>
<b>Máximo</b>	<b>31.0000</b>	<b>38.0000</b>
<b>Cuartil Inferior</b>	<b>11.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>17.0000</b>	<b>27.0000</b>
<b>Rango Intercuartílico</b>	<b>6.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>0.5665</b>	<b>-0.4241</b>

Tabla 38. Estadísticos para la variable «Preposiciones» NN-1D / N-1D

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta la muestra del grupo de N,  $s=6.4983$ . Mientras que la muestra del grupo de NN registra un valor  $s=5.1946$ . Vemos que la diferencia entre ambas muestras no es demasiado grande. En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas, así como los valores extremos que presentan las dos muestras.

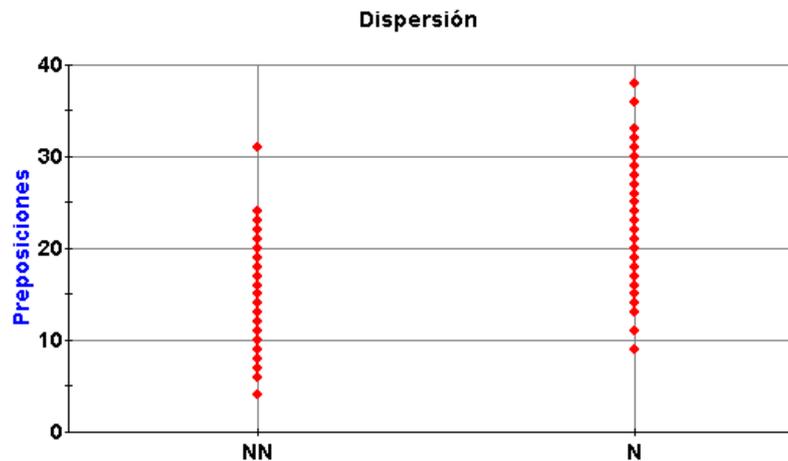


Gráfico 61. Gráfico de dispersión para la variable « Preposiciones» NN-1D / N-1D

El otro estadístico que informa de la dispersión, el rango intercuartílico, presenta el valor más alto en la muestra del grupo de N (IQR=9.0000); mientras que la muestra del grupo de NN registra un valor de IQR=6.0000. Por tanto, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de N que en la de NN, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

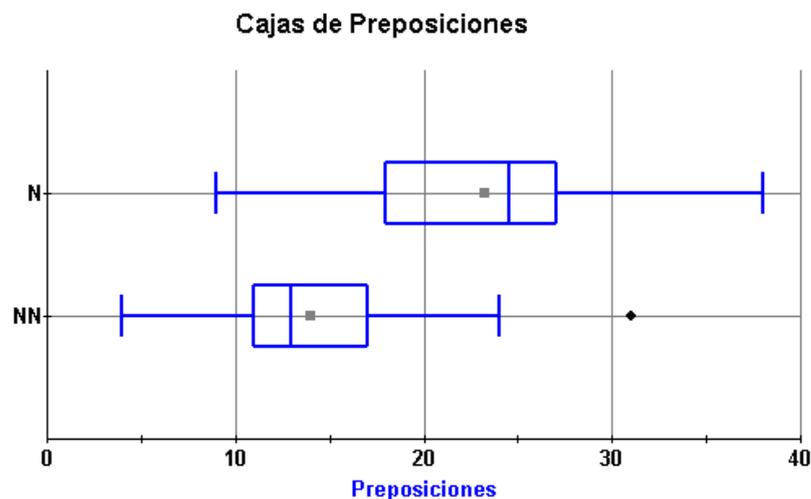


Gráfico 62. Diagrama de Cajas para la variable « Preposiciones» NN-1D / N-1D

Lo primero que llama la atención al observar el diagrama de cajas es, por un lado, la gran diferencia en el tamaño de las cajas. Como ya hemos indicado, el mayor tamaño de la caja de la muestra del grupo N hace que los valores incluidos en ella presenten una mayor dispersión que los de la muestra del grupo de NN. Por otro lado, comprobamos la distancia

que hay entre media y mediana en cada una de las muestras. Esa pequeña diferencia indica que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a los descriptores «mínimo» y «máximo», llama la atención cómo el estadístico «máximo» de la muestra del grupo de N se aleja de la muestra del grupo de NN.

En cuanto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 38) que la curtosis es negativa en la muestra del grupo de N (curtosis= -0.4241), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del grupo de NN es positivo, curtosis=0.5665, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

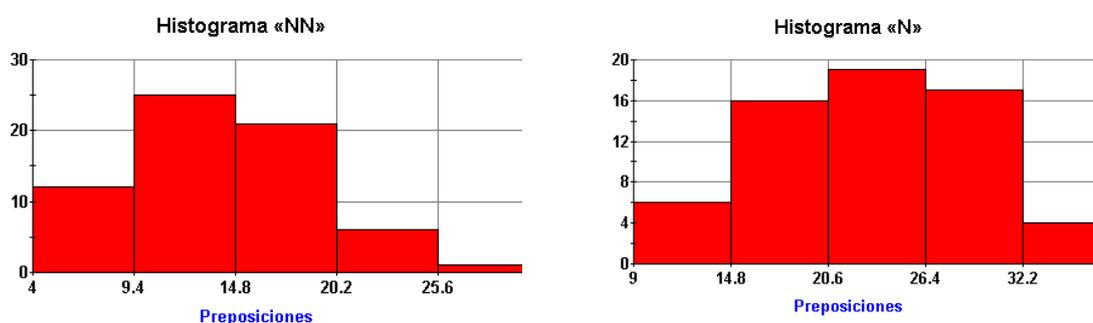


Gráfico 63. Histograma para la variable « Preposiciones» NN-1D / N-1D

(9) Artículos: Total de artículos.

Muestras «ID»		
	NN-1D (n=65)	N-1D (n=62)
Artículos	151	223

Comprobamos una vez más que el número más alto de artículos lo encontramos en la muestra del grupo de nativos.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Artículos**

<b>Grupos</b>	<b>NN</b>	<b>N</b>
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>9.3538</b>	<b>15.4032</b>
<b>Mediana</b>	<b>9.0000</b>	<b>14.0000</b>
<b>Varianza</b>	<b>14.5760</b>	<b>28.3757</b>
<b>Desviación Típica</b>	<b>3.8178</b>	<b>5.3269</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>6.0000</b>
<b>Máximo</b>	<b>18.0000</b>	<b>33.0000</b>
<b>Cuartil Inferior</b>	<b>6.0000</b>	<b>12.0000</b>
<b>Cuartil Superior</b>	<b>12.0000</b>	<b>19.0000</b>
<b>Rango Intercuartílico</b>	<b>6.0000</b>	<b>7.0000</b>
<b>Curtosis</b>	<b>-0.5977</b>	<b>0.8238</b>

Tabla 39. Estadísticos para la variable « Artículos» NN-1D / N-1D

Cuando observamos los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos en la muestra del grupo de N. Por un lado, la media de la muestra del grupo de N es unos seis puntos superior a la del grupo de NN, y la mediana más de cinco puntos. Por otro lado, vemos que la muestra del grupo de N presenta una diferencia mayor entre los descriptores de «media» y «mediana». Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 65).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del grupo de N,  $s=5.3269$ . Mientras que la muestra del grupo de NN registra un valor  $s=3.8178$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas, así como los valores extremos que presenta la muestra del grupo de N.

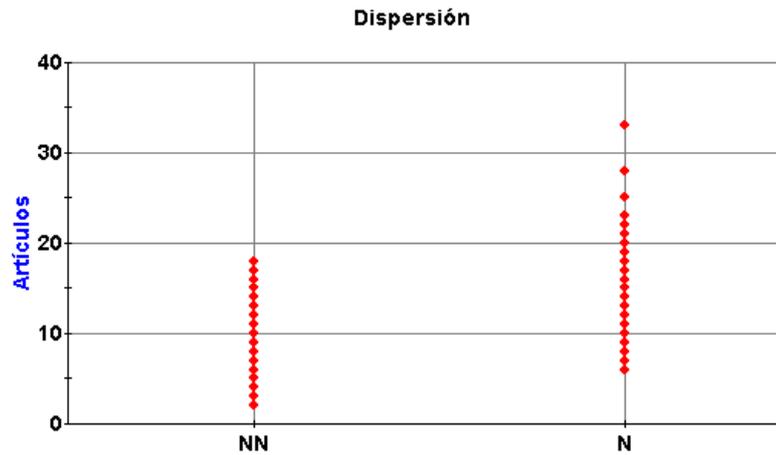


Gráfico 64. Gráfico de dispersión para la variable « Artículos» NN-1D / N-1D

Respecto al valor del rango intercuartílico, el otro estadístico que indica la dispersión de los datos, presenta el valor más alto en la muestra del grupo de N (IQR=7.0000); mientras que la muestra del grupo de NN registra un valor de IQR=6.0000. Por tanto, aunque sea poca la diferencia, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de N que en la de NN, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

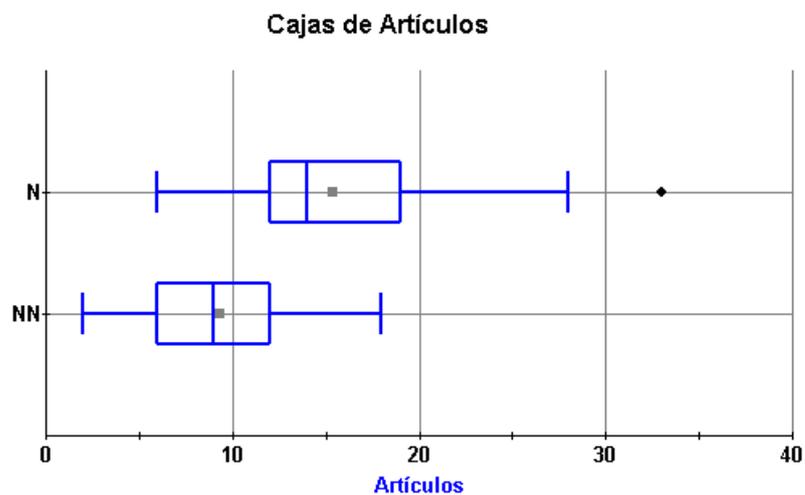


Gráfico 65. Diagrama de Cajas para la variable « Artículos» NN-1D / N-1D

Cuando observamos el diagrama de cajas, lo primero que llama la atención es, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, comprobamos la distancia que hay entre media y mediana en cada una de las muestras. Mientras que en la muestra del grupo de NN las dos medidas son casi coincidentes, en la

muestra del grupo de N se aprecia una distancia algo mayor. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a los descriptores «mínimo» y «máximo», igual que en las variables anteriores, llama la atención cómo el estadístico «máximo» de la muestra del grupo de N se aleja de la muestra del grupo de NN.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 39) que la curtosis es negativa en la muestra del grupo de NN (curtosis= -0.5977), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del grupo de N es positivo, curtosis=0.8238, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

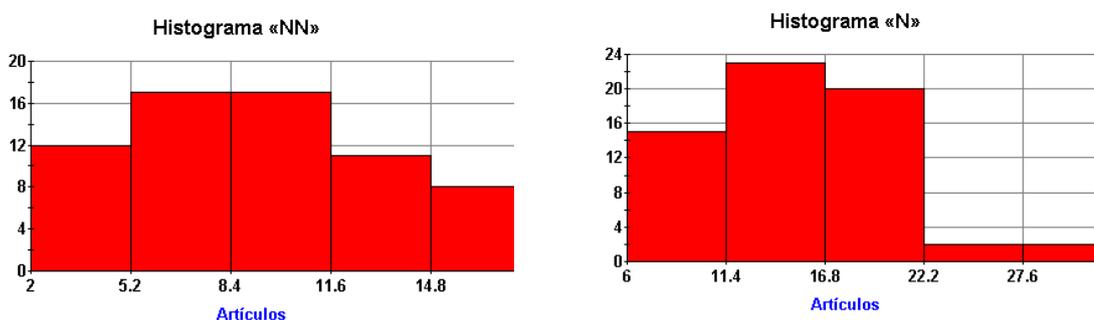


Gráfico 66. Histograma para la variable «Artículos» NN-1D / N-1D

(10) Conjunciones: Total de conjunciones.

Muestras «1D»		
	NN-1D (n=65)	N-1D (n=62)
Conjunciones	258	620

Comprobamos una vez más que el número más alto de conjunciones lo encontramos en la muestra del grupo de nativos.

A continuación presentamos la tabla de estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Conjunciones**

<b>Grupos</b>	<b>NN</b>	<b>N</b>
<b>N</b>	<b>65</b>	<b>62</b>
<b>Media</b>	<b>11.9538</b>	<b>23.0323</b>
<b>Mediana</b>	<b>11.0000</b>	<b>23.0000</b>
<b>Varianza</b>	<b>38.9822</b>	<b>44.1629</b>
<b>Desviación Típica</b>	<b>6.2436</b>	<b>6.6455</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>11.0000</b>
<b>Máximo</b>	<b>33.0000</b>	<b>46.0000</b>
<b>Cuartil Inferior</b>	<b>8.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>15.0000</b>	<b>27.0000</b>
<b>Rango Intercuartílico</b>	<b>7.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>1.5824</b>	<b>0.9065</b>

Tabla 40. Estadísticos para la variable «Conjunciones» NN-1D / N-1D

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos, una vez más, en la muestra del grupo de N. Por un lado, la media de la muestra del grupo de N es unos once puntos superior a la del grupo de NN, y la mediana más de doce puntos. Por otro lado, vemos que la muestra del grupo de N presenta una diferencia mínima entre los descriptores de «media» y «mediana», mientras que la muestra del grupo de NN presenta una distancia algo mayor entre esos mismos descriptores. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 68).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que las dos muestras presentan valores muy similares. La muestra del grupo de NN registra un valor de  $s=6.2436$ ; mientras que la muestra del grupo de N presenta un valor  $s=6.6455$ . En el siguiente gráfico se puede visualizar la dispersión de cada una de las muestras y la pequeña diferencia que se da entre ambas, así como los valores extremos que presenta la muestra del grupo de N.

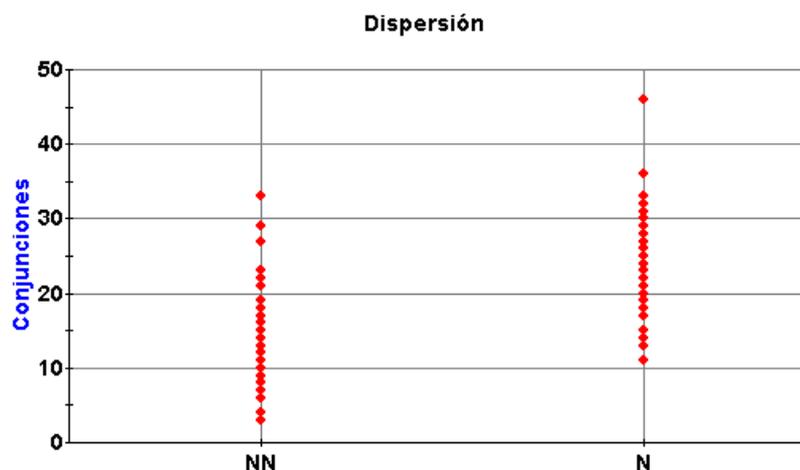


Gráfico 67. Gráfico de dispersión para la variable « Conjunctiones» NN-1D / N-1D

El otro estadístico que indica la dispersión de los datos, el rango intercuartílico, presenta el valor más alto en la muestra del grupo de N (IQR=9.0000); mientras que la muestra del grupo de NN registra un valor de IQR=7.0000. Por tanto, aunque sea poca la diferencia, vamos a encontrar una mayor dispersión en la caja de la muestra del grupo de N que en la de NN, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que presentamos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

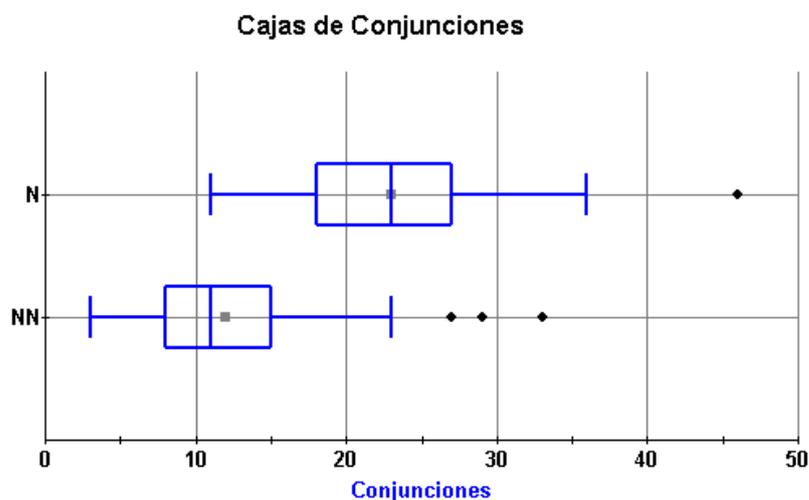


Gráfico 68. Diagrama de Cajas para la variable « Conjunctiones» NN-1D / N-1D

Al observar el diagrama de cajas, lo primero que llama la atención es, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, comprobamos la distancia que hay entre media y mediana en cada una de las muestras. Mientras que en la muestra del grupo de N las dos medidas son casi coincidentes, en la

muestra del grupo de NN se aprecia una distancia algo mayor. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a los descriptores «mínimo» y «máximo», igual que en las variables anteriores, llama la atención cómo el estadístico «máximo» de la muestra del grupo de N se aleja del «máximo» de la muestra del grupo de NN.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 40) que la curtosis es positiva en ambas muestras. Por tanto, la curva de las dos muestras tiende a una forma más aguda que una curtosis cero. No obstante, vemos que los valores son muy diferentes; mientras que la muestra del grupo de NN presenta una curtosis=1.5824, la del grupo de N (curtosis=0.9065) registra una curtosis más próxima al cero. Los valores nos indican que la forma de la curva de la muestra del grupo de N va a estar más cercana a la forma de una campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

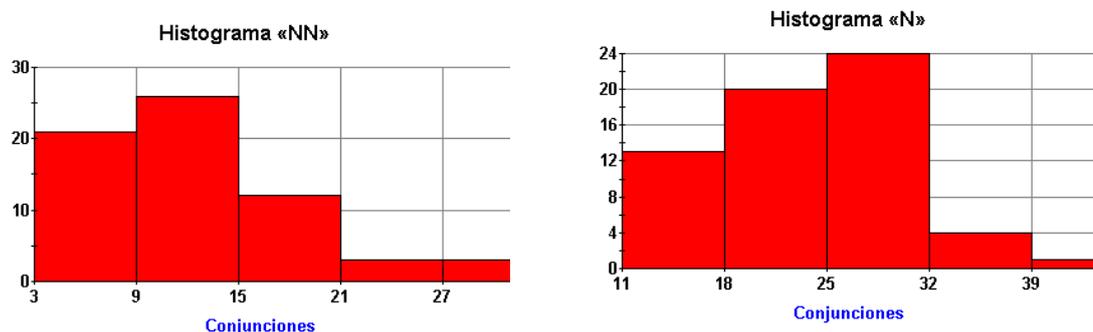


Gráfico 69. Histograma para la variable «Conjunciones» NN-1D / N-1D

### 6.1.3 Estadística descriptiva para el M1 y M2 del grupo de NN

Como podemos ver en la tabla siguiente, hay un aumento en los totales de todas las variables en el M-2 (final del curso) respecto del M-1 (inicio del curso). A continuación vamos a ver cómo se comporta cada variable, qué dicen los estadísticos sobre las medidas de centralización, dispersión y de forma de cada una de las muestras.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
<i>Tokens</i>	12.185	16.715
<i>Types</i>	8.405	11.164
Verbos	2.190	3.231
Sustantivos	5.198	7.042
Adjetivos	1.742	2.232
Adverbios	785	949
Pronombres	880	1.432
Preposiciones	579	755
Artículos	235	307
Conjunciones	576	769

Tabla 41. Totales de los «M-1» y «M-2» NN-M1 / NN-M2

(1) *Tokens*: Total de palabras.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
<i>Tokens</i>	12.185	16.715

Observamos que se produce un aumento del número total de palabras en el M-2 (final del curso) respecto al M-1 (inicio del curso).

A continuación presentamos la tabla de estadísticos para ver de qué manera esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable *Tokens* :**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>116.0476</b>	<b>147.9204</b>
<b>Mediana</b>	<b>119.0000</b>	<b>142.0000</b>
<b>Varianza</b>	<b>842.2381</b>	<b>1701.1632</b>
<b>Desviación Típica</b>	<b>29.0213</b>	<b>41.2452</b>
<b>Mínimo</b>	<b>52.0000</b>	<b>72.0000</b>
<b>Máximo</b>	<b>191.0000</b>	<b>283.0000</b>
<b>Cuartil Inferior</b>	<b>100.0000</b>	<b>122.0000</b>
<b>Cuartil Superior</b>	<b>133.0000</b>	<b>175.0000</b>
<b>Rango Intercuartílico</b>	<b>33.0000</b>	<b>53.0000</b>
<b>Curtosis</b>	<b>-0.1460</b>	<b>0.2229</b>

Tabla 42. Estadísticos para la variable «*Tokens*» NN-M1 / NN-M2

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos también en la muestra del M-2, como en los totales. Por un lado, la media de la muestra del M-2 supera en casi treinta y dos puntos la muestra del M-1, y la mediana en veintitrés puntos. Por otro lado, vemos que la distancia entre los descriptores de «media» y «mediana» en cada una de las muestras es mayor en la muestra del M-2. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 71).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=41.2452$ . Mientras que la muestra del M-1 registra un valor  $s=29.0213$ . Es decir, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el gráfico 70, así como los valores extremos.

En cuanto al valor del rango intercuartílico, el otro estadístico que indica la dispersión de los datos, comprobamos igualmente que es la muestra del M-2 la que presenta el valor más alto ( $IQR=53.0000$ ); mientras que la muestra del M-1 registra un valor de  $IQR=33.0000$ . Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas (Gráfico 71) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

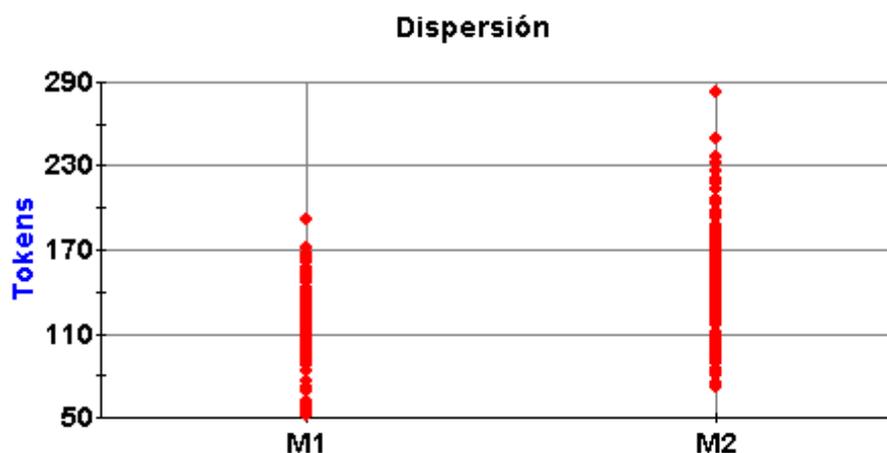


Gráfico 70. Gráfico de dispersión para la variable «Tokens» NN-M1 / NN-M2

Si observamos el diagrama de cajas, lo primero que llama la atención es, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, comprobamos la distancia que hay entre media y mediana en cada una de las muestras. Mientras que en la muestra del M-1 las dos medidas están muy próximas, en la muestra del

M-2 se aprecia una distancia algo mayor. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

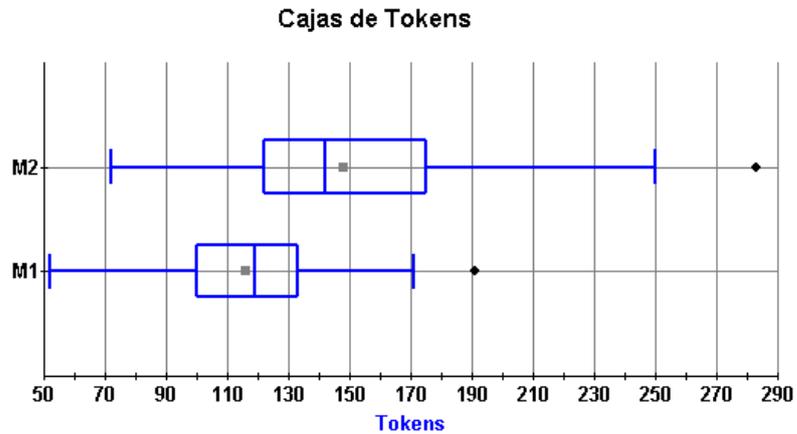


Gráfico 71. Diagrama de Cajas para la variable «Tokens» NN-M1 / NN-M2

En cuanto a los descriptores «mínimo» y «máximo», llama la atención la diferencia del estadístico «máximo» entre las dos muestras. Es llamativo el aumento que se produce en la muestra del M-2.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 42) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.1460), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.2229, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

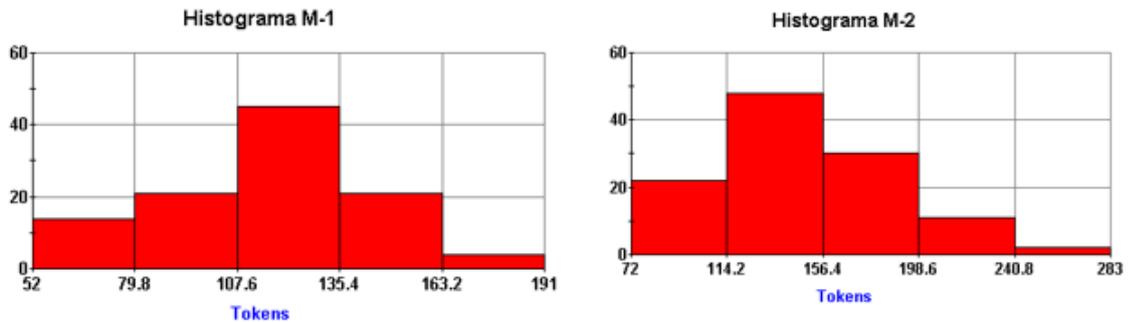


Gráfico 72. Histograma para la variable «Tokens» NN-M1 / NN-M2.

(2) *Types*: Total de palabras diferentes.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
<i>Types</i>	8.405	11.164

También en el número total de palabras diferentes se produce un aumento considerable en el M-2 respecto al M-1.

En la tabla 43 mostramos los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

También en la variable «*Types*» vemos que los valores de los estadísticos de centralización, media y mediana, son más altos en la muestra del M-2. Por un lado, la media de la muestra del M-2 supera en casi diecinueve puntos la muestra del M-1, y la mediana en diecisiete puntos. Por otro lado, vemos igualmente que la mayor distancia entre los descriptores de «media» y «mediana» la encontramos en la muestra del M-2. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 74).

**Estadísticos para la variable *Types***

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>80.0476</b>	<b>98.7965</b>
<b>Mediana</b>	<b>82.0000</b>	<b>99.0000</b>
<b>Varianza</b>	<b>289.8535</b>	<b>466.5564</b>
<b>Desviación Típica</b>	<b>17.0251</b>	<b>21.5999</b>
<b>Mínimo</b>	<b>42.0000</b>	<b>51.0000</b>
<b>Máximo</b>	<b>115.0000</b>	<b>153.0000</b>
<b>Cuartil Inferior</b>	<b>70.0000</b>	<b>83.0000</b>
<b>Cuartil Superior</b>	<b>92.0000</b>	<b>112.0000</b>
<b>Rango Intercuartílico</b>	<b>22.0000</b>	<b>29.0000</b>
<b>Curtosis</b>	<b>-0.3656</b>	<b>-0.3445</b>

Tabla 43. Estadísticos para la variable «*Types*» NN-M1 / NN-M2

En cuanto a los estadísticos que informan sobre la dispersión, la desviación típica y el rango intercuartílico, vemos que las diferencias no son tan marcadas como en la variable anterior, aun así, es la muestra del M-2 la que registra los valores más altos. Para la

desviación típica presenta un valor  $s=21.5999$ ; Mientras que la muestra del M-1 registra un valor  $s=17.0251$ . Es decir, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el siguiente gráfico, así como los valores extremos.

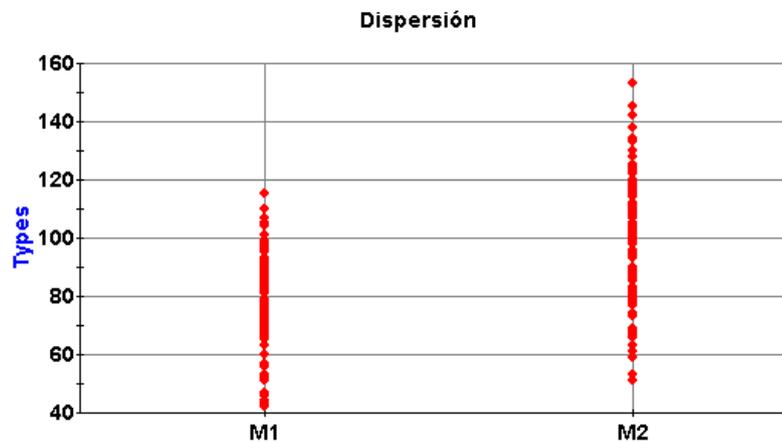


Gráfico 73. Gráfico de dispersión para la variable «Types» NN-M1 / NN-M2

También el valor del rango intercuartílico presenta un valor más alto en la muestra del M-2 (IQR=29.0000); mientras que la muestra del M-1 registra un valor de IQR=22.0000. Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que mostramos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

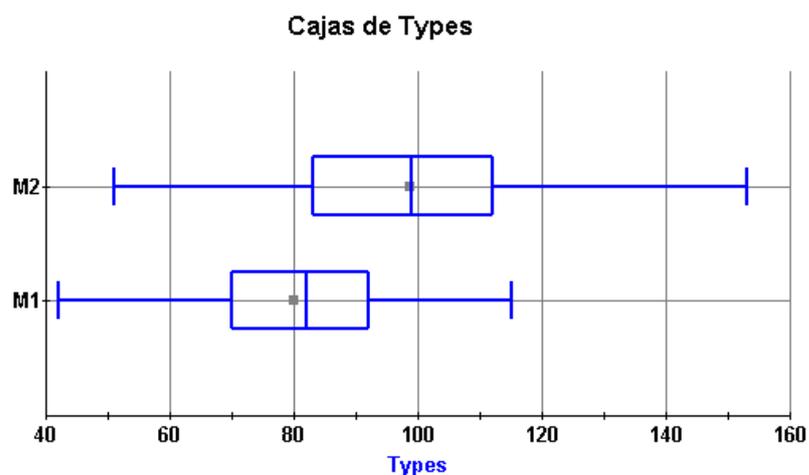


Gráfico 74. Diagrama de Cajas para la variable «Types» NN-M1 / NN-M2

Comprobamos al visualizar el diagrama de cajas el mayor tamaño de la caja de la muestra del M-2. Por tanto, no solo la mayor dispersión de los datos la encontramos en el

conjunto de la muestra, sino también dentro de la caja, donde se encuentra el 50% de los datos de toda la muestra.

Otro rasgo que llama la atención de la muestra del M-2 es la casi coincidencia entre media y mediana. Mientras que en la muestra del M-1 las dos medidas presentan una distancia algo mayor. Esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

También podemos apreciar en el diagrama de cajas la gran diferencia que existe entre el M-1 y el M-2 respecto al estadístico «máximo».

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 43) que la curtosis es negativa en ambas muestras y con valores muy próximos el uno al otro: M-1 curtosis= -0.3656; M-2 curtosis= -0.3445. Eso significa que las curvas tienden a una forma más plana que una curtosis cero. Los valores nos indican que la forma de las curvas no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

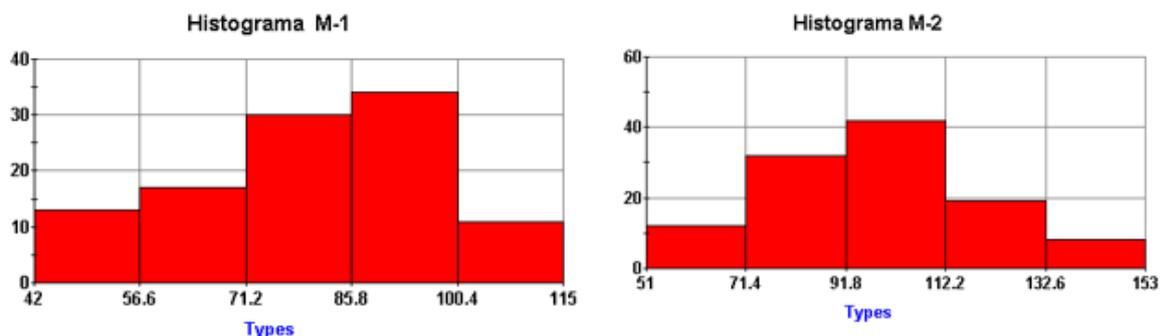


Gráfico 75. Histograma para la variable «Types» NN-M1 / NN-M2

(3) Verbos: Total de verbos.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Verbos	2.190	3.231

A partir de los totales de verbos que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Verbos por V1**

<b>Grupos</b>	<b>M1</b>	<b>M2</b>
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>31.0286</b>	<b>40.9292</b>
<b>Mediana</b>	<b>31.0000</b>	<b>40.0000</b>
<b>Varianza</b>	<b>130.2973</b>	<b>248.8878</b>
<b>Desviación Típica</b>	<b>11.4148</b>	<b>15.7762</b>
<b>Mínimo</b>	<b>8.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>62.0000</b>	<b>95.0000</b>
<b>Cuartil Inferior</b>	<b>24.0000</b>	<b>29.0000</b>
<b>Cuartil Superior</b>	<b>39.0000</b>	<b>50.0000</b>
<b>Rango Intercuartílico</b>	<b>15.0000</b>	<b>21.0000</b>
<b>Curtosis</b>	<b>-0.3264</b>	<b>0.2703</b>

Tabla 44. Estadísticos para la variable «Verbos» NN-M1 / NN-M2

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos una vez más en la muestra del M-2, como en las variables anteriores. Advertimos que la media de la muestra del M-2 supera en casi diez puntos la muestra del M-1, y la mediana en nueve puntos.

Llama la atención, asimismo, la proximidad entre los descriptores de «media» y «mediana» en cada una de las muestras. No obstante, comprobamos que es en la muestra del M-2 donde se da una diferencia algo mayor. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 77).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=15.7762$ . Mientras que la muestra del M-1 registra un valor  $s=11.4148$ . Es decir, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el gráfico de dispersión (Gráfico 76), así como los valores extremos.

El otro descriptor que informa de la dispersión, el rango intercuartílico, presenta un valor más alto en la muestra del M-2 (IQR=21.0000); mientras que la muestra del M-1 registra un valor de IQR=15.0000. Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el

diagrama de cajas (Gráfico 77) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

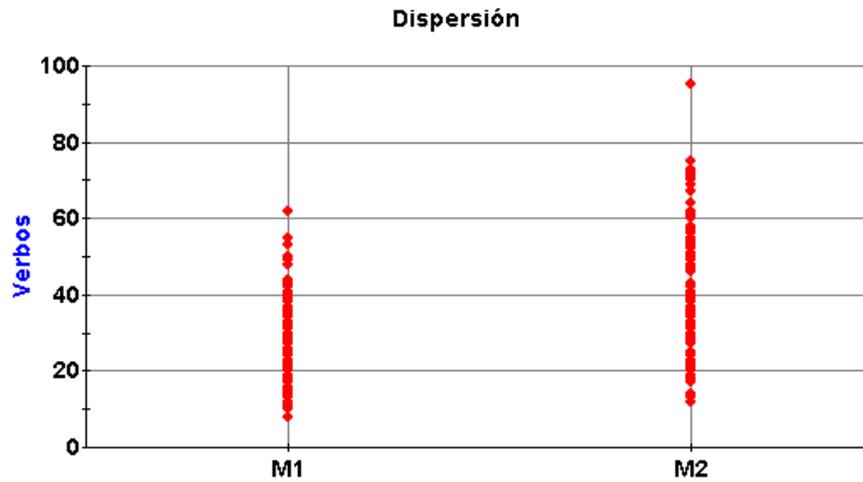


Gráfico 76. Gráfico de dispersión para la variable «Verbos» NN-M1 / NN-M2

Al visualizar el diagrama de cajas, lo primero que llama la atención es, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, comprobamos la distancia que hay entre media y mediana en cada una de las muestras. Ambos valores, como hemos indicado más arriba, están muy próximos el uno al otro, pero comprobamos que en la muestra del M-1 son prácticamente coincidentes.

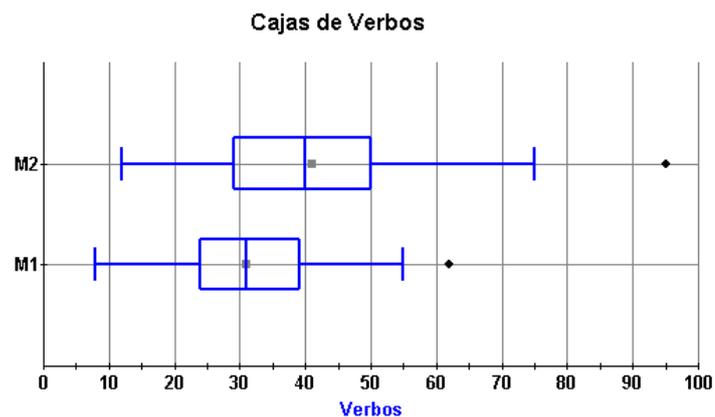


Gráfico 77. Diagrama de Cajas para la variable «Verbos» NN-M1 / NN-M2

En cuanto a los descriptores «mínimo» y «máximo», vemos que los valores del «mínimo» están próximos, sin embargo, los del «máximo» marcan diferencia entre el M-1 y el M-2.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 44) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.3264), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.2703, por tanto, la curva tiende a una forma más aguda que una curtosis cero. Los valores nos indican que la forma de las curvas de cada una de las muestras, aun siendo valores opuestos, uno negativo y otro positivo, no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

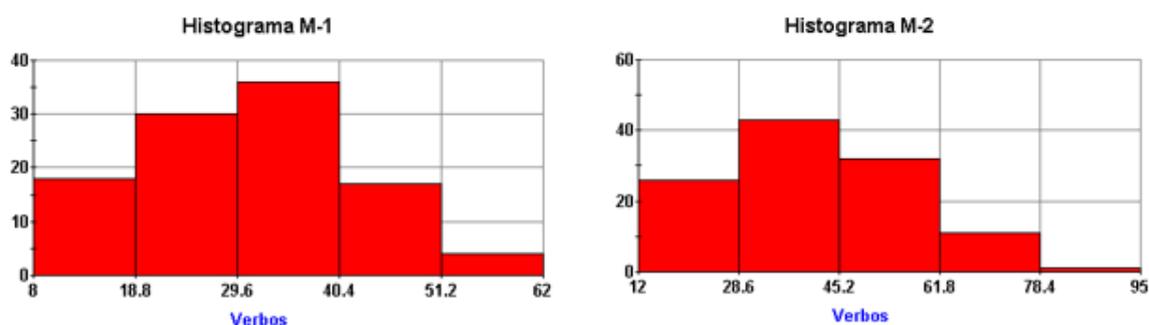


Gráfico 78. Histograma para la variable «Verbos» NN-M1 / NN-M2

(4) Sustantivos: Total de sustantivos.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Sustantivos	5.198	7.042

A partir de los totales de sustantivos que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

Cuando observamos la tabla de estadísticos (Tabla 45), vemos que los valores más altos de las medidas de centralización, media y mediana, corresponden una vez más a la muestra del M-2, como en las variables anteriores. Comprobamos que la media de la muestra del M-2 supera en casi quince puntos la muestra del M-1, y la mediana en once puntos.

En cuanto a los descriptores de «media» y «mediana» en cada una de las muestras, observamos que la distancia en ambos casos es mínima. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 80).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=21.1195$ . Mientras que la muestra del M-1 registra un valor  $s=17.6480$ . Es decir, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el gráfico de dispersión (Gráfico 79), así como los valores extremos.

**Estadísticos para la variable Sustantivos**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>60.8571</b>	<b>75.8053</b>
<b>Mediana</b>	<b>63.0000</b>	<b>74.0000</b>
<b>Varianza</b>	<b>311.4505</b>	<b>446.0332</b>
<b>Desviación Típica</b>	<b>17.6480</b>	<b>21.1195</b>
<b>Mínimo</b>	<b>26.0000</b>	<b>36.0000</b>
<b>Máximo</b>	<b>100.0000</b>	<b>132.0000</b>
<b>Cuartil Inferior</b>	<b>50.0000</b>	<b>63.0000</b>
<b>Cuartil Superior</b>	<b>74.0000</b>	<b>88.0000</b>
<b>Rango Intercuartílico</b>	<b>24.0000</b>	<b>25.0000</b>
<b>Curtosis</b>	<b>-0.4779</b>	<b>-0.0919</b>

Tabla 45. Estadísticos para la variable «Sustantivos» NN-M1 / NN-M2

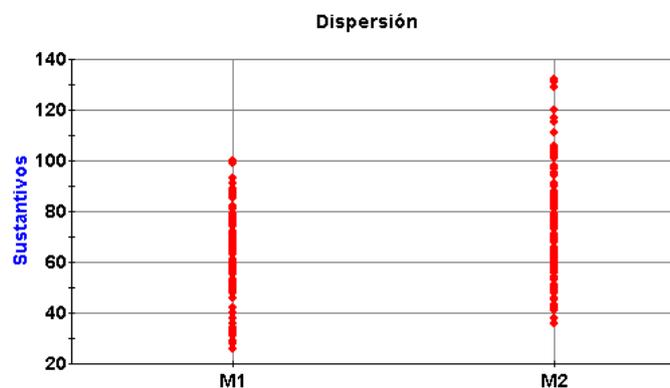


Gráfico 79. Gráfico de dispersión para la variable «Sustantivos» NN-M1 / NN-M2

Respecto al otro descriptor que informa de la dispersión, el rango intercuartílico, vemos que el valor es muy similar en las dos muestras: M-1 IQR=24.0000; M-2, IQR=25.0000. Por tanto, en esta variable, la dispersión de los datos dentro de las cajas va a ser casi la misma. En el diagrama de cajas (Gráfico 80) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

Comprobamos, al visualizar el diagrama, que el tamaño de las cajas es prácticamente igual. También podemos ver que la distancia entre los descriptores de «media» y «mediana» es mínima en las dos muestras. Aunque esa pequeña diferencia indica que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

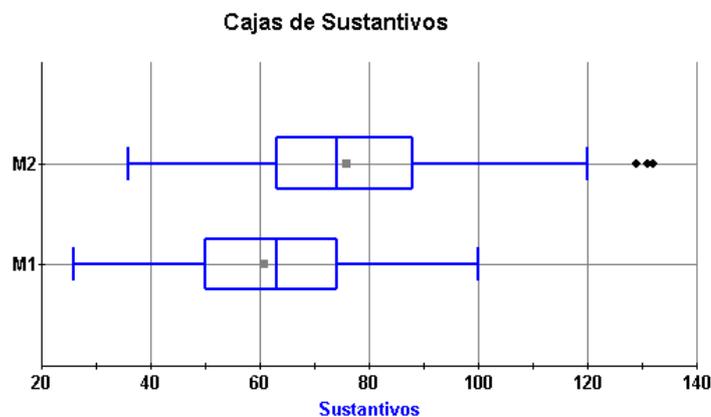


Gráfico 80. Diagrama de Cajas para la variable «Sustantivos» NN-M1 / NN-M2

También podemos apreciar en el diagrama de cajas la diferencia que se mantiene respecto al estadístico «máximo» entre la muestra del M-1 y M-2, aunque es una diferencia menor a las anteriores variables.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 45) que la curtosis es negativa en ambas muestras: M-1 curtosis= -0.4779; M-2 curtosis= -0.0919. Eso significa que las curvas tienden a una forma más plana que una curtosis cero. No obstante, los valores nos indican que la forma de las curvas no se aleja demasiado de una campana de Gauss, pues ambos están próximos a una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

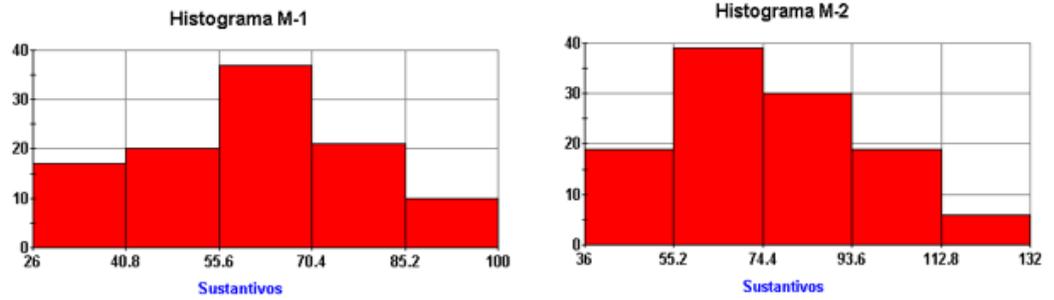


Gráfico 81. Histograma para la variable «Sustantivos» NN-M1 / NN-M2

(5) Adjetivos: Total de adjetivos.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Adjetivos	1.742	2.232

A partir de los totales de adjetivos que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos (Tabla 46) para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos una vez más en la muestra del M-2, como en las variables anteriores. Advertimos que en esta variable las diferencias son menores que en las anteriores. Aquí la media de la muestra del M-2 supera en casi siete puntos la muestra del M-1, y la mediana en cinco puntos.

**Estadísticos para la variable Adjetivos**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>27.1810</b>	<b>33.9823</b>
<b>Mediana</b>	<b>27.0000</b>	<b>32.0000</b>
<b>Varianza</b>	<b>88.0535</b>	<b>134.5175</b>
<b>Desviación Típica</b>	<b>9.3837</b>	<b>11.5982</b>
<b>Mínimo</b>	<b>7.0000</b>	<b>13.0000</b>
<b>Máximo</b>	<b>46.0000</b>	<b>76.0000</b>
<b>Cuartil Inferior</b>	<b>20.0000</b>	<b>26.0000</b>
<b>Cuartil Superior</b>	<b>34.0000</b>	<b>42.0000</b>
<b>Rango Intercuartílico</b>	<b>14.0000</b>	<b>16.0000</b>
<b>Curtosis</b>	<b>-0.7436</b>	<b>0.6337</b>

Tabla 46. Estadísticos para la variable «Adjetivos» NN-M1 / NN-M2

Llama la atención, asimismo, la proximidad entre los descriptores de «media» y «mediana» en cada una de las muestras. No obstante, comprobamos que es en la muestra del M-2 donde se da una diferencia algo mayor, casi de dos puntos. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 83).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=11.5982$ . Mientras que la muestra del M-1 registra un valor  $s=9.3837$ . También en este descriptor advertimos diferencias menores que en las variables anteriores. No obstante, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el gráfico siguiente, así como los valores extremos.

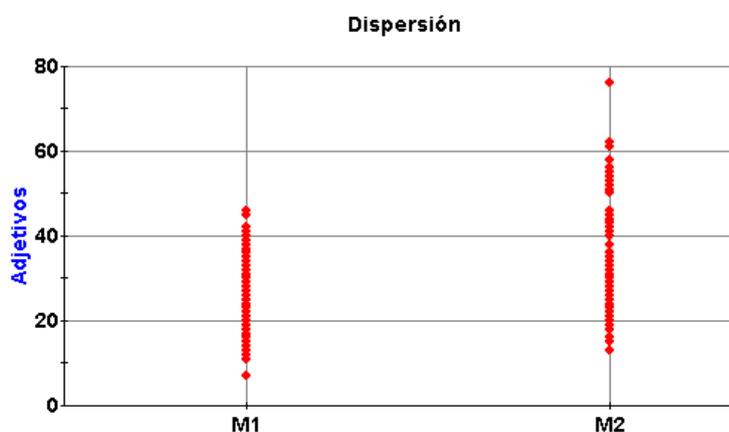


Gráfico 82. Gráfico de dispersión para la variable «Adjetivos» NN-M1 / NN-M2

El otro descriptor que informa de la dispersión, el rango intercuartílico, presenta un valor más alto en la muestra del M-2 ( $IQR=16.0000$ ); mientras que la muestra del M-1 registra un valor de  $IQR=14.0000$ . Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas que mostramos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

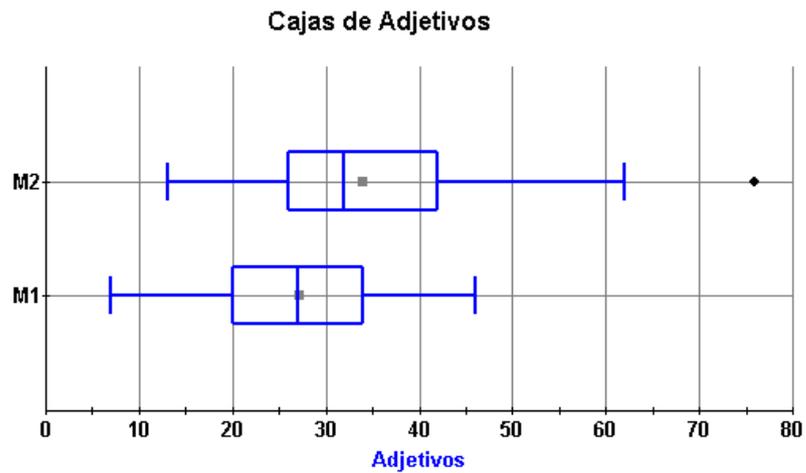


Gráfico 83. Diagrama de Cajas para la variable «Adjetivos» NN-M1 / NN-M2

Comprobamos al visualizar el diagrama de cajas, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, comprobamos la distancia que hay entre media y mediana en cada una de las muestras. Ambos valores, como hemos indicado más arriba, están muy próximos el uno al otro, pero constatamos que en la muestra del M-1 son prácticamente coincidentes.

En cuanto al descriptor «máximo», vemos que sigue la misma tónica que en las anteriores variables, la muestra del M-2 se distancia de la del M-1.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 46) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.7436), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.6337, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

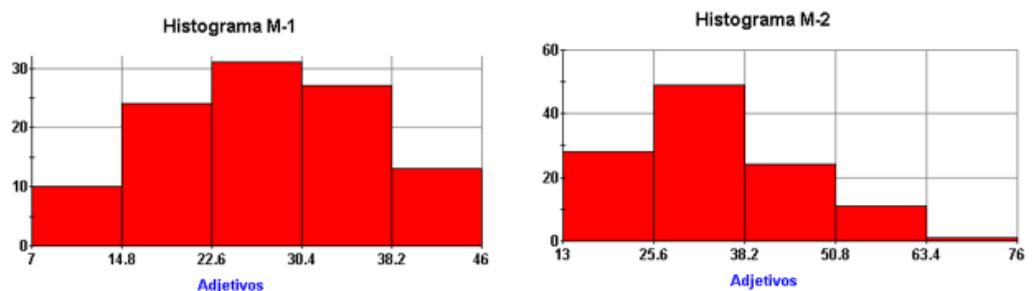


Gráfico 84. Histograma para la variable «Adjetivos» NN-M1 / NN-M2.

(6) Adverbios: Total de adverbios.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Adverbios	785	949

A partir de los totales de adverbios que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Adverbios**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>16.7143</b>	<b>20.9912</b>
<b>Mediana</b>	<b>17.0000</b>	<b>19.0000</b>
<b>Varianza</b>	<b>44.0137</b>	<b>72.2053</b>
<b>Desviación Típica</b>	<b>6.6343</b>	<b>8.4974</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>7.0000</b>
<b>Máximo</b>	<b>36.0000</b>	<b>49.0000</b>
<b>Cuartil Inferior</b>	<b>12.0000</b>	<b>15.0000</b>
<b>Cuartil Superior</b>	<b>21.0000</b>	<b>27.0000</b>
<b>Rango Intercuartílico</b>	<b>9.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.2159</b>	<b>0.6140</b>

Tabla 47. Estadísticos para la variable «Adverbios» NN-M1 / NN-M2

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos una vez más en la muestra del M-2, como en las variables anteriores, aunque comprobamos que las diferencias se reducen. La media de la muestra del M-2 supera solo en poco más de cuatro puntos la muestra del M-1, y la mediana en dos puntos.

Llama la atención, sin embargo, la distancia entre los descriptores de «media» y «mediana» en la muestra del M-2, casi dos puntos; mientras que la distancia en la muestra del M-1 es mínima. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 86).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta,

igualmente, la muestra del M-2,  $s=8.4974$ . Mientras que la muestra del M-1 registra un valor  $s=6.6343$ . Es decir, en el M-2 los datos presentan una mayor dispersión que podemos visualizar en el gráfico de dispersión (Gráfico 85), así como los valores extremos.

El otro descriptor que informa de la dispersión, el rango intercuartílico, presenta un valor más alto en la muestra del M-2 ( $IQR=12.0000$ ); mientras que la muestra del M-1 registra un valor de  $IQR=9.0000$ . Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas (Gráfico 86) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

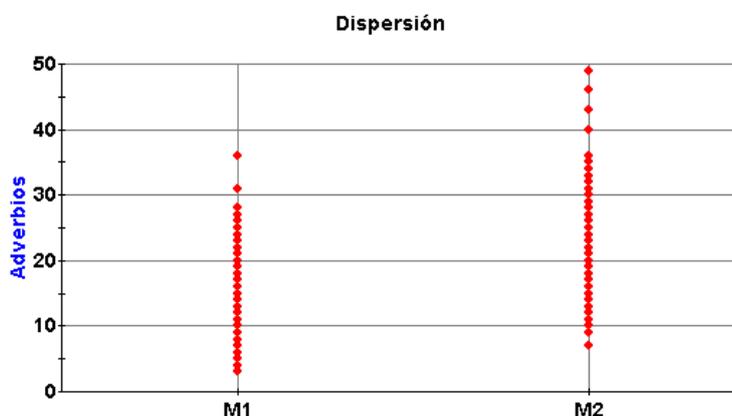


Gráfico 85. Gráfico de dispersión para la variable «Adverbios» NN-M1 / NN-M2

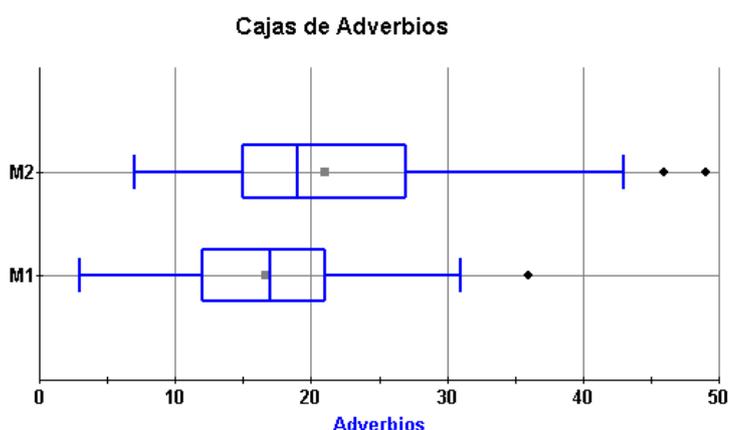


Gráfico 86. Diagrama de Cajas para la variable «Adverbios» NN-M1 / NN-M2

Comprobamos al visualizar el diagrama de cajas, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado. Por otro lado, confirmamos que la

distancia que hay entre media y mediana en la muestra del M-2 es mayor que en la muestra del M-1. Esas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a los descriptores «mínimo» y «máximo», vemos que, mientras el primero mantiene cierta proximidad entre el M-1 y el M-2, los valores del estadístico «máximo» se alejan entre ambos momentos.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 47) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.2159), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.6140, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

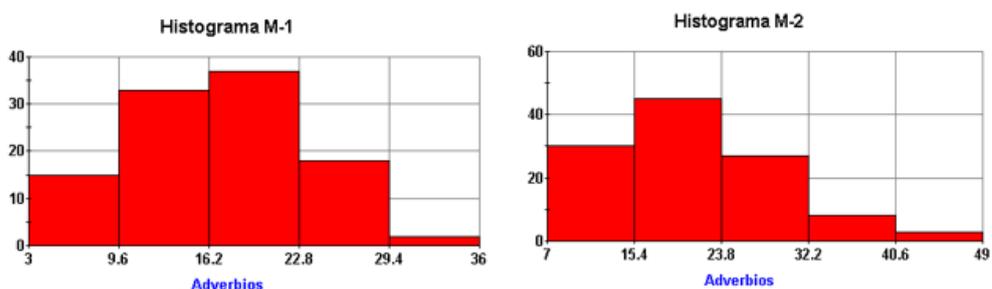


Gráfico 87. Histograma para la variable «Adverbios» NN-M1 / NN-M2

(7) Pronombres: Total de pronombres.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Pronombres	880	1.432

A partir de los totales de pronombres que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Pronombres :**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>20.1238</b>	<b>27.5487</b>
<b>Mediana</b>	<b>19.0000</b>	<b>26.0000</b>
<b>Varianza</b>	<b>38.3980</b>	<b>86.8748</b>
<b>Desviación Típica</b>	<b>6.1966</b>	<b>9.3207</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>38.0000</b>	<b>64.0000</b>
<b>Cuartil Inferior</b>	<b>17.0000</b>	<b>21.0000</b>
<b>Cuartil Superior</b>	<b>25.0000</b>	<b>33.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.1365</b>	<b>1.3328</b>

Tabla 48. Estadísticos para la variable «Pronombres» NN-M1 / NN-M2

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos una vez más en la muestra del M-2, como en las variables anteriores. Aquí la media de la muestra del M-2 supera en algo más de siete puntos la muestra del M-1, y la mediana en siete puntos.

Vemos, asimismo, que hay cierta distancia entre los valores de los descriptores de «media» y «mediana» en cada una de las muestras. Comprobamos que es en la muestra del M-2 donde se da una diferencia algo mayor. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 89).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=9.3207$ . Mientras que la muestra del M-1 registra un valor  $s=6.1966$ . Podemos visualizar en el gráfico siguiente estas diferencias, así como los valores extremos.

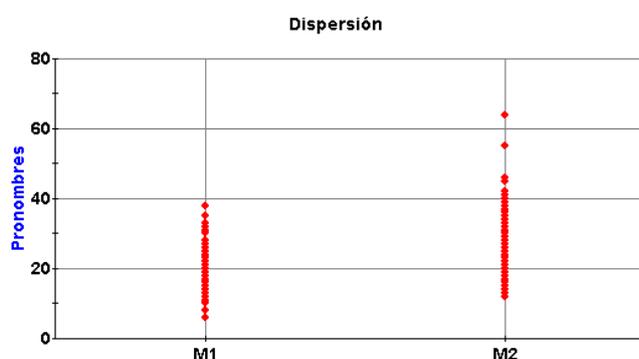


Gráfico 88. Gráfico de dispersión para la variable «Pronombres» NN-M1 / NN-M2

El otro descriptor que informa de la dispersión, el rango intercuartílico, presenta un valor más alto en la muestra del M-2 (IQR=12.0000); mientras que la muestra del M-1 registra un valor de IQR=8.0000. Por tanto, también los valores incluidos en la caja de la muestra del M-2 presentan mayor dispersión que los de la caja de la muestra del M-1, en la que los valores van a estar más concentrados en torno a las medidas de centralización. En el diagrama de cajas (Gráfico 89) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

Al observar el diagrama de cajas comprobamos, por un lado, la diferencia en el tamaño de las cajas, como ya hemos mencionado, más grande la del M-2. Por otro lado, advertimos cierta distancia entre media y mediana en cada una de las muestras. Esas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto al descriptor «máximo», vemos que la muestra del M-2 se distancia de la del M-1, igual que en las variables anteriores.

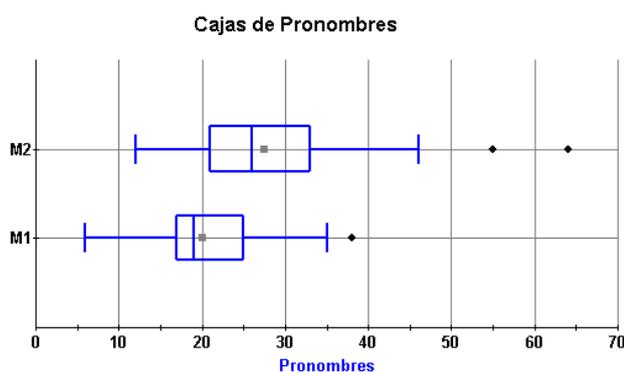


Gráfico 89. Diagrama de Cajas para la variable «Pronombres» NN-M1 / NN-M2

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 48) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.1365), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=1.3328, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

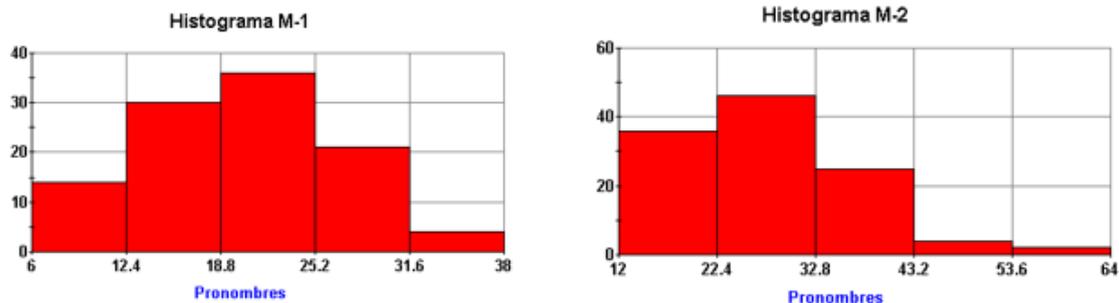


Gráfico 90. Histograma para la variable «Pronombres» NN-M1 / NN-M2

(8) Preposiciones: Total de preposiciones.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Preposiciones	579	755

A partir de los totales de preposiciones que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

**Estadísticos para la variable Preposiciones**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>16.2857</b>	<b>20.7257</b>
<b>Mediana</b>	<b>16.0000</b>	<b>19.0000</b>
<b>Varianza</b>	<b>33.8022</b>	<b>60.2366</b>
<b>Desviación Típica</b>	<b>5.8140</b>	<b>7.7612</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>4.0000</b>
<b>Máximo</b>	<b>32.0000</b>	<b>41.0000</b>
<b>Cuartil Inferior</b>	<b>12.0000</b>	<b>16.0000</b>
<b>Cuartil Superior</b>	<b>21.0000</b>	<b>25.0000</b>
<b>Rango Intercuartílico</b>	<b>9.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>-0.2782</b>	<b>0.2520</b>

Tabla 49. Estadísticos para la variable «Preposiciones» NN-M1 / NN-M2

Observamos en la tabla de estadísticos que los valores más altos de las medidas de centralización, media y mediana, corresponden una vez más a la muestra del M-2, como en las variables anteriores. No obstante, advertimos que en esta variable las diferencias son menores que en las anteriores. Aquí la media de la muestra del M-2 supera en algo más de cuatro puntos la muestra del M-1, y la mediana en tres puntos.

En cuanto a los valores de los descriptores de «media» y «mediana» en cada una de las muestras, comprobamos que es en la muestra del M-2 donde se da una diferencia algo mayor que en la del M-1. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 92).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=7.7612$ . Mientras que la muestra del M-1 registra un valor  $s=5.8140$ . En el siguiente gráfico podemos visualizar las pequeñas diferencias, así como los valores extremos.

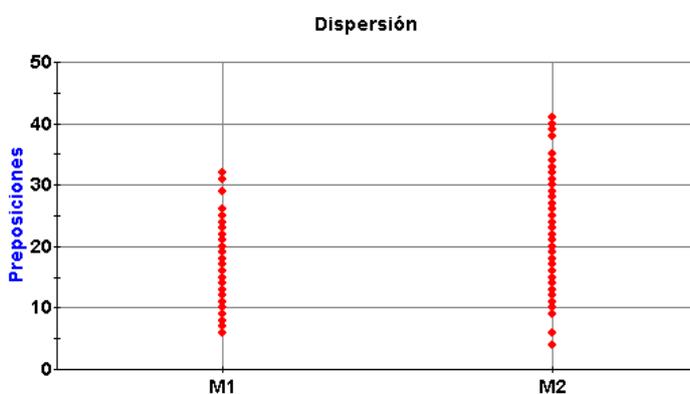


Gráfico 91. Gráfico de dispersión para la variable «Preposiciones» NN-M1 / NN-M2

En cuanto al otro descriptor que informa de la dispersión, el rango intercuartílico, vemos que presenta el mismo valor en ambas muestras ( $IQR=9.0000$ ), algo que no se ha dado en ninguna de las otras variables. Por tanto, los valores incluidos dentro de las cajas, recordemos que son el 50% del total de la muestra, tienen la misma dispersión. En el diagrama de cajas (Gráfico 92) podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

Comprobamos en el diagrama de cajas, por un lado, la coincidencia de tamaño de las dos cajas. Por otro lado, advertimos cierta distancia entre media y mediana en la muestra del M-2; mientras que en la muestra del M-1 hay casi coincidencia entre esas dos medidas de centralización. Aun así, esas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

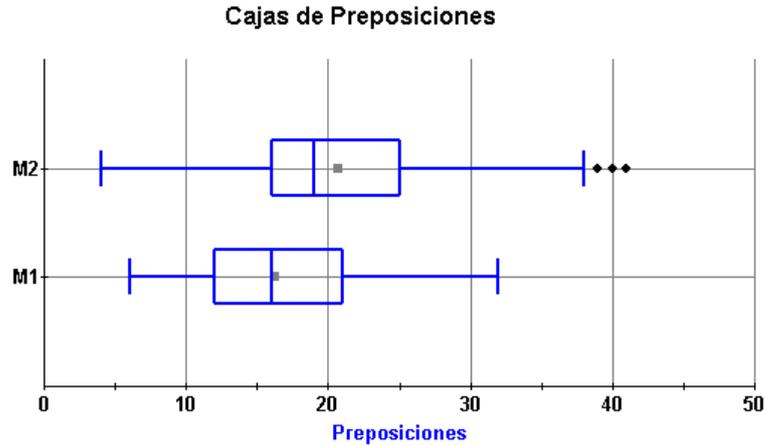


Gráfico 92. Diagrama de Cajas para la variable «Preposiciones» NN-M1 / NN-M2

En cuanto al descriptor «mínimo», vemos que la muestra del M-2 presenta un valor inferior al del M-1; mientras que el descriptor «máximo», al igual que en el resto de variables, registra un valor mayor al del M-1.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 49) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.27.82), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.2520, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

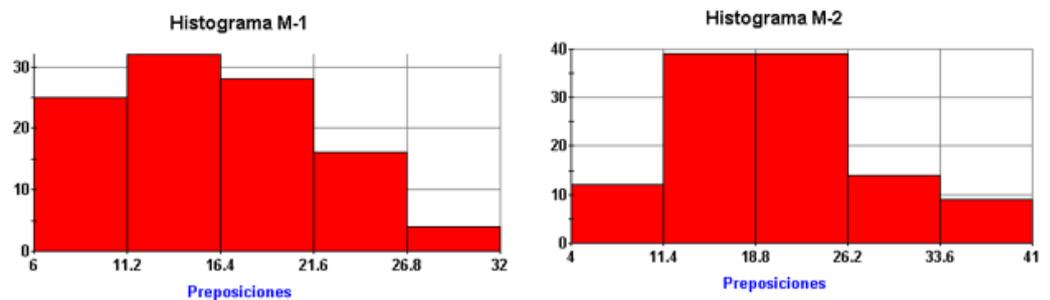


Gráfico 93. Histograma para la variable «Preposiciones» NN-M1 / NN-M2

(9) Artículos: Total de artículos.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Artículos	235	307

A partir de los totales de artículos que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos (Tabla 50) para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

Comprobamos una vez más que los valores de los estadísticos de centralización, media y mediana, corresponden a la muestra del M-2, como en las variables anteriores. Aquí la media de la muestra del M-2 supera en algo más de cuatro puntos la muestra del M-1, y la mediana en cuatro puntos.

En cuanto a los valores de los descriptores de «media» y «mediana» en cada una de las muestras, comprobamos que es en la muestra del M-1 donde se da una diferencia algo mayor en comparación con la del M-2. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 95).

**Estadísticos para la variable Artículos**

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>10.4000</b>	<b>14.7965</b>
<b>Mediana</b>	<b>11.0000</b>	<b>15.0000</b>
<b>Varianza</b>	<b>19.3192</b>	<b>28.2171</b>
<b>Desviación Típica</b>	<b>4.3954</b>	<b>5.3120</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>5.0000</b>
<b>Máximo</b>	<b>22.0000</b>	<b>29.0000</b>
<b>Cuartil Inferior</b>	<b>7.0000</b>	<b>11.0000</b>
<b>Cuartil Superior</b>	<b>14.0000</b>	<b>18.0000</b>
<b>Rango Intercuartílico</b>	<b>7.0000</b>	<b>7.0000</b>
<b>Curtosis</b>	<b>-0.6628</b>	<b>0.0198</b>

Tabla 50. Estadísticos para la variable «Artículos» NN-M1 / NN-M2

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=5.3120$ . Mientras que la muestra del M-1 registra un valor  $s=4.3954$ . Como vemos, la diferencia es mínima. En el siguiente gráfico podemos visualizar las pequeñas diferencias, así como los valores extremos.

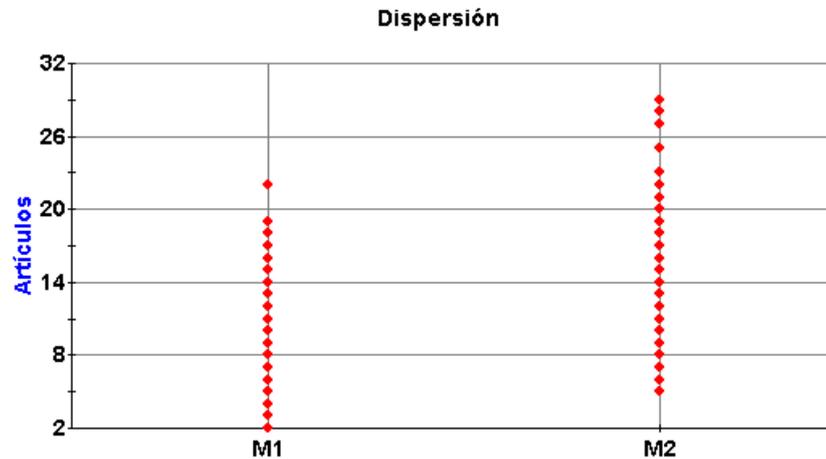


Gráfico 94. Gráfico de dispersión para la variable «Artículos» NN-M1 / NN-M2

En cuanto al otro descriptor que informa de la dispersión, el rango intercuartílico, al igual que en la variable anterior, vemos que presenta el mismo valor en ambas muestras (IQR=7.0000). Por tanto, los valores incluidos dentro de las cajas tienen la misma dispersión. En el diagrama de cajas que mostramos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

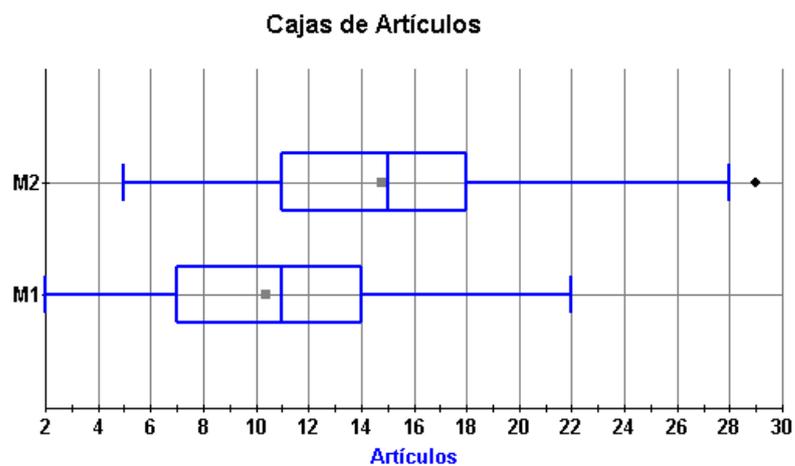


Gráfico 95. Diagrama de Cajas para la variable «Artículos» NN-M1 / NN-M2

Al visualizar el diagrama de cajas comprobamos, por un lado, la coincidencia de tamaño de las dos cajas. Por otro lado, advertimos cierta distancia entre media y mediana en

la muestra del M-1; mientras que en la muestra del M-2 hay casi coincidencia entre esas dos medidas de centralización. Aun así, esas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto al descriptor «mínimo» y «máximo», vemos que la muestra del M-2 presenta valores superiores al del M-1.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 50) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.6628), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=0.0198, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

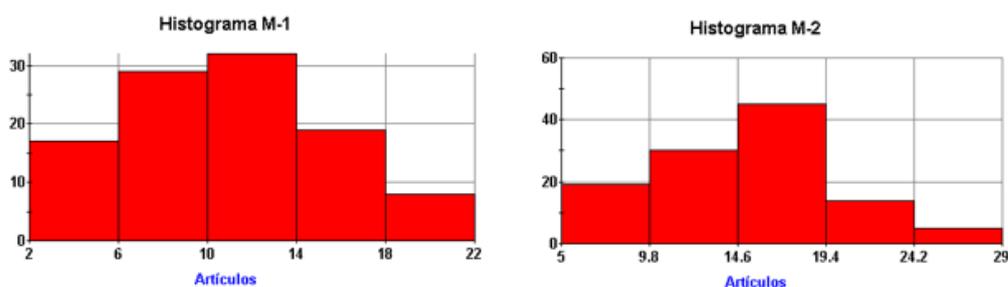


Gráfico 96. Histograma para la variable «Artículos» NN-M1 / NN-M2

(10) Conjunciones: Total de conjunciones.

Muestras «M1» y «M-2»		
	NN M-1(n=105)	NN M-2 (n=113)
Conjunciones	576	769

A partir de los totales de conjunciones que ofrecen las dos muestras del grupo de NN, vamos a ver qué dicen los estadísticos para comprobar si esa diferencia se refleja en las medidas de centralización, dispersión y en la forma de la distribución de las muestras.

Al observar los estadísticos de centralización, media y mediana, vemos que los valores más altos los encontramos una vez más en la muestra del M-2, como en las variables anteriores. Aquí la media de la muestra del M-2 supera en poco más de cinco puntos la muestra del M-1, y la mediana en cinco puntos.

### Estadísticos para la variable Conjunciones

Grupos	M1	M2
<b>N</b>	<b>105</b>	<b>113</b>
<b>Media</b>	<b>16.1524</b>	<b>21.2389</b>
<b>Mediana</b>	<b>16.0000</b>	<b>21.0000</b>
<b>Varianza</b>	<b>50.7266</b>	<b>90.0942</b>
<b>Desviación Típica</b>	<b>7.1223</b>	<b>9.4918</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>4.0000</b>
<b>Máximo</b>	<b>36.0000</b>	<b>68.0000</b>
<b>Cuartil Inferior</b>	<b>12.0000</b>	<b>15.0000</b>
<b>Cuartil Superior</b>	<b>21.0000</b>	<b>27.0000</b>
<b>Rango Intercuartílico</b>	<b>9.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.2330</b>	<b>4.2664</b>

Tabla 51. Estadísticos para la variable «Conjunciones» NN-M1 / NN-M2

Vemos, asimismo, que los valores de los descriptores de «media» y «mediana» dentro de cada una de las muestras casi coinciden. Podemos visualizar estas diferencias en el diagrama de cajas (Gráfico 98).

Respecto a la desviación típica, el estadístico que informa de la dispersión de los valores que conforman las muestras, comprobamos que la mayor dispersión la presenta, igualmente, la muestra del M-2,  $s=9.4918$ . Mientras que la muestra del M-1 registra un valor  $s=7.1223$ . Podemos visualizar en el gráfico siguiente estas diferencias, así como los valores extremos.

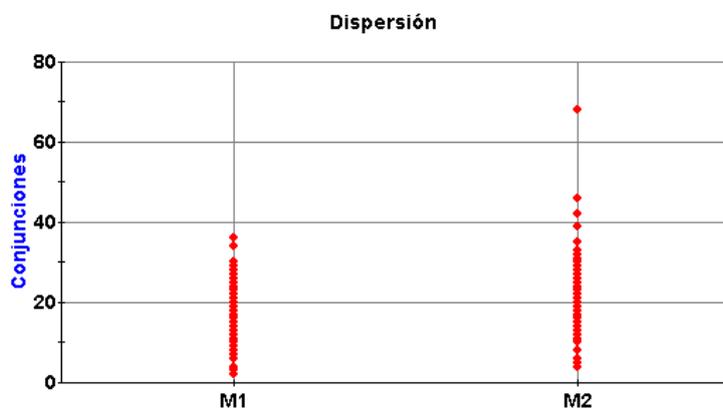


Gráfico 97. Gráfico de dispersión para la variable «Conjunciones» NN-M1 / NN-M2

En cuanto al otro descriptor que informa de la dispersión, el rango intercuartílico, comprobamos que en esta variable vuelve a presentar el valor más alto la muestra del M-2 (IQR=12.0000). La muestra del M-1 registra un valor IQR=9.0000. Así, la caja de la muestra del M-2 es más grande y por tanto los valores incluidos en ella presentan una mayor dispersión. En el diagrama de cajas que mostramos a continuación podemos visualizar estos datos y el resto de valores de la tabla de estadísticos.

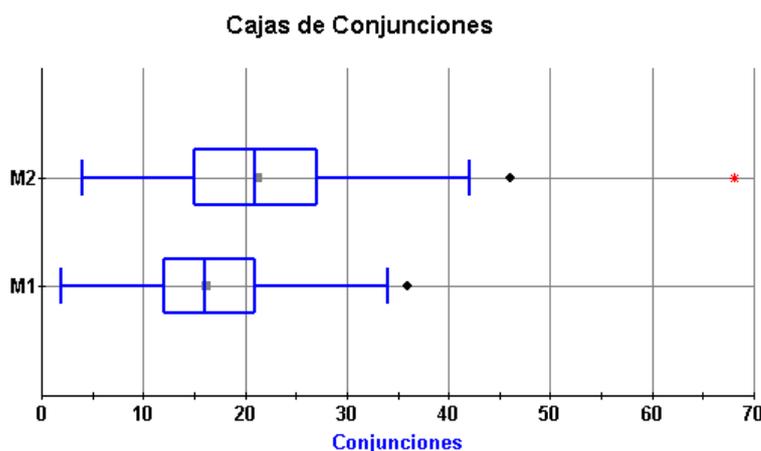


Gráfico 98. Diagrama de Cajas para la variable «Conjunciones» NN-M1 / NN-M2

Comprobamos en el diagrama de cajas, por un lado, el mayor tamaño de la caja de la muestra del M-2. Por otro lado, advertimos la casi coincidencia entre las medidas de centralización de cada una de las muestras. Aun así, las pequeñas diferencias que presentan las muestras indican una tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto al descriptor «mínimo», vemos que los valores de la muestra del M-2 y los del M-1 están muy próximos entre sí.

Respecto a la forma que adopta la curva de cada una de las muestras, vemos en la tabla de estadísticos (Tabla 51) que la curtosis es negativa en la muestra del M-1 (curtosis= -0.2330), eso significa que la curva tiende a una forma más plana que una curtosis cero. Mientras que el valor de la muestra del M-2 es positivo, curtosis=4.2664, por tanto, la curva tiende a una forma más aguda que una curtosis cero. En los histogramas siguientes podemos visualizar esa tendencia.

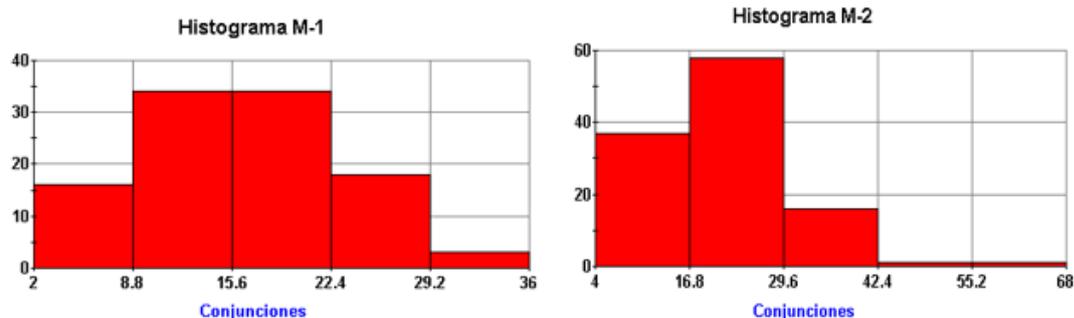


Gráfico 99. Histograma para la variable «Conjunciones» NN-M1 / NN-M2

#### 6.1.4 Estadística descriptiva para la muestra «F» por estadios

En la tabla 52, presentamos los datos de los que partimos. Observando los totales que tenemos, vemos que los del grupo «NN-M1-F» (no nativos, momento 1) están muy próximos al grupo de nativos, «N-F». Mientras que la muestra «NN-M2-F» (no nativos, momento 2) supera en valores absolutos al grupo de nativos. Sin embargo, hay que tener en cuenta también la diferencia de producciones en un caso y en el otro.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
<i>Tokens</i>	9.088	13.144	9.661
<i>Types</i>	6.157	8.592	6.341
Verbos	1.686	2.683	1.546
Sustantivos	3.920	5.467	3.998
Adjetivos	1.285	1.694	1.182
Adverbios	566	771	376
Pronombres	611	1.092	981
Preposiciones	404	577	695
Artículos	161	230	218
Conjunciones	455	630	665

Tabla 52. Totales de las variables para las muestras «F» NN-M1-F / NN-M2-F / N-F

A continuación presentamos el análisis descriptivo de estos datos por variables.

(11) *Tokens*: Total de palabras.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
<i>Tokens</i>	9.088	13.144	9.661

A partir de estos valores absolutos, vamos a ver qué nos dicen los estadísticos.

En la tabla 53 podemos observar los diferentes descriptores que nos informan de las medidas de tendencia central, de dispersión y de forma.

Como comprobamos en la tabla 53, las medias de las tres muestras son muy diferentes, siendo la más baja la del grupo de no nativos en el M1 (media=128.000) y la más alta la del grupo de nativos (media=175.6545). En cuanto a la desviación típica, la más baja vuelve a ser la del grupo NN-M1-F (s=21.3334), sin embargo, la más alta corresponde, en este caso, al grupo NN-M2-F (s=39.6262). Esto nos indica que el grupo NN-M2-F es el que mayor dispersión muestra. Esto se puede observar en el gráfico 100 de dispersión y también en el gráfico 101 de cajas.

Como hemos indicado más arriba, es en el gráfico de cajas donde quedan reflejados gráficamente los estadísticos para cada una de las muestras. Como se puede observar, en los tres casos la mediana está por debajo de la media, aunque con poca diferencia. Eso significa que estamos ante casos en los que se dan algunos valores extremos.

**Estadísticos para la variable Tokens .**

<b>Grupos</b>	<b>NN_M1_F</b>	<b>NN_M2_F</b>	<b>N_F</b>
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>128.0000</b>	<b>160.2927</b>	<b>175.6545</b>
<b>Mediana</b>	<b>126.0000</b>	<b>157.0000</b>	<b>174.0000</b>
<b>Varianza</b>	<b>455.1143</b>	<b>1570.2343</b>	<b>882.5636</b>
<b>Desviación Típica</b>	<b>21.3334</b>	<b>39.6262</b>	<b>29.7080</b>
<b>Mínimo</b>	<b>76.0000</b>	<b>72.0000</b>	<b>127.0000</b>
<b>Máximo</b>	<b>191.0000</b>	<b>283.0000</b>	<b>250.0000</b>
<b>Cuartil Inferior</b>	<b>112.0000</b>	<b>131.0000</b>	<b>152.0000</b>
<b>Cuartil Superior</b>	<b>141.0000</b>	<b>184.0000</b>	<b>191.0000</b>
<b>Rango Intercuartílico</b>	<b>29.0000</b>	<b>53.0000</b>	<b>39.0000</b>
<b>Curtosis</b>	<b>0.2517</b>	<b>0.4223</b>	<b>-0.0715</b>

Tabla 53. Estadísticos para la variable «Tokens» NN-M1-F / NN-M2-F / N-F

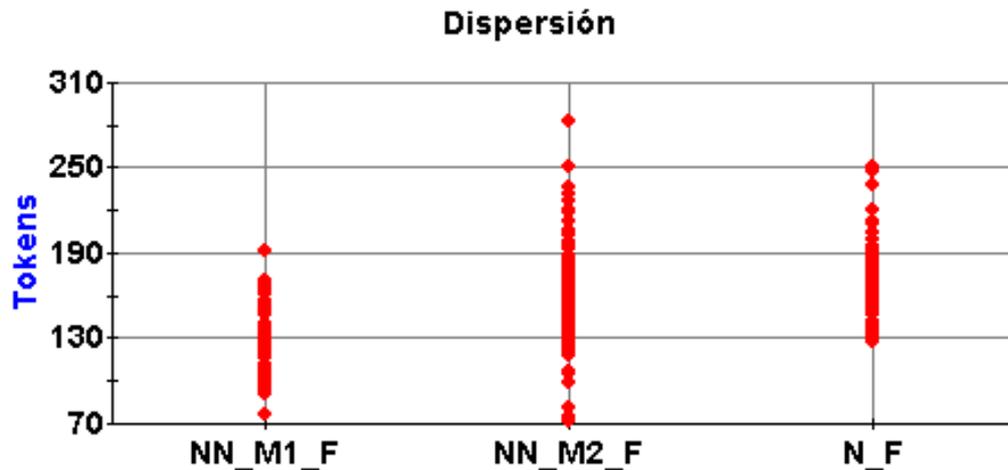


Gráfico 100. Gráfico de dispersión para la variable «Tokens» NN-M1-F / NN-M2-F / N-F

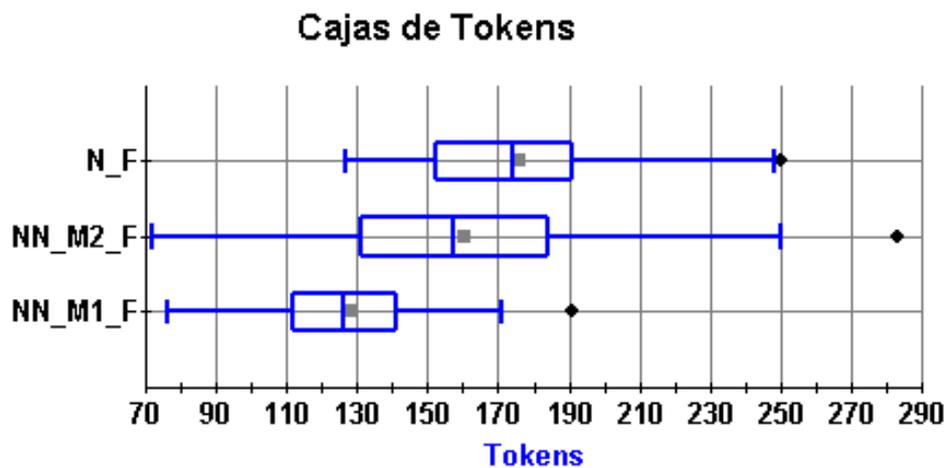


Gráfico 101. Diagrama de Cajas para la variable «Tokens» NN-M1-F / NN-M2-F / N-F

Al observar en la tabla de estadísticos y en el gráfico de cajas el otro valor que habla de la dispersión, el descriptor «rango intercuartílico», vemos que el valor más alto es el correspondiente a la muestra «NN-M2-F» (no Nativos M2). Es decir, en esta muestra la concentración en torno a la media y la mediana es menor que en las otras dos muestras.

Otro dato a resaltar, al observar el diagrama de cajas, es que el grupo «NN-M2-F» es el que muestra una diferencia algo mayor entre media y mediana respecto a los otros dos grupos, como se aprecia en la tabla de estadísticos y en el gráfico de cajas. No obstante, podemos observar que ambos valores están muy próximos en las tres muestras. Esto indica que la tendencia en los tres casos es a una distribución normal o campana de Gauss.

Cómo evolucionan los estadísticos «mínimo» y «máximo» entre las muestras del grupo de NN entre el M1 y el M2 es un dato que hay que tener en consideración. El grupo

«NN-M2-F» presenta un valor del estadístico «mínimo» inferior al grupo «NN-M1-F»; sin embargo, en el primero hay valores del estadístico «máximo» muy superiores a la muestra del grupo de NN del M1 e incluso superiores a la muestra del grupo de Nativos. Por otro lado, también observamos que el estadístico «mínimo» de la muestra del grupo de Nativos es sensiblemente superior a las otras dos muestras.

También llama la atención el estadístico curtosis. Vemos que en las muestras del grupo de NN este valor es positivo, mientras que en la muestra del grupo de Nativos, es negativo. Recordemos que este valor hace referencia a la forma de la curva. En este caso, la curtosis de las muestras del grupo de NN va a ser más aguda que la de la muestra de N, pues los valores negativos tienden a tener menor número de valores extremos y, por tanto, la forma es más plana. Esto lo podemos observar en los histogramas correspondientes.

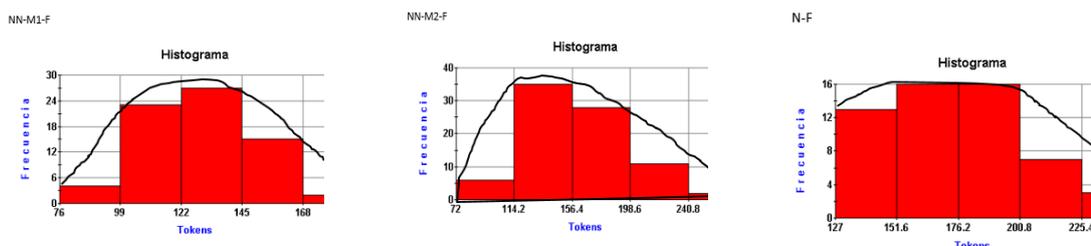


Gráfico 102. Histograma para la variable «Tokens» NN-M1-F / NN-M2-F / N-F

(12) *Types*: Es el número de palabras diferentes.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
<i>Types</i>	6.157	8.592	6.341

Como ya hemos indicado, los valores de las medidas de centralización y de dispersión ofrecen una idea básica de cómo se distribuye la muestra. Como se puede observar en la tabla de estadísticos (Tabla 54), las diferencias en las medias de las dos muestras del grupo de NN son bastante llamativas. Mientras que para el grupo «NN-M1-F» (grupo no nativo, momento 1) la media es 87.7183, para el grupo «NN-M2-F» (grupo no nativo, momento 2) la media es 104.7805. También la desviación típica es muy superior en el caso del grupo de NN en el M2 ( $s=12.6595$  /  $s=20.7382$ , respectivamente). Esta distribución contrasta con la del grupo de nativos que aun mostrando una media superior a la del grupo «NN-M2-F», su desviación típica es inferior ( $s=17.5875$ ).

Estadísticos para la variable Types

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>86.7183</b>	<b>104.7805</b>	<b>115.2909</b>
<b>Mediana</b>	<b>87.0000</b>	<b>107.0000</b>	<b>115.0000</b>
<b>Varianza</b>	<b>160.2624</b>	<b>430.0747</b>	<b>309.3212</b>
<b>Desviación Típica</b>	<b>12.6595</b>	<b>20.7382</b>	<b>17.5875</b>
<b>Mínimo</b>	<b>56.0000</b>	<b>51.0000</b>	<b>82.0000</b>
<b>Máximo</b>	<b>115.0000</b>	<b>153.0000</b>	<b>156.0000</b>
<b>Cuartil Inferior</b>	<b>77.0000</b>	<b>90.0000</b>	<b>104.0000</b>
<b>Cuartil Superior</b>	<b>97.0000</b>	<b>118.0000</b>	<b>124.0000</b>
<b>Rango Intercuartilico</b>	<b>20.0000</b>	<b>28.0000</b>	<b>20.0000</b>
<b>Curtosis</b>	<b>-0.4210</b>	<b>0.0675</b>	<b>-0.1796</b>

Tabla 54. Estadísticos para la variable «Types» NN-M1-F / NN-M2-F / N-F

En el gráfico de dispersión se observa claramente lo que acabamos de comentar. Mientras que en el grupo «NN-M1-F» los valores quedan más concentrados en torno a la media y mediana, en el grupo «NN-M2-F» la dispersión es mayor.

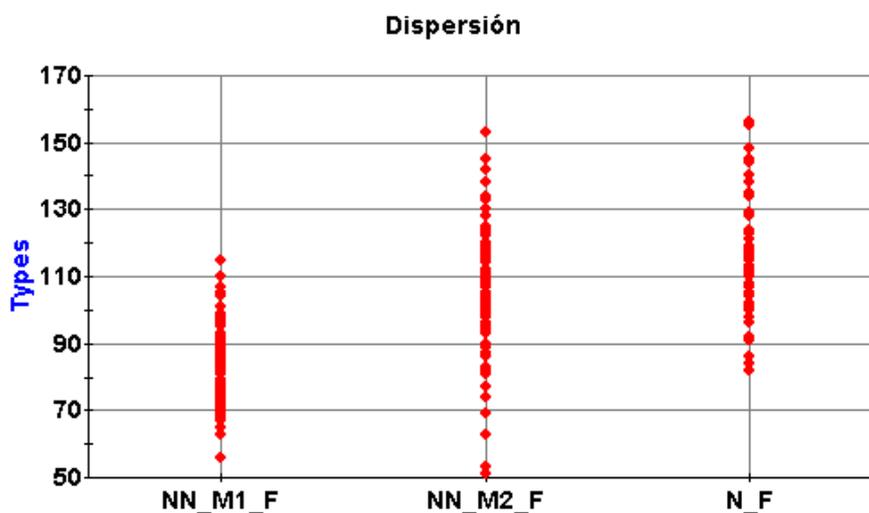


Gráfico 103. Gráfico de dispersión para la variable «Types» NN-M1-F / NN-M2-F / N-F

A continuación mostramos el diagrama de cajas para ver gráficamente cómo quedan representados los estadísticos.

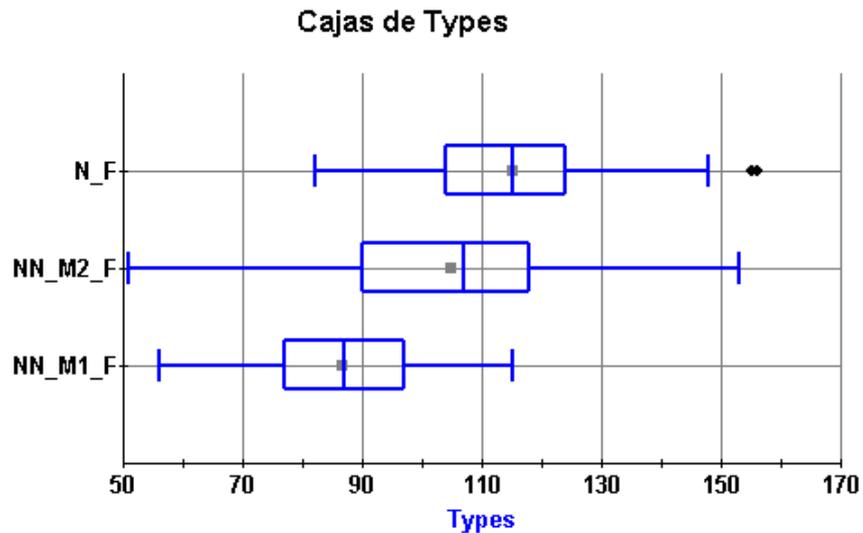


Gráfico 104. Diagrama de Cajas para la variable «Types» NN-M1-F / NN-M2-F / N-F

Al observar el diagrama de cajas, llama la atención la casi coincidencia de la media y la mediana en la muestra del grupo de Nativos y la del grupo de no Nativos en el M1, mientras que en el grupo «NN-M2-F» (grupo no Nivo, momento 2), al igual que sucede con la variable «Tokens», presenta una diferencia algo mayor entre la media y la mediana.

El valor del rango intercuartílico también nos habla de la dispersión de los datos y, como se ve en el diagrama de cajas (Gráfico 104), y coincidiendo con la variable «Tokens», el de la muestra del grupo «NN-M2-F» es mayor al de las otras dos muestras. En este caso, llama la atención que el rango intercuartílico del grupo de no Nativos M1 y el de Nativos sean exactamente iguales.

En esta variable, y a pesar de la mayor dispersión de los datos en el grupo «NN-M2-F», solo en el grupo de Nativos vemos datos que quedan fuera de los «bigotes».

En cuanto a la forma, el estadístico de la «curtosis» presenta valores negativos en las muestras del grupo de Nativos y en el grupo de no Nativos M1, mientras que el valor en el grupo de no Nativos M2 es positivo. Así, como se puede apreciar en el histograma que se muestra a continuación, la forma de la muestra del grupo «NN-M2-F» es más aguda que la de las otras dos muestras.

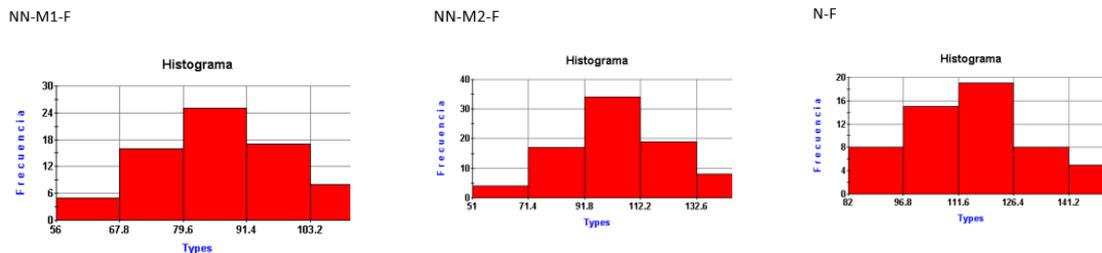


Gráfico 105. Histograma para la variable «Types» NN-M1-F / NN-M2-F / N-F

(13) Verbos: Número total de verbos.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Verbos	1.686	2.683	1.546

A partir de los valores absolutos, vamos a ver qué dicen los estadísticos y ver de qué manera se centralizan los datos, qué dispersión presentan y qué forma tienen.

**Estadísticos para la variable Verbos**

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>36.3380</b>	<b>46.6098</b>	<b>41.4000</b>
<b>Mediana</b>	<b>36.0000</b>	<b>46.5000</b>	<b>41.0000</b>
<b>Varianza</b>	<b>71.7412</b>	<b>200.4878</b>	<b>68.0593</b>
<b>Desviación Tipica</b>	<b>8.4700</b>	<b>14.1594</b>	<b>8.2498</b>
<b>Mínimo</b>	<b>22.0000</b>	<b>17.0000</b>	<b>28.0000</b>
<b>Máximo</b>	<b>62.0000</b>	<b>95.0000</b>	<b>65.0000</b>
<b>Cuartil Inferior</b>	<b>29.0000</b>	<b>37.0000</b>	<b>36.0000</b>
<b>Cuartil Superior</b>	<b>41.0000</b>	<b>54.0000</b>	<b>44.0000</b>
<b>Rango Intercuartilico</b>	<b>12.0000</b>	<b>17.0000</b>	<b>8.0000</b>
<b>Curtosis</b>	<b>0.2490</b>	<b>0.7374</b>	<b>0.7899</b>

Tabla 55. Estadísticos para la variable «Verbos» NN-M1-F / NN-M2-F / N-F

Lo primero que llama la atención de las medidas de centralización es que los valores de los estadísticos de la media y la mediana están muy próximos entre sí en cada una de las tres muestras analizadas. Pero una vez más, es el valor de la muestra del grupo «NN-M2-F» el que supera los valores de las otras dos muestras. Vemos que sobrepasa en más de diez

puntos los valores de la media y la mediana del grupo de no Nativos del M1, y en seis los valores del grupo de Nativos.

En cuanto a la desviación típica, mientras que los valores del grupo de no Nativos M1 ( $s=8.4700$ ) y el grupo de Nativos ( $s=8.2498$ ) están muy próximos, el valor que presenta el grupo de no Nativos M2 es muy superior ( $s=14.1594$ ). Esto lo podemos visualizar claramente en el gráfico de dispersión.

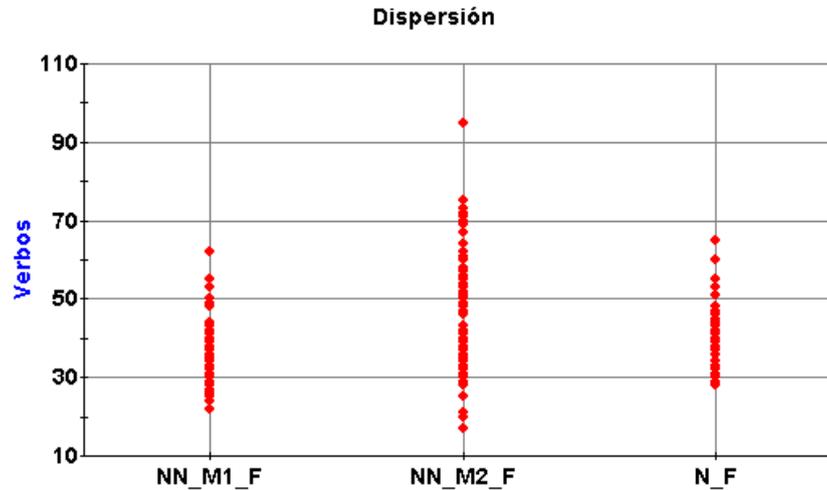


Gráfico 106. Gráfico de dispersión para la variable «Verbos» NN-M1-F / NN-M2-F / N-F

También en el diagrama de cajas (Gráfico 107) podemos comprobar ese grado de dispersión mayor que presenta el grupo «NN-M2-F».

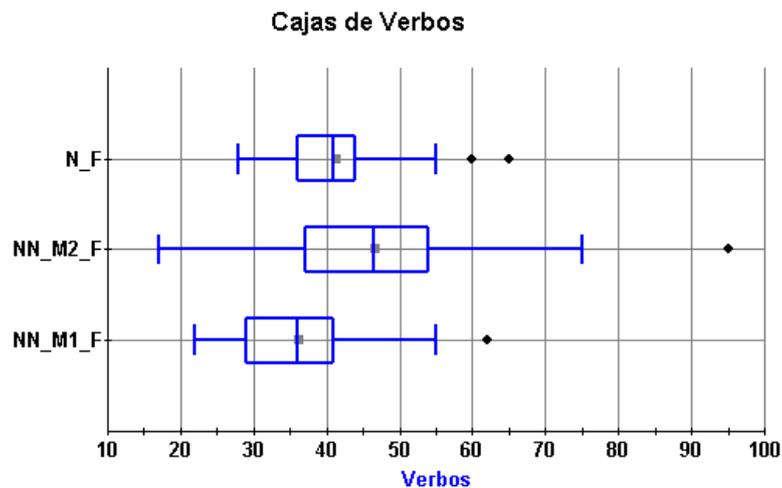


Gráfico 107. Diagrama de cajas para la variable «Verbos» NN-M1-F / NN-M2-F / N-F

La otra medida de dispersión que nos muestra la tabla de estadísticos es el rango intercuartílico. Podemos ver claramente en el diagrama de cajas que el tamaño de la caja del grupo de no Nativos M2 es mayor al de los otros dos grupos. En esta ocasión es el grupo de Nativos el que presenta la mayor concentración de datos en torno a la media y mediana con un valor  $IQR=8.0000$  inferior al del grupo «NN-M1-F» ( $IQR=12.0000$ ) y al grupo «NN-M2-F» ( $IQR=17.0000$ ).

También observamos en el diagrama de cajas los datos extremos que presentan las tres muestras. Recordemos que esos datos se sitúan fuera de los «bigotes» de la caja pues van más allá del valor de  $1.5 IQR$ . Así, vemos que el grupo de Nativos muestra dos datos (marcados con dos puntos) que quedan fuera de esa medida. También el grupo de no Nativos, tanto en el M1 como en el M2, presenta datos fuera de los «bigotes» de la caja, aunque es el grupo «NN-M2-F» el que tiene el dato más alejado del estadístico «máximo».

En cuanto al estadístico referido a la forma, la curtosis, tenemos que decir que presenta valores positivos en las tres muestras. Así, igual que en la variable anterior, estamos ante una forma más aguda de la curva que si fueran valores negativos. Los histogramas que mostramos a continuación pueden dar una idea de cómo es la curva con estos valores positivos de «curtosis».

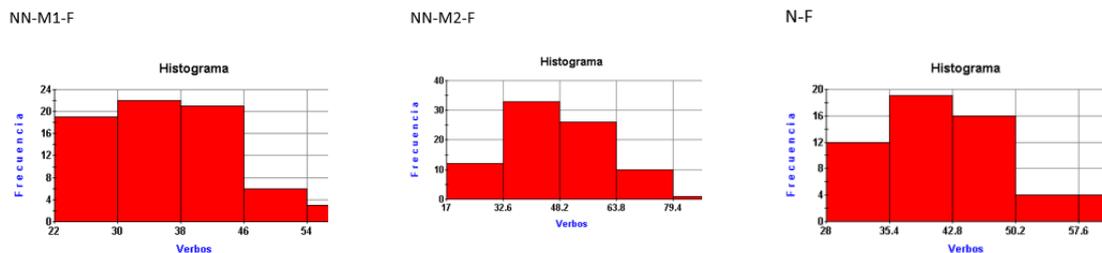


Gráfico 108. Histograma para la variable «Verbos» NN-M1-F / NN-M2-F / N-F

(14) Sustantivos: Número total de sustantivos.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Sustantivos	3.920	5.467	3.998

Vemos una vez más que el número total de sustantivos en las muestras de no Nativos M1 y de Nativos están próximos, aunque hay que tener presente la diferencia en cuanto al total de sujetos de cada una de las muestras. Comprobamos, asimismo, que el número total de sustantivos del grupo de no Nativos M2 es muy superior a las otras dos muestras. Con estos datos de valores absolutos, vamos a ver qué dicen los estadísticos (Tabla 56).

Al observar las medidas de centralización, vemos que los valores de la media y la mediana del grupo de «NN-M2-F» (no Nativos, M2) y el de Nativos, «N-F», están muy próximos, por encima de ochenta en ambos casos. Sin embargo, el grupo de no Nativos en el M1 presenta un valor de media=68.5493 y de mediana=67.0000 muy por debajo de las otras dos muestras.

**Estadísticos para la variable Sustantivos**

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>68.5493</b>	<b>82.0122</b>	<b>87.3273</b>
<b>Mediana</b>	<b>67.0000</b>	<b>82.5000</b>	<b>89.0000</b>
<b>Varianza</b>	<b>182.9082</b>	<b>415.6912</b>	<b>238.7057</b>
<b>Desviación Típica</b>	<b>13.5244</b>	<b>20.3885</b>	<b>15.4501</b>
<b>Mínimo</b>	<b>33.0000</b>	<b>38.0000</b>	<b>56.0000</b>
<b>Máximo</b>	<b>100.0000</b>	<b>132.0000</b>	<b>132.0000</b>
<b>Cuartil Inferior</b>	<b>60.0000</b>	<b>66.0000</b>	<b>77.0000</b>
<b>Cuartil Superior</b>	<b>76.0000</b>	<b>97.0000</b>	<b>97.0000</b>
<b>Rango Intercuartílico</b>	<b>16.0000</b>	<b>31.0000</b>	<b>20.0000</b>
<b>Curtosis</b>	<b>-0.1135</b>	<b>-0.1062</b>	<b>0.5846</b>

Tabla 56. Estadísticos para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F

Hay que llamar la atención sobre la proximidad entre los valores de media y mediana que hay en cada una de las muestras.

Si observamos los valores de los estadísticos que hacen referencia a la dispersión, vemos que es la muestra del grupo «NN-M2-F» (no Nativos, M2) el que presenta una mayor desviación típica,  $s=20.3885$ , frente a  $s=13.5244$  para el grupo de no Nativos M1 y

s=15.0000 para el grupo de Nativos. A continuación podemos visualizar estos datos en el gráfico de dispersión.

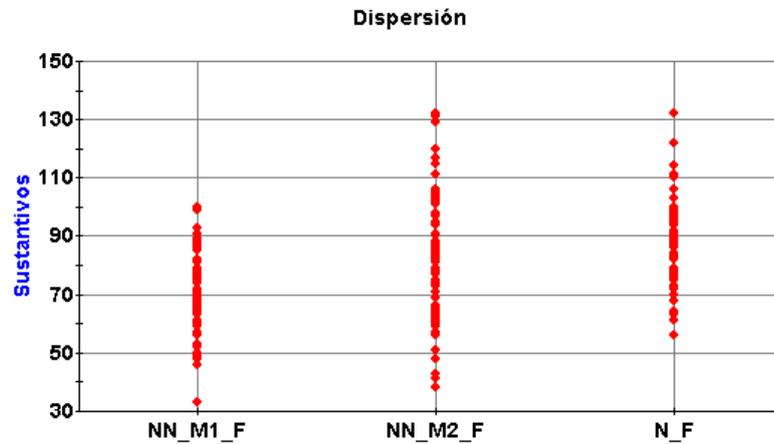


Gráfico 109. Gráfico de dispersión para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F

El otro descriptor que informa de la dispersión, el rango intercuartílico, también presenta valores muy diferentes para cada una de las tres muestras, como podemos comprobar en la tabla de estadísticos. Nuevamente, es en la muestra del grupo «NN-M2-F» (no Nativos, M2) donde encontramos el valor más alto, con un IQR=31.0000, frente a un IQR=16.0000 para el grupo de no Nativos M1, y un IQR=20.0000 para el grupo de Nativos. Recordemos que este valor hace referencia al tamaño de la caja y que en ella quedan concentrados el 50% de los valores de la muestra. Así, comprobamos, una vez más, que es la muestra del grupo de no Nativos M2 el que presenta una menor concentración de los datos en torno a las medidas de centralización. En el siguiente gráfico de cajas podremos visualizar más claramente toda esta información.

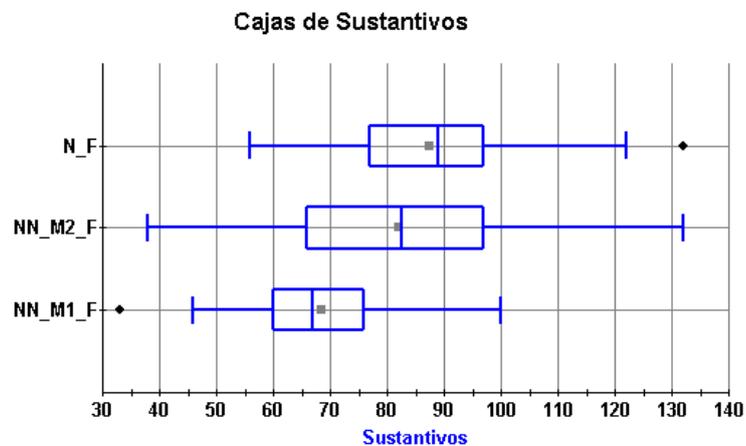


Gráfico 110. Diagrama de cajas para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F

Como se ve en el gráfico, la caja correspondiente a la muestra «NN-M2-F» es sensiblemente más grande que las otras dos, y llama la atención que es casi el doble de grande que la otra muestra de los no Nativos en el M1.

Otro elemento de dispersión que llama la atención son los valores extremos (marcados con puntos negros) que aparecen en las muestras de los no Nativos en el M1 y en el de los Nativos. Mientras que en el grupo «NN-M1-F» este valor está más allá del descriptor «mínimo», en el grupo «N-F» el valor extremo aparece más allá del descriptor «máximo».

En cuanto a la forma que adopta la curva de la muestra viene dada por el valor del estadístico curtosis. Vemos en la tabla de estadísticos que este valor es negativo en el grupo de no Nativos, tanto en el M1 como en el M2, mientras que en el grupo de Nativos este valor es positivo. Es decir, que las dos muestras del grupo de no Nativos van a presentar una forma más plana que la de la muestra del grupo de Nativos. En los histogramas siguientes podemos ver esa tendencia.

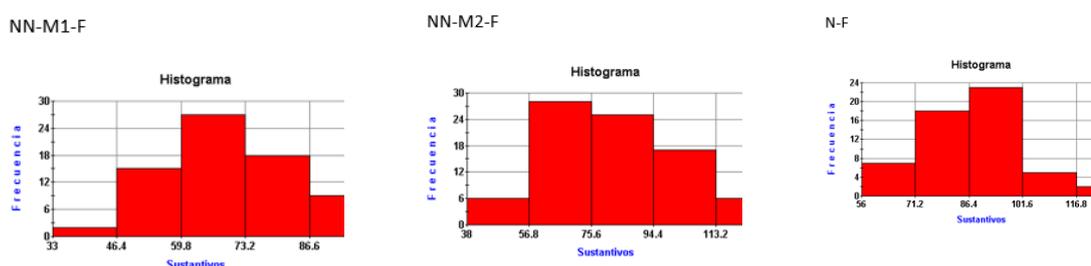


Gráfico 111. Histograma para la variable «Sustantivos» NN-M1-F / NN-M2-F / N-F

(15) Adjetivos: Número total de adjetivos.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Adjetivos	1.285	1.694	1.182

A continuación vamos a ver qué dicen los estadísticos respecto a la variable de adjetivos. Como vemos en el número total, es la muestra del grupo de Nativos el que presenta menor número total de adjetivos.

**Estadísticos para la variable Adjetivos**

<b>Grupos</b>	<b>NN_M1_F</b>	<b>NN_M2_F</b>	<b>N_F</b>
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>30.7042</b>	<b>36.7561</b>	<b>36.1636</b>
<b>Mediana</b>	<b>31.0000</b>	<b>35.5000</b>	<b>35.0000</b>
<b>Varianza</b>	<b>62.8113</b>	<b>143.5447</b>	<b>74.6579</b>
<b>Desviación Típica</b>	<b>7.9254</b>	<b>11.9810</b>	<b>8.6405</b>
<b>Mínimo</b>	<b>15.0000</b>	<b>13.0000</b>	<b>20.0000</b>
<b>Máximo</b>	<b>46.0000</b>	<b>76.0000</b>	<b>72.0000</b>
<b>Cuartil Inferior</b>	<b>26.0000</b>	<b>28.0000</b>	<b>31.0000</b>
<b>Cuartil Superior</b>	<b>37.0000</b>	<b>44.0000</b>	<b>41.0000</b>
<b>Rango Intercuartílico</b>	<b>11.0000</b>	<b>16.0000</b>	<b>10.0000</b>
<b>Curtosis</b>	<b>-0.5564</b>	<b>0.3587</b>	<b>4.5193</b>

Tabla 57. Estadísticos para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F

En la tabla podemos observar que los estadísticos de centralización, la media y la mediana, están muy próximos en las muestras del grupo de no Nativos M2 (media=36.7561; mediana=35.5000) y del grupo de Nativos (media=36.1636; mediana=35.0000), a pesar de la gran diferencia en el número total de adjetivos que presenta cada una de las muestras. En cambio la muestra del grupo de no Nativos M1 presenta un valor de media=30.7042 y mediana=31.0000 inferior a las otras dos muestras, incluso es inferior a la muestra de Nativos aunque ésta presenta un total de adjetivos menor a la de los no Nativos M1.

El valor más bajo respecto al estadístico de dispersión «desviación típica» es el de la muestra del grupo «NN-M1-F», con un valor de  $s=7.9254$ , seguido por la muestra del grupo de Nativos,  $s=8.6405$ , y el valor más alto de «desviación típica» es, una vez más, la muestra del grupo de no Nativos M2,  $s=11.9810$ . En el gráfico de dispersión (Gráfico 112) podemos constatar estos datos.

En el diagrama de cajas (Gráfico 113) tenemos otra perspectiva de la dispersión de las muestras. Observamos que se dan valores extremos en las muestras de Nativos y en las muestras de no Nativos M2. También podemos ver que el tamaño de la caja del grupo «NN-M2-F» es mayor, con un valor de  $IQR=16.0000$ , que las de las otras dos muestras:  $IQR=11.0000$  para el grupo «NN-M1-F»;  $IQR=10.0000$  para el grupo «N-F».

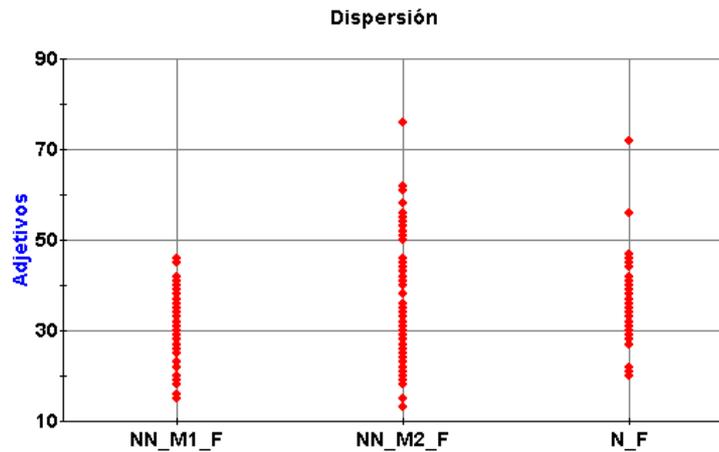


Gráfico 112. Gráfico de dispersión para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F

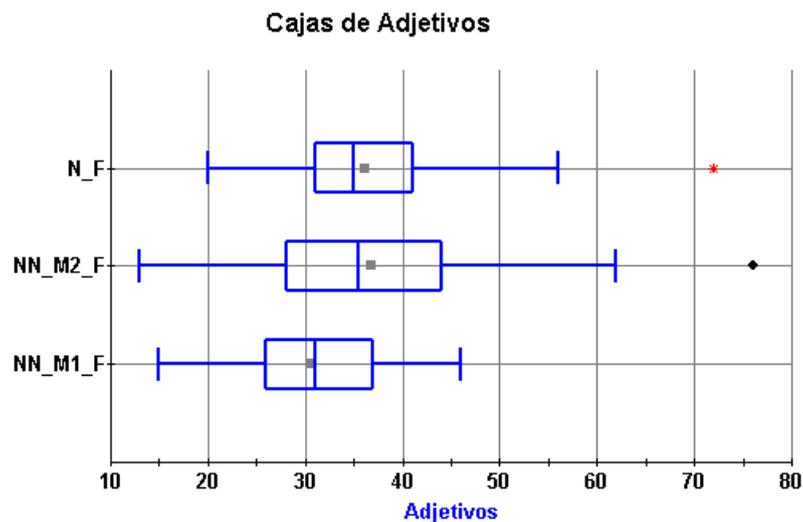


Gráfico 113. Diagrama de cajas para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F

A continuación mostramos los histogramas de las tres muestras que dan una idea de la tendencia de la curva a ser más plana o más aguda. Esa tendencia viene marcada por el estadístico curtosis que marca valores positivos para las muestras de los grupos «NN-M2-F» (curtosis=0.3587) y «N-F» (curtosis=4.5193), es decir, curva más aguda; mientras que el valor del grupo «NN-M2-F» es negativo (curtosis= -0.5564), tendencia a una curva más plana.

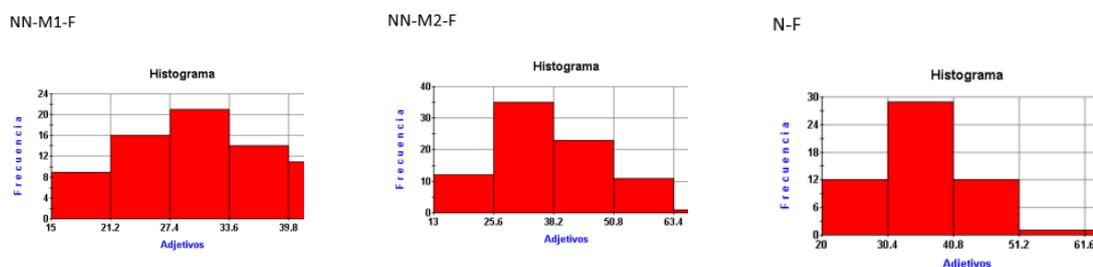


Gráfico 114. Histograma para la variable «Adjetivos» NN-M1-F / NN-M2-F / N-F

(16) Adverbios: Número total de adverbios

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Adverbios	566	771	376

Esta variable, como podemos ver, no presenta tantas diferencias en el número total como las variables anteriores. A continuación mostramos la tabla de estadísticos para ver qué información nos da sobre los datos que tenemos y comprobar si esas diferencias mínimas se constatan o no.

**Estadísticos para la variable Adverbios**

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>19.1549</b>	<b>23.4756</b>	<b>21.4000</b>
<b>Mediana</b>	<b>19.0000</b>	<b>22.0000</b>	<b>21.0000</b>
<b>Varianza</b>	<b>29.0757</b>	<b>68.1043</b>	<b>27.0593</b>
<b>Desviación Típica</b>	<b>5.3922</b>	<b>8.2525</b>	<b>5.2019</b>
<b>Mínimo</b>	<b>8.0000</b>	<b>7.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>36.0000</b>	<b>49.0000</b>	<b>36.0000</b>
<b>Cuartil Inferior</b>	<b>15.0000</b>	<b>18.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>23.0000</b>	<b>29.0000</b>	<b>24.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>11.0000</b>	<b>6.0000</b>
<b>Curtosis</b>	<b>0.3481</b>	<b>0.6718</b>	<b>0.3006</b>

Tabla 58. Estadísticos para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F

Vemos que los valores de centralización, la media y la mediana, de las tres muestras están próximos, aunque mantienen unas pequeñas diferencias. Igual que en las variables anteriores, el valor de la media y mediana de la muestra del grupo «NN-M2-F» es superior al de las otras dos muestras. También hay que destacar que los valores de la media y mediana

del grupo «N-F» son superiores a los del grupo «NN-M1-F», aunque este presenta un número total de adverbios mayor.

En cuanto a la dispersión de las muestras, vemos que los valores del grupo «NN-M1-F» (s=5.3922) y del grupo «N-F» (s=5.2019) están muy cercanos. Nuevamente, es el grupo «NN-M2-F» el que presenta una mayor dispersión de sus datos como podemos comprobar en el gráfico de dispersión (Gráfico 115).

También vemos esa mayor dispersión de la muestra del grupo «NN-M2-F» en el valor del rango intercuartílico, IQR=11.0000, frente a IQR=8.0000 para el grupo «NN-M1-F», y IQR=6.0000 para el grupo «N-F». Estos valores los podemos visualizar más claramente en el diagrama de cajas (Gráfico 116).

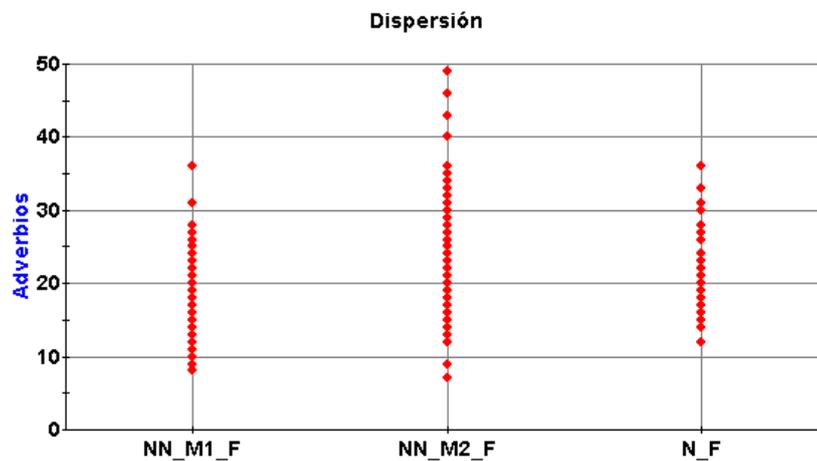


Gráfico 115. Gráfico de dispersión para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F

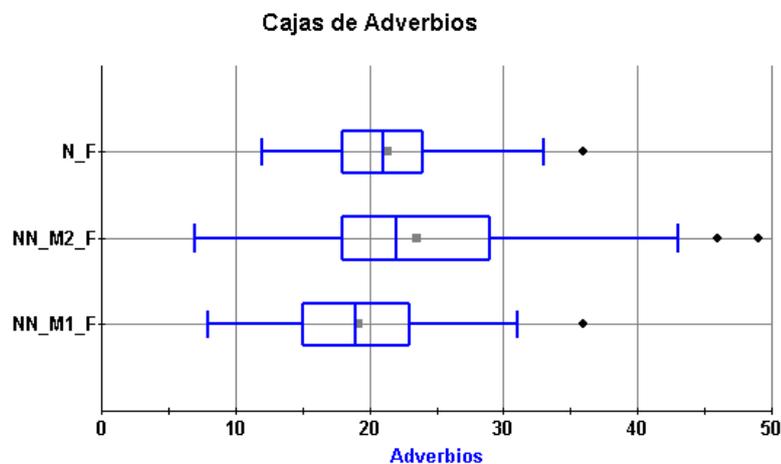


Gráfico 116. Diagrama de cajas para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F

Con el diagrama de cajas podemos visualizar las diferencias en los valores del rango intercuartílico de las tres muestras. Así, constatamos que es el grupo de Nativos el que mayor concentración de datos presenta en torno a las medidas de centralización.

También podemos observar en el diagrama de cajas, por un lado, los valores extremos que presentan las tres muestras y, por otro lado, la diferencia entre la mediana y la media de la muestra del grupo «NN-M2-F». Recordemos que la mediana divide la muestra en dos mitades iguales, mientras que la media es el resultado de dividir todos los valores por el total de la muestra, de ahí que sea más vulnerable a los datos extremos. El hecho de que estas dos medidas no coincidan indica que la distribución se aleja de una distribución normal o de campana de Gauss.

En esta variable, el valor del estadístico curtosis es positivo en las tres muestras, dando así una curva con tendencia más aguda que la normal o curva de Gauss. En los histogramas siguientes se puede percibir esa tendencia.

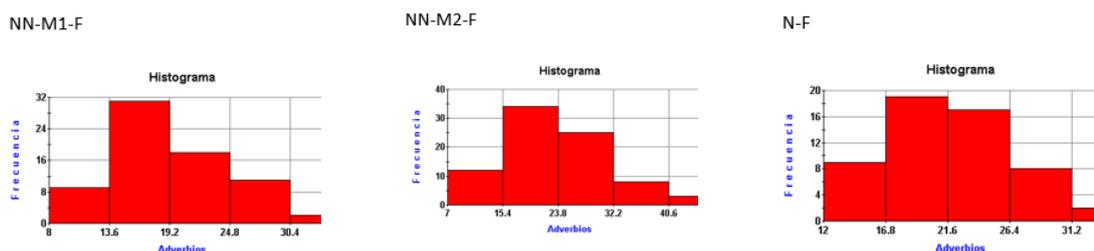


Gráfico 117. Histograma para la variable «Adverbios» NN-M1-F / NN-M2-F / N-F

(17) Pronombres: Número total de pronombres.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Pronombres	611	1.092	981

A partir del número total de pronombres por muestra, vamos a ver qué dicen los estadísticos sobre la distribución de las mismas.

Las medidas de centralización, la media y la mediana, presentan valores muy diferentes para cada muestra. En esta variable, es el grupo de Nativos el que tiene el valor de la media y mediana más alto, aun teniendo un número total de pronombres inferior al del grupo de no Nativos M2.

En cuanto al descriptor para indicar la dispersión de la muestra, la desviación típica, vemos que aun por poca diferencia, pero es la muestra del grupo «NN-M2-F» la que tiene un valor más alto ( $s=9.5264$ ), frente al grupo N que presenta un valor de  $s=9.1240$  y a la muestra del grupo no N M1 con un valor  $s=5.9718$ . Así, al igual que en las variables anteriores, es la muestra del grupo de no Nativos M2 el que presenta más dispersión de los datos. En el gráfico 118 de dispersión podemos visualizar esta información.

**Estadísticos para la variable Pronombres**

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>21.7746</b>	<b>29.8049</b>	<b>33.4182</b>
<b>Mediana</b>	<b>20.0000</b>	<b>29.0000</b>	<b>33.0000</b>
<b>Varianza</b>	<b>35.6628</b>	<b>90.7516</b>	<b>83.2478</b>
<b>Desviación Típica</b>	<b>5.9718</b>	<b>9.5264</b>	<b>9.1240</b>
<b>Mínimo</b>	<b>10.0000</b>	<b>12.0000</b>	<b>12.0000</b>
<b>Máximo</b>	<b>38.0000</b>	<b>64.0000</b>	<b>56.0000</b>
<b>Cuartil Inferior</b>	<b>17.0000</b>	<b>23.0000</b>	<b>27.0000</b>
<b>Cuartil Superior</b>	<b>26.0000</b>	<b>37.0000</b>	<b>39.0000</b>
<b>Rango Intercuartílico</b>	<b>9.0000</b>	<b>14.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.1927</b>	<b>1.1054</b>	<b>0.5550</b>

Tabla 59. Estadísticos para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F

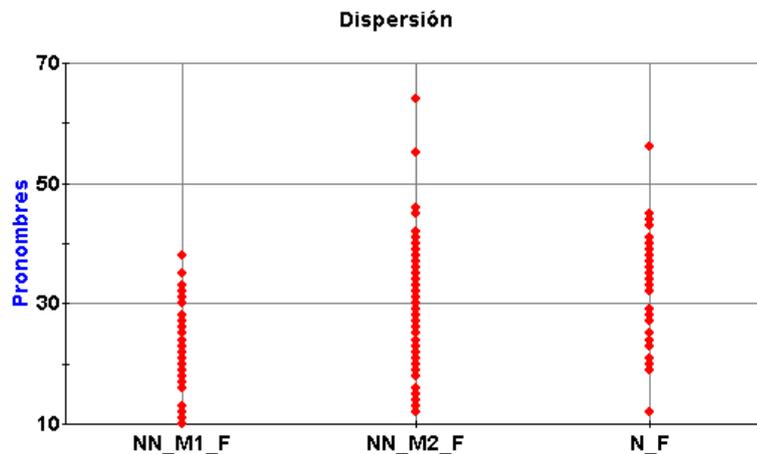


Gráfico 118. Gráfico de dispersión para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F

También el valor del rango intercuartílico, el otro descriptor que nos informa de la dispersión de los datos, presenta un valor superior en el grupo de no Nativos M2.

En el diagrama de cajas (Gráfico 119) queda representado en el tamaño de la caja el valor del rango intercuartílico. Podemos observar, por un lado, que la caja de la muestra del grupo «NN-M2-F» es algo más grande que la del grupo de Nativos, por tanto, hay mayor dispersión, es decir, menos centralización en torno a la media y la mediana. Por otro lado, constatamos las diferencias observadas en los valores de los estadísticos de la muestra del grupo de no Nativos M1 en la representación del diagrama de cajas para dichos valores. Mientras que el estadístico «mínimo» es casi el mismo para las tres muestras, en el «máximo» hay una gran diferencia entre la muestra del grupo de no Nativos M1 y las otras dos muestras, la de no Nativos M2 y la de Nativos. Éstas, muy próximas entre sí.

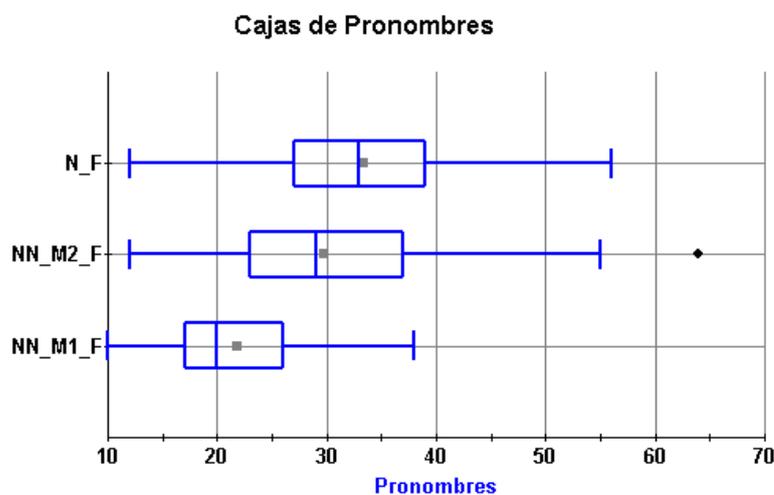
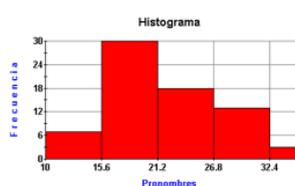


Gráfico 119. Diagrama de cajas para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F

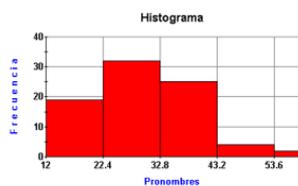
También llama la atención al observar el diagrama de cajas, la diferencia entre la media y la mediana de la muestra de no Nativos M1.

En cuanto a la forma de la distribución de los datos, el valor del estadístico curtosis marca un valor negativo para la muestra del grupo «NN-M1-F» y valor positivo para las otras dos muestras. En el primer caso, valor negativo, la tendencia es una curva más plana respecto a la distribución normal o campana de Gauss y, en el segundo caso, una curva con tendencia más aguda. En los histogramas siguientes se puede visualizar esa tendencia.

NN-M1-F



NN-M2-F



N-F

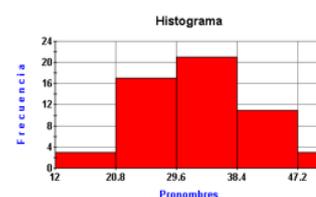


Gráfico 120. Histograma para la variable «Pronombres» NN-M1-F / NN-M2-F / N-F

(18) Preposiciones: Número total de preposiciones.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Preposiciones	404	577	695

Como en el resto de variables, es la muestra de «NN-M2-F» la que presenta el número total de preposiciones más alto. A continuación vamos a ver en la tabla de estadísticos cómo quedan distribuidos los datos.

#### Estadísticos para la variable Preposiciones

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>17.8028</b>	<b>22.8902</b>	<b>28.7273</b>
<b>Mediana</b>	<b>18.0000</b>	<b>22.0000</b>	<b>29.0000</b>
<b>Varianza</b>	<b>28.6463</b>	<b>57.8026</b>	<b>34.0539</b>
<b>Desviación Típica</b>	<b>5.3522</b>	<b>7.6028</b>	<b>5.8356</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>10.0000</b>	<b>18.0000</b>
<b>Máximo</b>	<b>32.0000</b>	<b>41.0000</b>	<b>40.0000</b>
<b>Cuartil Inferior</b>	<b>14.0000</b>	<b>18.0000</b>	<b>25.0000</b>
<b>Cuartil Superior</b>	<b>21.0000</b>	<b>28.0000</b>	<b>33.0000</b>
<b>Rango Intercuartílico</b>	<b>7.0000</b>	<b>10.0000</b>	<b>8.0000</b>
<b>Curtosis</b>	<b>-0.2313</b>	<b>-0.1598</b>	<b>-0.6398</b>

Tabla 60. Estadísticos para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F

Como podemos observar en la tabla de estadísticos, el valor más alto para el descriptor de media y mediana, medidas de centralización, es el de la muestra del grupo de Nativos. El siguiente valor es para la muestra del grupo de no Nativos M2 y el valor más bajo es el de la muestra del grupo de no Nativos M1. Asimismo, comprobamos que las diferencias entre la media y la mediana en cada una de las muestras es mínima. La mayor diferencia, aunque

pequeña también, la encontramos en la muestra del grupo «NN-M2-F». En el diagrama de cajas podemos visualizar esas pequeñas diferencias (Gráfico 122).

En cuanto a las medidas de dispersión, observamos que el valor más alto del descriptor «desviación típica» lo volvemos a encontrar en la muestra del grupo «NN-M2-F»,  $s=7.6028$ , en comparación con la muestra del mismo grupo no Nivo en el M1,  $s=5.3522$ , y con la muestra del grupo Nivo,  $s=5.8356$ . En el gráfico de dispersión podemos visualizar estos datos (Gráfico 121).

El valor más alto del otro estadístico que informa de la dispersión de la muestra, el rango intercuartílico, también corresponde a la muestra del grupo «NN-M2-F» con un valor de  $IQR=10.0000$ , frente a un  $IQR=7.0000$  para el grupo de no Nativos en el M1 y un  $IQR=8.0000$  para el grupo de Nativos. Estas pequeñas diferencias las podemos constatar en el diagrama de cajas (Gráfico 122).

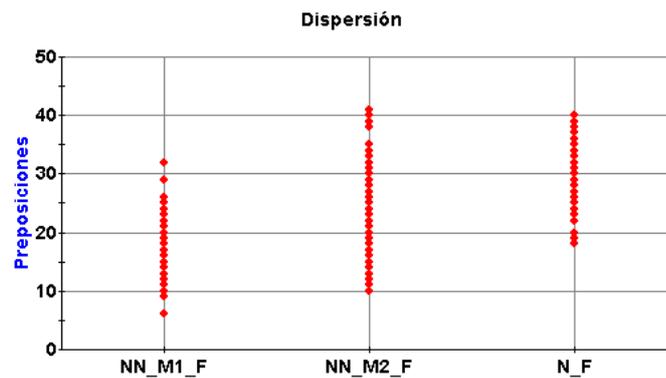


Gráfico 121. Gráfico de dispersión para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F

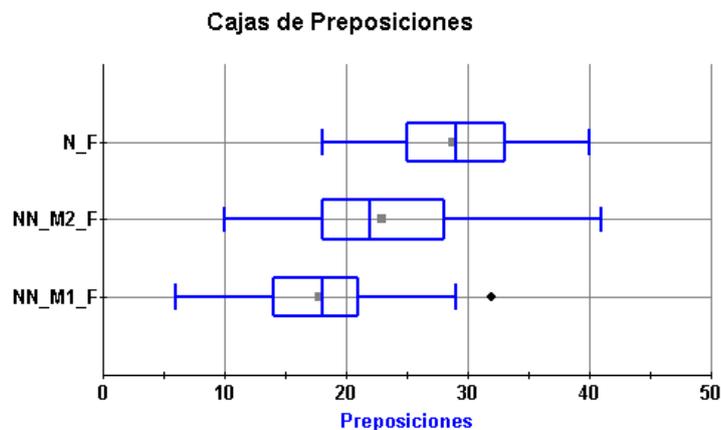


Gráfico 122. Diagrama de cajas para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F

Como podemos ver en los dos gráficos superiores, y como ya hemos indicado, es la muestra del grupo de no Nativos M2 el que presenta mayor dispersión de los datos.

Al observar los diagramas de cajas de las tres muestras, llama la atención los estadísticos de «mínimo» y «máximo». Por un lado, podemos ver que el «mínimo» del grupo de no Nativos en el M1 y en el M2 están relativamente cerca, sin embargo, el «máximo» de la muestra del M2 se aleja bastante del «máximo» del M1. Por otro lado, llama la atención que el «mínimo» del grupo de Nativos presenta el valor más alto y se aleja de los valores de las dos muestras del grupo de no Nativos, sin embargo, el «máximo» está cercano a la muestra del M2 del grupo de no Nativos.

Respecto a la forma que adopta cada una de las muestras, vemos en la tabla de estadísticos (Gráfico 60) que la curtosis es negativa en los tres casos. Eso nos indica una tendencia a una curva más plana en relación con la distribución normal o campana de Gauss. En los histogramas siguientes podemos visualizar esa tendencia.

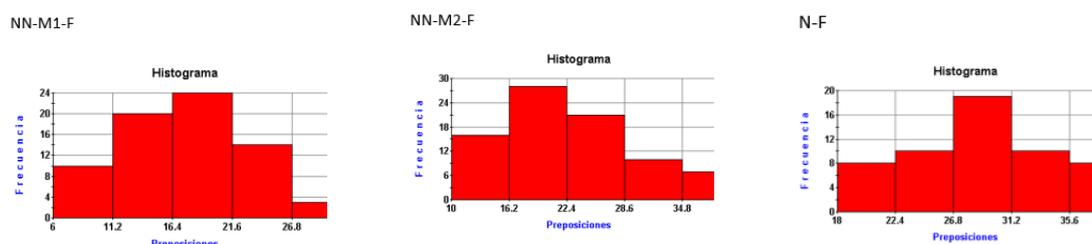


Gráfico 123. Histograma para la variable «Preposiciones» NN-M1-F / NN-M2-F / N-F

(19) Artículos: Número total de artículos.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Artículos	161	230	218

A continuación mostramos la tabla de estadísticos para obtener la información básica sobre las medidas de centralización, dispersión y forma de los datos de la variable «Artículos».

Estadísticos para la variable Artículos

Grupos	NN_M1_F	NN_M2_F	N_F
N	71	82	55
Media	11.5915	16.2561	18.1455
Mediana	12.0000	16.0000	18.0000
Varianza	18.8736	26.1188	18.2377
Desviación Típica	4.3444	5.1107	4.2706
Mínimo	3.0000	5.0000	10.0000
Máximo	22.0000	29.0000	31.0000
Cuartil Inferior	8.0000	12.0000	16.0000
Cuartil Superior	15.0000	19.0000	20.0000
Rango Intercuartílico	7.0000	7.0000	4.0000
Curtosis	-0.6485	0.1068	0.3607

Tabla 61. Estadísticos para la variable «Artículos» NN-M1-F / NN-M2-F / N-F

Como vemos en la tabla superior, en esta variable, el valor más alto de la media y la mediana la encontramos en la muestra del grupo de Nativos. El grupo de no Nativos presenta en el descriptor «media» una diferencia de casi cinco puntos entre el M1 y el M2. En cambio, entre el grupo «NN-M2-F» y el grupo «N-F» la diferencia es solo de dos puntos.

En el descriptor «desviación típica», las semejanzas las encontramos entre las muestras del grupo «NN-M1-F»,  $s=4.3444$ , y el grupo de Nativos,  $s=4.2706$ . El grupo «NN-M2-F» supera a las otras dos muestras, aunque por menos de un punto. Aun así, una vez más, es este grupo el que presenta mayor dispersión de los datos, aunque es menor que en otras variables. En el gráfico de dispersión se puede visualizar esta información.

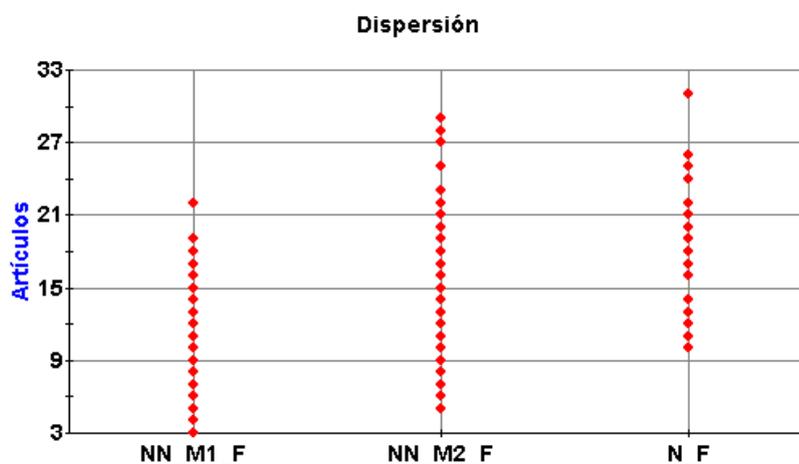


Gráfico 124. Gráfico de dispersión para la variable «Artículos» NN-M1-F / NN-M2-F / N-F

Llama la atención la coincidencia del valor del rango intercuartílico entre el M1 y el M2 del grupo de no Nativos,  $IQR=7.0000$ . En la muestra del grupo de Nativos, el valor de  $IQR=4.0000$ . Así, vemos que esta muestra presenta una mayor concentración de los datos en torno a la media y la mediana, medidas de centralización, que las dos muestras del grupo de no Nativos. Estos datos los vamos a visualizar en el diagrama de cajas (Gráfico 125).

Cuando observamos los diagramas de cajas de las muestras del grupo de no Nativos en el M1 y en el M2, llama la atención el estadístico «máximo» por la diferencia que hay entre ellos, siete puntos, mientras que la diferencia en el estadístico «mínimo» es solo de dos puntos.

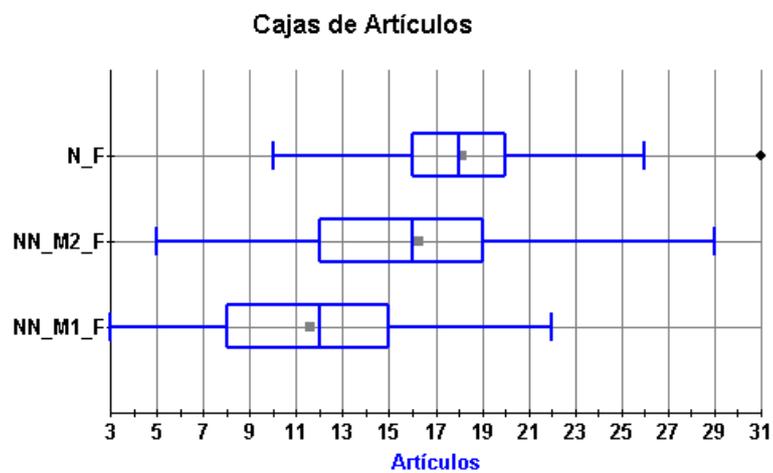


Gráfico 125. Diagrama de cajas para la variable «Artículos» NN-M1-F / NN-M2-F / N-F

Respecto a la forma de la distribución de las muestras, una curva más aguda o más plana en relación a la distribución normal o campana de Gauss, vemos que el valor del estadístico «curtosis», tanto en la muestra del grupo «NN-M2-F» como del grupo «N-F», es positivo, por tanto presenta una tendencia a una curva más aguda; mientras que el valor en la muestra del grupo «NN-M1-F» es negativa, por tanto, tiende a una curva más plana. En los siguientes histogramas podemos observar esa tendencia.

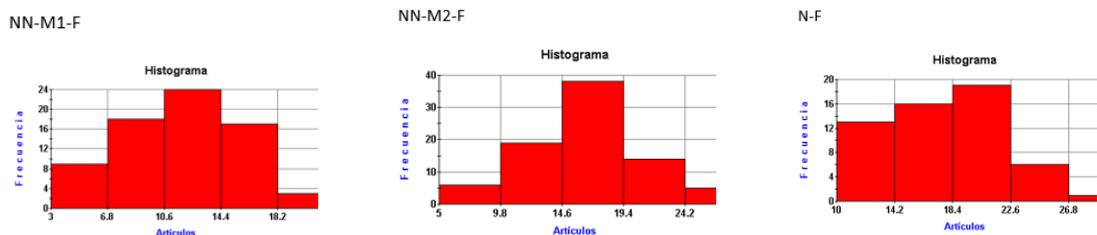


Gráfico 126. Histograma para la variable «Artículos» NN-M1-F / NN-M2-F / N-F

(20) Conjunciones: Número total de conjunciones.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Conjunciones	455	630	665

En la tabla 62 mostramos los estadísticos para obtener información sobre las medidas de centralización, dispersión y forma que proporcionan los datos de las tres muestras analizadas.

Los valores de las medidas de centralización de las tres muestras son muy diferentes. En esta variable, el valor más alto corresponde a la muestra del grupo de Nativos, como vemos en la tabla de estadísticos. En el grupo de no Nativos la diferencia entre el M1 y el M2 es de algo más de cinco puntos; mientras que la distancia entre el grupo «NN-M2-F» y el grupo «N-F» es de poco más de tres puntos.

En cuanto a la medida de dispersión, hay cierta coincidencia en el valor de la desviación típica de las muestras del grupo «NN-M1-F»,  $s=6.0928$ , y del grupo «N-F»,  $s=6.5965$ . La muestra del grupo «NN-M2-F» tiene un valor de  $s=8.8123$ , superando, así, en dos puntos la desviación típica de las otras dos muestras. Veamos en el gráfico de dispersión estos resultados.

**Estadísticos para la variable Conjunciones**

Grupos	NN_M1_F	NN_M2_F	N_F
<b>N</b>	<b>71</b>	<b>82</b>	<b>55</b>
<b>Media</b>	<b>18.8592</b>	<b>24.1463</b>	<b>27.3091</b>
<b>Mediana</b>	<b>18.0000</b>	<b>23.0000</b>	<b>27.0000</b>
<b>Varianza</b>	<b>37.1227</b>	<b>77.6573</b>	<b>43.5138</b>
<b>Desviación Típica</b>	<b>6.0928</b>	<b>8.8123</b>	<b>6.5965</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>5.0000</b>	<b>14.0000</b>
<b>Máximo</b>	<b>36.0000</b>	<b>68.0000</b>	<b>52.0000</b>
<b>Cuartil Inferior</b>	<b>15.0000</b>	<b>18.0000</b>	<b>23.0000</b>
<b>Cuartil Superior</b>	<b>23.0000</b>	<b>29.0000</b>	<b>32.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>11.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>0.5387</b>	<b>6.7171</b>	<b>2.4058</b>

Tabla 62. Estadísticos para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F

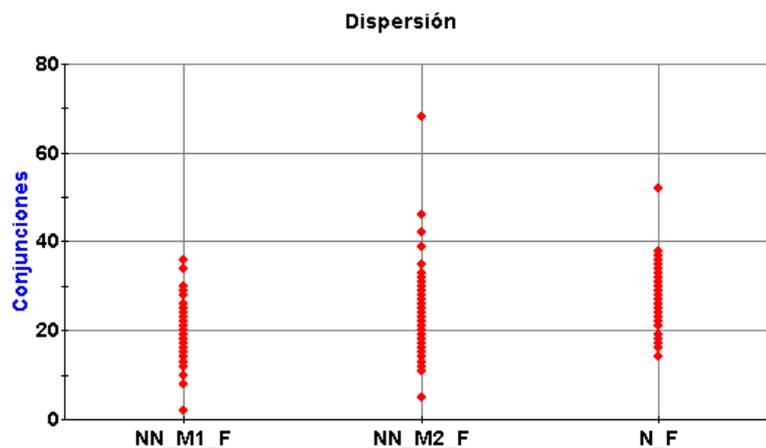


Gráfico 127. Gráfico de dispersión para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F

El valor más alto del rango intercuartílico, la otra medida de dispersión, corresponde al M2 del grupo de no Nativos,  $IQR=11.0000$ . Mientras que las otras dos muestras presentan valores inferiores:  $IQR=9.0000$  para la muestra del grupo de Nativos, y  $IQR=8.0000$  para el M1 del grupo de no Nativos. Así, es la muestra del grupo «NN-M1-F» la que presenta mayor concentración de valores en torno a las medidas de centralización. Estas pequeñas diferencias se pueden apreciar con más detalle en el diagrama de cajas (Gráfico 128).

Tanto en el gráfico de dispersión como en el diagrama de cajas, podemos observar, también, los elementos extremos que poseen las tres muestras.

En cuanto a la forma, el valor de la curtosis es positiva en las tres muestras, pero con valores muy diferentes. Mientras que el valor de la curtosis de la muestra del grupo de no Nativos en el M1 está próximo a cero ( $curtosis=0.5387$ ), por tanto, cercano a la distribución normal o campana de Gauss; el valor de la curtosis del M2 del grupo de no Nativos es  $curtosis=6.7171$ , por tanto, con una tendencia a ser más aguda. El valor de la muestra del grupo de Nativos es  $curtosis=2.4058$ . Así, en los tres casos, estamos ante curvas con tendencia a ser agudas, pero con diferentes grados. En los histogramas siguientes podemos visualizar esta tendencia.

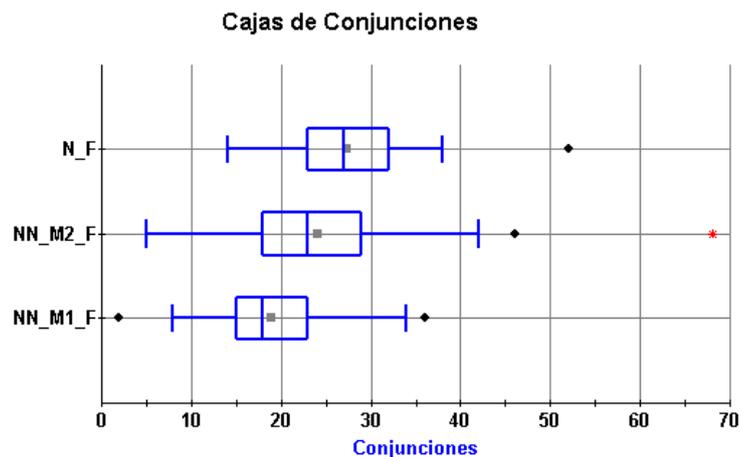


Gráfico 128. Diagrama de cajas para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F

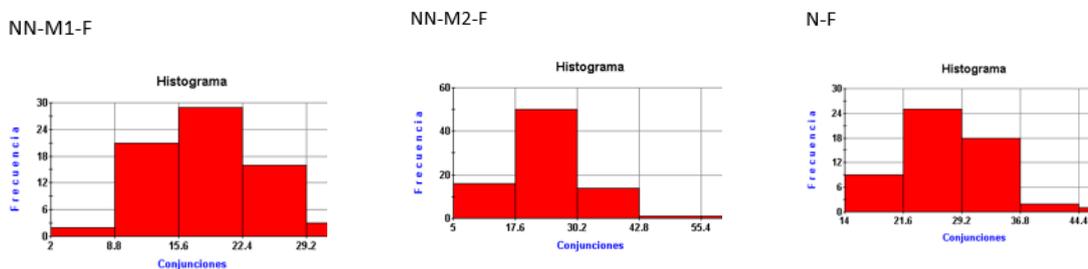


Gráfico 129. Histograma para la variable «Conjunciones» NN-M1-F / NN-M2-F / N-F

### 6.1.5 Estadística descriptiva para la muestra «1D»

Como indicamos al inicio de este capítulo, en este apartado vamos a presentar la segunda parte de la estadística descriptiva de nuestro corpus. Es decir, vamos a desglosar las tres muestras correspondientes al texto codificado como «1D», una vez analizado en el apartado anterior las muestras correspondientes al texto «F».

Los grupos son los mismos, (i) un grupo formado por sujetos no Nativos de la lengua española que realizaron la prueba en dos momentos de su proceso de aprendizaje (M1; M2) y (ii) un grupo formado por sujetos Nativos de la lengua española que realizaron la prueba en un solo momento.

Como se puede ver en la tabla siguiente (Tabla 63), el número de sujetos que realizaron la prueba correspondiente al texto «1D» varía respecto a los que realizaron el escrito «F». En esta ocasión, el número más alto de participantes lo encontramos en la muestra del grupo

de Nativos, «N-1D», con un total de n=62. Mientras que en el grupo de no Nativos M1 hay un total de n=34 y en el M2, n=31. Así, vemos que el número total por variables es muy superior en la muestra del grupo de Nativos en comparación con el grupo de no Nativos en los M1 y M2.

En cuanto a los resultados del grupo de no Nativos, llama la atención los totales algo superiores en la muestra del grupo de no Nativos M2 en comparación con el M1, aun teniendo éste tres participantes más en la muestra.

Así, partiendo de estos datos, de estos totales, vamos a realizar a continuación la estadística descriptiva de cada una de las variables para entender cómo se distribuyen en relación a las medidas de centralización, dispersión y la forma.

Muestras «1D»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Tokens</i>	3.097	3.571	9.746
<i>Types</i>	2.248	2.572	6.607
Verbos	504	548	1.429
Sustantivos	1.278	1575	3.993
Adjetivos	457	538	1.206
Adverbios	219	178	507
Pronombres	269	340	1.149
Preposiciones	175	178	619
Artículos	74	77	223
Conjunciones	121	137	620

Tabla 63. Totales de las variables para las muestras «1D» NN-M1-1D / NN-M2-1D / N-1D

(1) *Tokens*: Es el número total de palabras:

Muestras «1D»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Tokens</i>	3.097	3.571	9.746

Partiendo de los totales de *tokens* en cada muestra, vamos a ver qué nos dicen los estadísticos sobre la forma, la centralización o la dispersión de los datos.

**Estadísticos para la variable Tokens**

<b>Grupos</b>	<b>NN_M1_1D</b>	<b>NN_M2_1D</b>	<b>N_1D</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>91.0882</b>	<b>115.1935</b>	<b>157.1935</b>
<b>Mediana</b>	<b>92.5000</b>	<b>111.0000</b>	<b>154.5000</b>
<b>Varianza</b>	<b>739.7193</b>	<b>586.2280</b>	<b>904.7816</b>
<b>Desviación Típica</b>	<b>27.1978</b>	<b>24.2121</b>	<b>30.0796</b>
<b>Mínimo</b>	<b>52.0000</b>	<b>74.0000</b>	<b>103.0000</b>
<b>Máximo</b>	<b>150.0000</b>	<b>165.0000</b>	<b>230.0000</b>
<b>Cuartil Inferior</b>	<b>62.0000</b>	<b>95.0000</b>	<b>130.0000</b>
<b>Cuartil Superior</b>	<b>113.0000</b>	<b>132.0000</b>	<b>182.0000</b>
<b>Rango Intercuartílico</b>	<b>51.0000</b>	<b>37.0000</b>	<b>52.0000</b>
<b>Curtosis</b>	<b>-0.9076</b>	<b>-0.5613</b>	<b>-0.6184</b>

Tabla 64. Estadísticos para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D

Las medidas de centralización, la media y la mediana, son muy diferentes en cada una de las tres muestras. El valor más alto lo encontramos en la muestra del grupo de Nativos con una media=157.1935 y una mediana=154.5000. No obstante, el dato que llama la atención aquí es la diferencia entre la media que presentan las muestras del grupo de no Nativos. Podemos observar que en la muestra del M2 el valor de la media=115.1935 es muy superior al del M1, media=91.0882.

Otro elemento a resaltar es la relación entre la media y la mediana en cada una de las muestras. Podemos observar que la diferencia no es mucha, pero es en el M2 del grupo de no Nativos donde encontramos una diferencia mayor, cuatro puntos, mientras que en el M1 es de uno, y en la muestra del grupo de Nativos es de poco más de dos.

El siguiente estadístico que nos ayuda a entender cómo se distribuyen las muestras es la desviación típica que nos habla de la dispersión de las mismas. Como podemos ver en la tabla de estadísticos, el valor más bajo, por tanto, menos dispersión, lo encontramos en la muestra «NN-M2-1D»,  $s=24.2121$ ; y el valor más alto corresponde a la muestra del grupo de Nativos,  $s=30.0796$ , la que presenta mayor dispersión. El M1 del grupo de no Nativos presenta un valor de  $s=27.1978$ . Estos datos los podemos visualizar en el gráfico de dispersión.

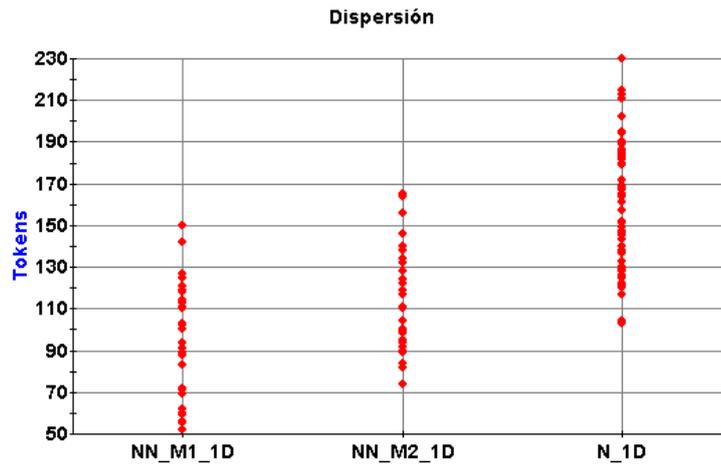


Gráfico 130. Gráfico de dispersión para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D

El otro estadístico que nos informa sobre la dispersión de las muestras es el rango intercuartílico. En este caso, recordemos, indica la dispersión o concentración de los datos en torno a las medidas de centralización, la media y la mediana. Podemos observar en la tabla de estadísticos que el valor más bajo es el de la muestra «NN-M2-1D»,  $IQR=37.0000$ . Es decir, el 50% de los datos de esta muestra los vamos a encontrar más concentrados en torno a la media y la mediana que en las otras dos muestras que presentan unos valores más altos, por tanto, más dispersión:  $IQR=51.0000$  para la muestra «NN-M1-1D» y  $IQR=52.0000$  para la muestra «N-1D». Estas diferencias las vamos a visualizar más claramente en el diagrama de cajas que mostramos a continuación.

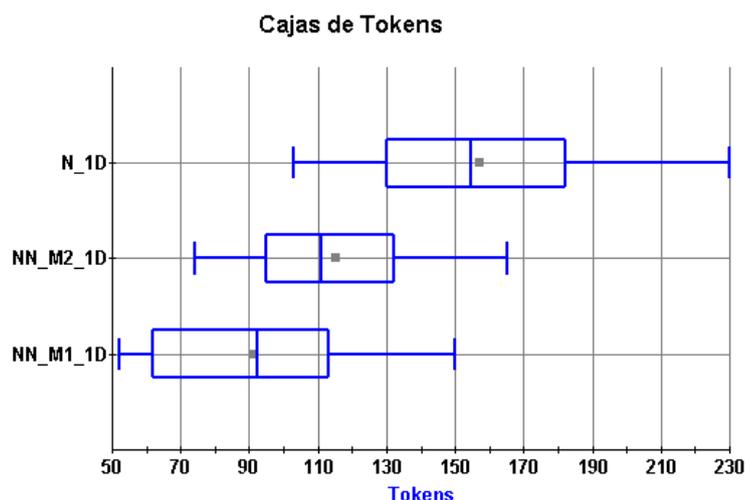


Gráfico 131. Diagrama de cajas para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D

Al observar el diagrama de cajas, constatamos la diferencia de tamaño entre la caja de la muestra del grupo de no Nativos M2 y las cajas de las otras dos muestras. También

podemos ver que el tamaño del diagrama de cajas de las dos muestras del grupo de no Nativos (caja y bigotes) es muy similar, la diferencia radica en los estadísticos «mínimo» y «máximo» que tienen valores muy distintos. Mientras que el diagrama de cajas de la muestra del grupo de Nativos presenta un tamaño mayor respecto a los otros dos.

También podemos visualizar las pequeñas diferencias que hemos comentado más arriba entre la media y la mediana en cada una de las muestras.

En cuanto a la forma de la distribución de las muestras en comparación con una distribución normal o de campana de Gauss, las tres muestras de nuestro estudio presentan una curtosis con valores negativos, es decir, que la curva tiene tendencia a ser más plana que en una distribución normal con valor cero. No obstante hay que señalar que aun siendo valores negativos, están muy próximos al cero: curtosis= -0.9076 («NN-M1-1D»); curtosis= -0.5613 («NN-M2-1D»); curtosis= -0.6184). En los histogramas siguientes podemos apreciar esta tendencia.

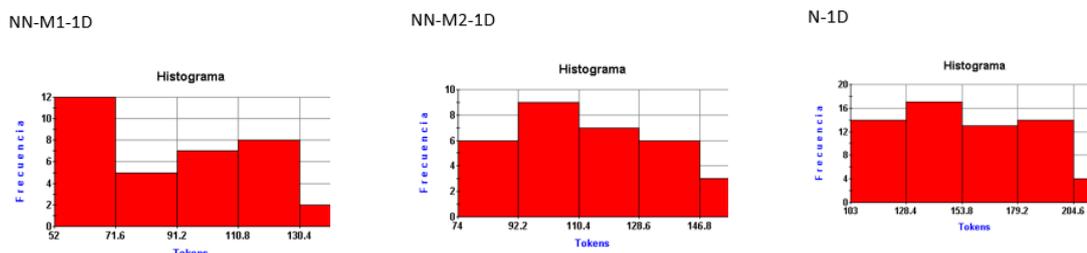


Gráfico 132. Histograma para la variable «Tokens» NN-M1-1D / NN-M2-1D / N-1D

(2) *Types*: Es el número total de palabras diferentes.

Muestras «1D»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Types</i>	2.248	2.572	6.607

Lo que llama la atención de estos totales es la diferencia que hay entre las dos muestras del grupo de no nativos teniendo en cuenta que el número de sujetos del M2 es inferior al del M1. A partir de estos datos, vamos a ver qué dicen los estadísticos de cada una de las muestras.

**Estadísticos para la variable *Types***

<b>Grupos</b>	<b>NN_M1_1D</b>	<b>NN_M2_1D</b>	<b>N_1D</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>66.1176</b>	<b>82.9677</b>	<b>106.5645</b>
<b>Mediana</b>	<b>66.5000</b>	<b>81.0000</b>	<b>103.5000</b>
<b>Varianza</b>	<b>277.8645</b>	<b>223.8323</b>	<b>316.6105</b>
<b>Desviación Típica</b>	<b>16.6693</b>	<b>14.9610</b>	<b>17.7936</b>
<b>Mínimo</b>	<b>42.0000</b>	<b>59.0000</b>	<b>76.0000</b>
<b>Máximo</b>	<b>105.0000</b>	<b>112.0000</b>	<b>148.0000</b>
<b>Cuartil Inferior</b>	<b>52.0000</b>	<b>69.0000</b>	<b>93.0000</b>
<b>Cuartil Superior</b>	<b>77.0000</b>	<b>94.0000</b>	<b>118.0000</b>
<b>Rango Intercuartílico</b>	<b>25.0000</b>	<b>25.0000</b>	<b>25.0000</b>
<b>Curtosis</b>	<b>-0.6568</b>	<b>-0.7308</b>	<b>-0.5335</b>

Tabla 65. Estadísticos para la variable «*Types*» NN-M1-1D / NN-M2-1D / N-1D

Los valores de los descriptores que señalan las medidas de centralización, la media y la mediana, indican que es la muestra del grupo de nativos la que presenta el valor más alto con gran diferencia respecto a las otras dos muestras, como podemos ver en la tabla de estadísticos. También observamos que la mayor diferencia entre media y mediana se da igualmente en la muestra del grupo de no nativos con tres puntos de diferencia; mientras que es en la muestra del grupo de no nativos M1 donde encontramos la diferencia más baja, menos de un punto.

Para conocer la dispersión de las muestras, comprobamos los valores del estadístico «desviación típica» y observamos que los valores de las tres muestras están muy próximos entre sí, pero el más alto corresponde a la del grupo de nativos,  $s=17.7936$ , seguida por el grupo «NN-M1-1D»,  $s=16.6693$ , y el valor más bajo corresponde al grupo «NN-M2-1D»,  $s=14.9610$ . Vemos, como en la variable anterior, que el grupo de no nativos M2 es el que presenta menor dispersión. Estas pequeñas diferencias las podemos visualizar en el gráfico 133 de dispersión.

Al comprobar el valor del otro estadístico que informa de la dispersión, el rango intercuartílico, vemos que es el mismo para las tres muestras ( $IQR=25.0000$ ). Así, el 50% de los valores que se encuentran en torno a la media y la mediana presentan la misma dispersión en las tres muestras. Es decir, que el tamaño de las cajas, como podemos comprobar en el diagrama de cajas (Gráfico 134), es el mismo.

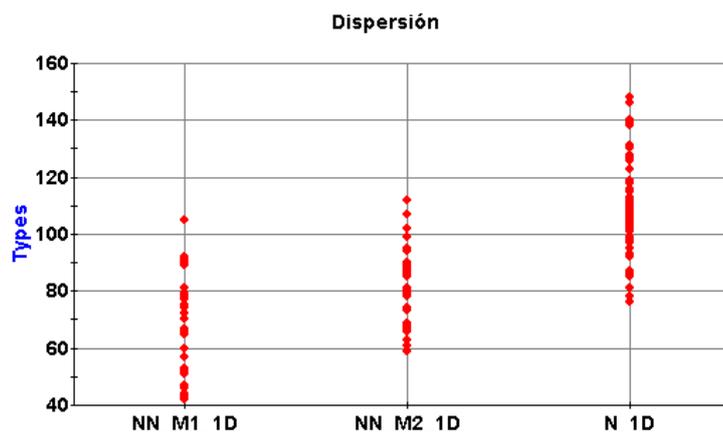


Gráfico 133. Gráfico de dispersión para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D

Al observar el diagrama de cajas, comprobamos las diferencias entre media y mediana que acabamos de señalar. Observamos que es la muestra del grupo de Nativos, «N-1D», la que presenta una mayor distancia. Eso significa que la muestra presenta una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

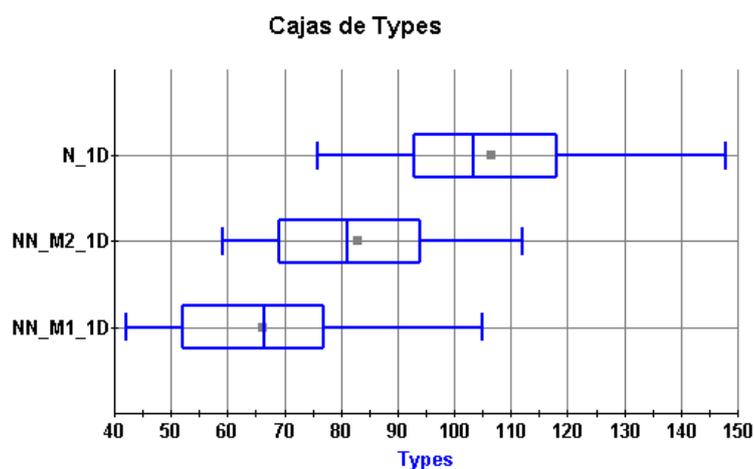


Gráfico 134. Diagrama de cajas para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D

También el estadístico «curtosis» informa de la forma de la muestra. En esta variable vemos que el valor es negativo para los dos grupos, Nativos y no Nativos, tanto en el M1 como en el M2. Las tres muestras están próximas al valor cero, valor que indica que se trata de una curva normal o de campana de Gauss. Pero es la muestra del grupo N el que presenta un valor más alejado del cero: curtosis= -0.5335 («N-1D»); curtosis= -0.6568 («NN-M1-1D»); curtosis= -0.7308 («NN-M2-1D»). Esta proximidad a una distribución normal lo podemos visualizar en los histogramas siguientes.

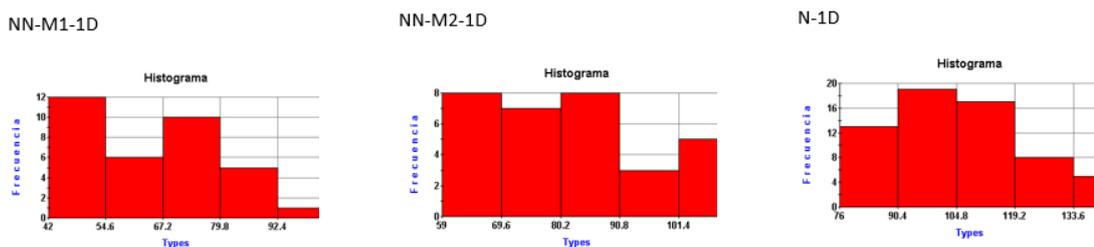


Gráfico 135. Histograma para la variable «Types» NN-M1-1D / NN-M2-1D / N-1D

(3) Verbos: Número total de verbos.

Muestras «ID»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Verbos	504	548	1.429

También en la variable «Verbos» comprobamos que hay un pequeño aumento en el número total entre las muestras del M1 y el M2 del grupo de no Nativos. A continuación presentamos la tabla de estadísticos (Gráfico 66) para conocer cómo se distribuye cada una de las muestras.

Al observar los estadísticos de centralización, la media y la mediana, llama la atención los valores tan dispares existentes entre las tres muestras. Entre el valor más alto, correspondiente a la muestra del grupo de Nativos, y el valor más bajo, que pertenece a la muestra M1 del grupo de no Nativos, hay una diferencia de más de catorce puntos. En cambio, la distancia entre la media y la mediana dentro de cada muestra no supera los dos puntos en ningún caso.

**Estadísticos para la variable Verbos**

Grupos	NN_M1_ID	NN_M2_ID	N_ID
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>19.9412</b>	<b>25.9032</b>	<b>34.3710</b>
<b>Mediana</b>	<b>18.0000</b>	<b>24.0000</b>	<b>33.0000</b>
<b>Varianza</b>	<b>71.1480</b>	<b>66.3570</b>	<b>71.9093</b>
<b>Desviación Típica</b>	<b>8.4349</b>	<b>8.1460</b>	<b>8.4799</b>
<b>Mínimo</b>	<b>8.0000</b>	<b>12.0000</b>	<b>17.0000</b>
<b>Máximo</b>	<b>44.0000</b>	<b>43.0000</b>	<b>51.0000</b>
<b>Cuartil Inferior</b>	<b>14.0000</b>	<b>21.0000</b>	<b>28.0000</b>
<b>Cuartil Superior</b>	<b>25.0000</b>	<b>31.0000</b>	<b>40.0000</b>
<b>Rango Intercuartílico</b>	<b>11.0000</b>	<b>10.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>0.9469</b>	<b>-0.3930</b>	<b>-0.7659</b>

Tabla 66. Estadísticos para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D

También llama la atención los resultados del estadístico «desviación típica» por la casi coincidencia de los valores de las tres muestras. Podemos ver esta similitud en el siguiente gráfico de dispersión.

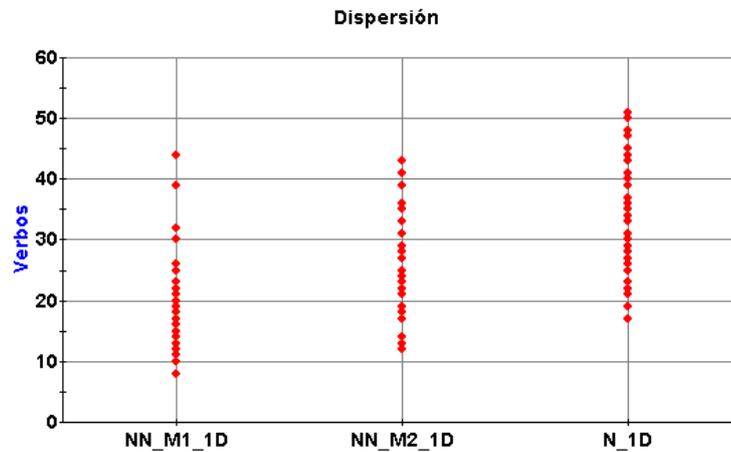


Gráfico 136. Gráfico de dispersión para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D

Las tres muestras presentan un valor del descriptor «rango intercuartílico», como vemos en la tabla de estadísticos, muy próximo. El valor más alto (IQR=12.0000) corresponde a la muestra del grupo Nivo, mientras que el valor más bajo (IQR=10.0000) lo observamos en la muestra del M2 del grupo de no Nativos. La muestra del grupo de no Nativos M1 presenta un valor IQR=11.0000. Podemos visualizar estos datos, con sus pequeñas diferencias, en los diagramas de cajas que mostramos a continuación y donde podemos ver con más claridad el tamaño de cada una de las cajas.

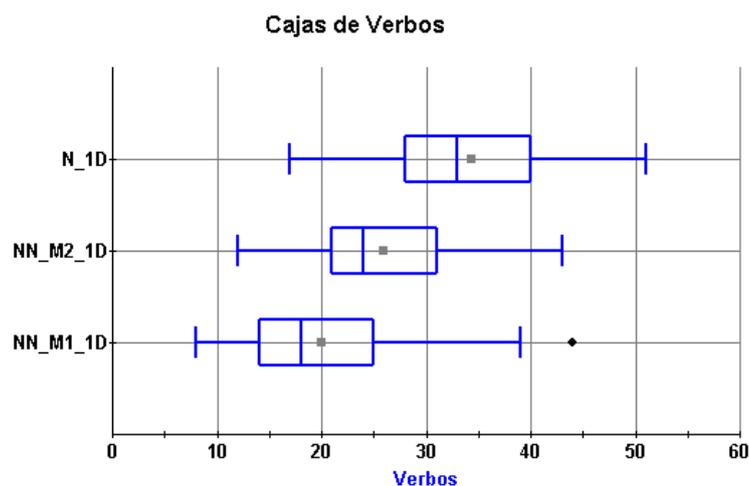


Gráfico 137. Diagrama de cajas para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D

También podemos ver la distancia entre la media y la mediana que presenta cada una de las muestras. Como hemos señalado más arriba, la diferencia en las tres muestras era muy similar y en el diagrama de cajas lo podemos constatar. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es negativo en las muestras del grupo «NN-M2-1D»,  $\text{curtosis} = -0.3930$ , y «N-1D»,  $\text{curtosis} = -0.7659$ , y positivo en la muestra del grupo de no Nativos «NN-M1-1D»,  $\text{curtosis} = 0.9469$ . Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan valores negativos («NN-M2-1D»; «N-1D») tienden a una curva más plana, mientras que la muestra que presenta el valor positivo («NN-M1-1D») tiende a una curva más aguda. Podemos visualizar estos datos en los histogramas siguientes.

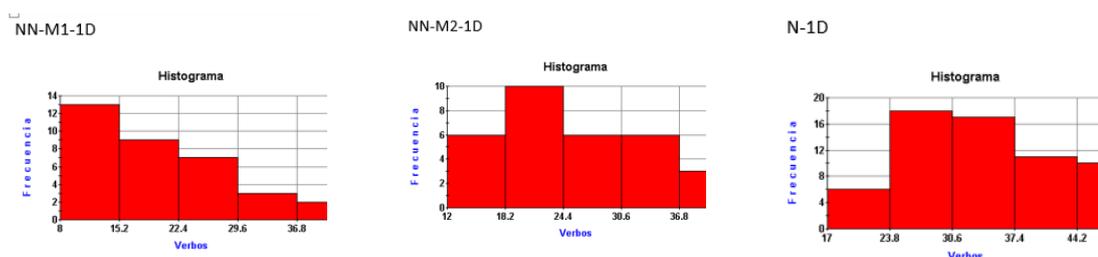


Gráfico 138. Histograma para la variable «Verbos» NN-M1-1D / NN-M2-1D / N-1D

(4) Sustantivos: Total de sustantivos.

Muestras «ID»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Sustantivos	1.278	1.575	3.993

Observamos una vez más que hay una pequeña diferencia entre las dos muestras del grupo de no Nativos. Vemos que la muestra del M2 es algo superior a la del M1. En la tabla 67 vamos a ver qué dicen los estadísticos a partir de estos totales.

Al observar los estadísticos de centralización, la media y la mediana, llama la atención, igual que en la variable anterior, los valores tan diferentes que se dan entre las tres muestras. Entre el valor más alto, correspondiente a la muestra del grupo de Nativos, y el valor más bajo, que pertenece a la muestra M1 del grupo de no Nativos, hay una diferencia de más de

31 puntos. Entre el valor intermedio, correspondiente a la muestra del M2 del grupo de no Nativos, y el valor más alto hay una diferencia de más de diecisiete puntos y con la muestra del M1 del grupo de no Nativos hay una diferencia de más de catorce puntos.

**Estadísticos para la variable Sustantivos**

Grupos	NN_M1_1D	NN_M2_1D	N_1D
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>44.7941</b>	<b>59.3871</b>	<b>76.6452</b>
<b>Mediana</b>	<b>42.0000</b>	<b>59.0000</b>	<b>77.0000</b>
<b>Varianza</b>	<b>200.4109</b>	<b>158.9785</b>	<b>300.8556</b>
<b>Desviación Típica</b>	<b>14.1567</b>	<b>12.6087</b>	<b>17.3452</b>
<b>Mínimo</b>	<b>26.0000</b>	<b>36.0000</b>	<b>44.0000</b>
<b>Máximo</b>	<b>72.0000</b>	<b>82.0000</b>	<b>116.0000</b>
<b>Cuartil Inferior</b>	<b>32.0000</b>	<b>50.0000</b>	<b>65.0000</b>
<b>Cuartil Superior</b>	<b>58.0000</b>	<b>70.0000</b>	<b>86.0000</b>
<b>Rango Intercuartílico</b>	<b>26.0000</b>	<b>20.0000</b>	<b>21.0000</b>
<b>Curtosis</b>	<b>-1.1176</b>	<b>-1.1304</b>	<b>-0.6025</b>

Tabla 67. Estadísticos para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D

Al comparar la distancia entre la media y la mediana dentro de cada muestra, vemos que salvo la muestra «NN-M1-1D» que presenta una diferencia de más de dos puntos, las otras dos muestras no superan el punto de diferencia.

La desviación típica de cada una de las muestras presenta valores muy diferentes. El valor más alto ( $s=17.3452$ ) lo encontramos en la muestra del grupo de Nativos y el valor más bajo ( $s=12.6087$ ) en la muestra del M2 del grupo de no Nativos. Vemos que la diferencia entre estos dos valores supera los cuatro puntos. El valor intermedio ( $s=14.1567$ ) pertenece a la muestra del M1 del grupo de no Nativos. Podemos visualizar estos datos en el gráfico de dispersión que mostramos a continuación.

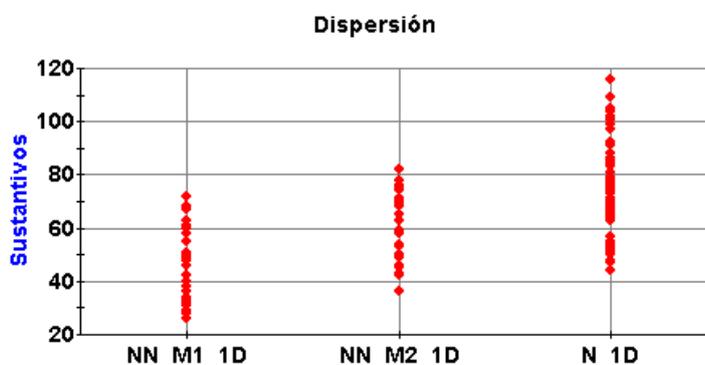


Gráfico 139. Gráfico de dispersión para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta valores diferentes para cada muestra. Podemos observar en la tabla de estadísticos que la muestra del grupo de Nativos (IQR=21.0000) y la muestra del grupo de no Nativos M2 (IQR=20.0000) presentan un valor muy próximo. En cambio, la muestra del grupo «NN-M1-1D» (IQR=26.0000) supera en cinco y seis puntos a las otras dos muestras respectivamente. Esto significa que la caja de esta muestra es mayor y, por tanto, el 50% de los valores de la muestra insertos en ella presentan una mayor dispersión. En el diagrama de cajas (Gráfico 140) que mostramos más abajo podemos ver gráficamente esta información.

También podemos ver la distancia entre la media y la mediana que presenta cada una de las muestras. Como hemos señalado más arriba, la diferencia en las muestras «NN-M2-1D» y «N-1D» era mínima y así lo podemos constatar en el diagrama de cajas. Es en la muestra del grupo «NN-M1-1D» donde se aprecia una mayor distancia entre la media y la mediana. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

También hay que destacar, al observar los diagramas de cajas, la diferencia en el descriptor «máximo». Podemos ver que el descriptor «máximo» de la muestra del grupo de Nativos se aleja bastante del descriptor «máximo» de las otras dos muestras.

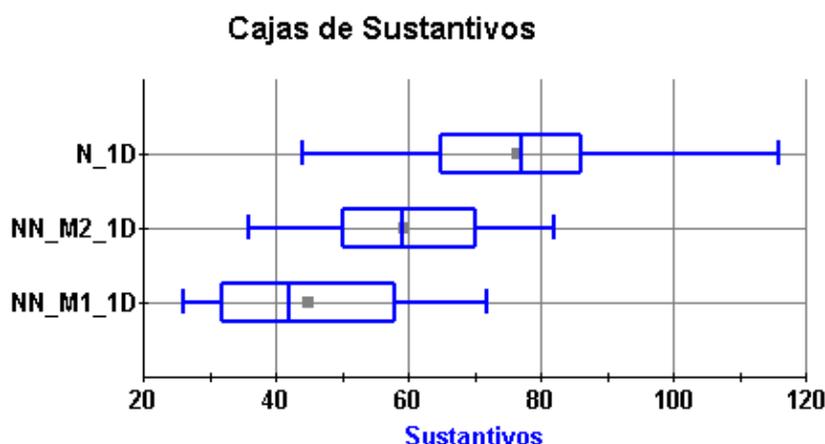


Gráfico 140. Diagrama de cajas para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es negativo en las tres muestras. La muestra «N-1D» presenta una curtosis= -0.6025; la muestra del grupo «NN-M2-1D» muestra una curtosis= -1.1304, y la muestra «NN-M1-1D», curtosis= -1.1176. Respecto a una curtosis cero, distribución normal o

campana de Gauss, las muestras que presentan valores negativos tienden a una curva más plana. Podemos visualizar estos datos en los histogramas siguientes.

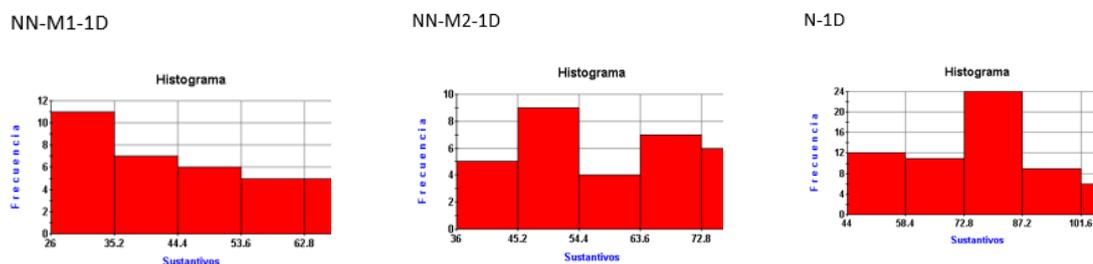


Gráfico 141. Histograma para la variable «Sustantivos» NN-M1-1D / NN-M2-1D / N-1D

(5) Adjetivos: Total de adjetivos.

Muestras «ID»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Adjetivos	457	538	1.206

Observamos nuevamente la pequeña diferencia entre las dos muestras del grupo de Nativos. Vemos que la muestra del M2 es algo superior a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos totales.

**Estadísticos para la variable Adjetivos :**

Grupos	NN_M1_1D	NN_M2_1D	N_1D
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>19.8235</b>	<b>26.6452</b>	<b>31.5968</b>
<b>Mediana</b>	<b>19.0000</b>	<b>25.0000</b>	<b>30.5000</b>
<b>Varianza</b>	<b>61.7861</b>	<b>37.9699</b>	<b>76.3429</b>
<b>Desviación Típica</b>	<b>7.8604</b>	<b>6.1620</b>	<b>8.7374</b>
<b>Mínimo</b>	<b>7.0000</b>	<b>16.0000</b>	<b>16.0000</b>
<b>Máximo</b>	<b>36.0000</b>	<b>44.0000</b>	<b>57.0000</b>
<b>Cuartil Inferior</b>	<b>13.0000</b>	<b>22.0000</b>	<b>26.0000</b>
<b>Cuartil Superior</b>	<b>25.0000</b>	<b>31.0000</b>	<b>36.0000</b>
<b>Rango Intercuartílico</b>	<b>12.0000</b>	<b>9.0000</b>	<b>10.0000</b>
<b>Curtosis</b>	<b>-0.5869</b>	<b>0.6209</b>	<b>0.8133</b>

Tabla 68. Estadísticos para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D

Los valores de los estadísticos que señalan las medidas de centralización, la media y la mediana, indican que es la muestra del grupo de Nativos la que presenta el valor más alto con una diferencia respecto a las otras dos muestras («NN-M1-1D»; «NN-M2-1D») de más

de once y cinco puntos respectivamente, como podemos ver en la tabla de estadísticos. Observamos, asimismo, que la mayor diferencia entre media y mediana se da en la muestra del grupo «NN-M1-1D» con casi dos puntos; mientras que es en la muestra del grupo de Nativos donde encontramos la diferencia más baja, poco más de un punto.

Para conocer la dispersión de las muestras, comprobamos los valores del estadístico «desviación típica» y observamos que los valores de las tres muestras están muy próximos entre sí, pero el más alto corresponde a la del grupo de Nativos,  $s=8.7374$ , seguida por el grupo «NN-M1-1D»,  $s=7.8604$ , y el valor más bajo corresponde al grupo «NN-M2-1D»,  $s=6.1620$ . Estas pequeñas diferencias las podemos visualizar en el gráfico de dispersión que mostramos a continuación.

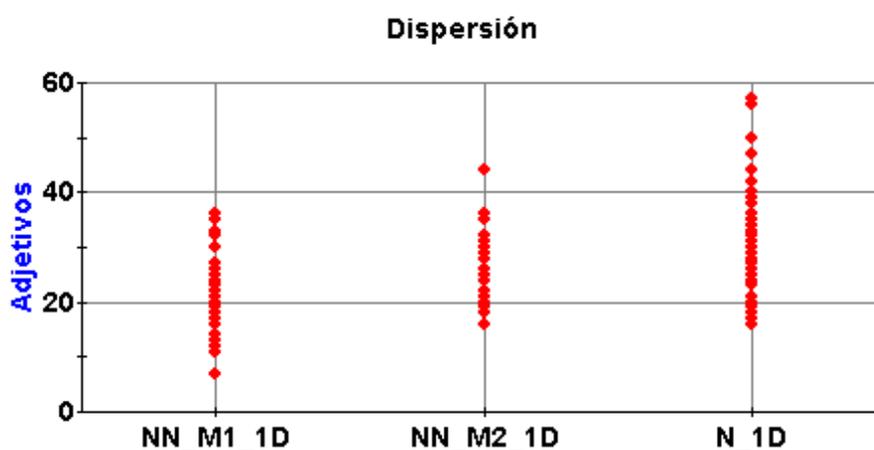


Gráfico 142. Gráfico de dispersión para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta valores diferentes para cada muestra. Podemos observar en la tabla de estadísticos que la muestra del grupo de no Nativos M1 presenta el valor más alto (IQR=12.0000). Mientras que el valor más bajo lo encontramos en la muestra del grupo de no Nativos M2 (IQR=9.0000). El valor intermedio lo presenta la muestra del grupo de Nativos (IQR=10.0000). Esto significa que la caja mayor es la de la muestra «NN-M1-1D», y, por tanto, el 50% de los valores de la muestra insertos en ella presentan una mayor dispersión en comparación con las otras dos muestras. En el diagrama de cajas (Gráfico 143) podemos ver gráficamente esta información.

Cuando observamos la distancia entre la media y la mediana que presenta cada una de las muestras, comprobamos que la diferencia en las muestras «NN-M1-1D» y «N-1D» es mínima y así lo podemos constatar en el diagrama de cajas. En esta variable es en la muestra

del grupo «NN-M2-1D» donde se aprecia una mayor distancia entre la media y la mediana. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

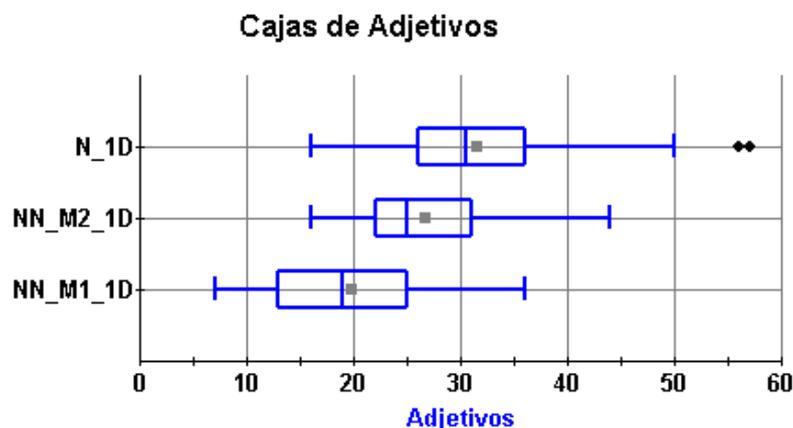


Gráfico 143. Diagrama de cajas para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D

Llama también la atención la coincidencia de valores en el descriptor «mínimo» entre la muestra «NN-M2-1D» y la muestra «N-1D». Sin embargo, vemos que el descriptor «máximo» de la muestra «N-1D» se distancia de la otra muestra en trece puntos como se observa en el diagrama.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es positivo en las muestras del grupo «NN-M2-1D»,  $\text{curtosis}=0.6209$ , y «N-1D»,  $\text{curtosis}=0.8133$ , y negativo en la muestra del grupo de no Nativos «NN-M1-1D»,  $\text{curtosis}=-0.5869$ . Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan los valores positivos («NN-M2-1D»; «N-1D») tienden a una curva más aguda, mientras que la muestra que presenta el valor negativo («NN-M1-1D») tiende a una curva más plana. Podemos visualizar estos datos en los histogramas siguientes.

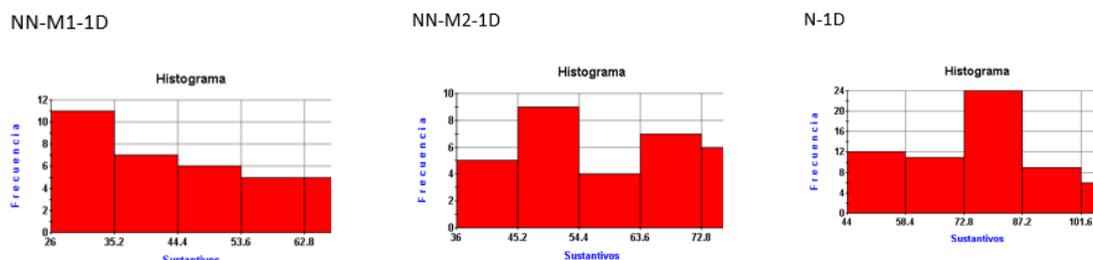


Gráfico 144. Histograma para la variable «Adjetivos» NN-M1-1D / NN-M2-1D / N-1D

(6) Adverbios: Totales de adverbios.

<b>Muestras «ID»</b>			
	<b>NN-M1-1D</b> (n=34)	<b>NN-M2-1D</b> (n=31)	<b>N-1D</b> (n=62)
Adverbios	219	178	507

Observamos, como en todas las variables, que hay una pequeña diferencia entre las dos muestras del grupo de no Nativos. Vemos que la muestra del M2 es algo superior a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos datos.

**Estadísticos para la variable Adverbios**

<b>Grupos</b>	<b>NN_M1_ID</b>	<b>NN_M2_ID</b>	<b>N_ID</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>11.6176</b>	<b>14.4194</b>	<b>20.3548</b>
<b>Mediana</b>	<b>11.0000</b>	<b>13.0000</b>	<b>19.5000</b>
<b>Varianza</b>	<b>37.4554</b>	<b>24.1849</b>	<b>51.1835</b>
<b>Desviación Típica</b>	<b>6.1201</b>	<b>4.9178</b>	<b>7.1543</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>7.0000</b>	<b>8.0000</b>
<b>Máximo</b>	<b>26.0000</b>	<b>27.0000</b>	<b>39.0000</b>
<b>Cuartil Inferior</b>	<b>7.0000</b>	<b>10.0000</b>	<b>15.0000</b>
<b>Cuartil Superior</b>	<b>17.0000</b>	<b>18.0000</b>	<b>24.0000</b>
<b>Rango Intercuartílico</b>	<b>10.0000</b>	<b>8.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>-0.3267</b>	<b>0.1527</b>	<b>0.0054</b>

Tabla 69. Estadísticos para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D

Al observar los estadísticos de centralización, la media y la mediana, observamos que las tres muestras presentan valores muy dispares. Entre el valor más alto, correspondiente a la muestra del grupo de Nativos, y el valor más bajo, que pertenece a la muestra M1 del grupo de no Nativos, hay una diferencia de más de 8 puntos. Entre el valor intermedio, correspondiente a la muestra del M2 del grupo de no Nativos, y el valor más alto, el del grupo de Nativos, hay una diferencia de más de seis puntos y con la muestra del M1 del grupo de no Nativos hay una diferencia de casi tres puntos.

Al comparar la distancia entre la media y la mediana dentro de cada muestra, vemos que las diferencias son mínimas. La muestra «N-1D», por ejemplo, presenta una diferencia inferior a un punto; las otras dos muestras presentan una diferencia de poco más de un punto.

Para conocer la dispersión de las muestras, comprobamos los valores del estadístico «desviación típica» y observamos que los valores de las muestras «NN-M1-1D»,  $s=6.1201$ ,

y «N-1D»,  $s=7.1543$ , están muy próximos entre sí, siendo la muestra del grupo Nivo el que presenta mayor dispersión. Sin embargo, la muestra «NN-M2-1D»,  $s=4.9178$ , es algo más baja que las otras dos. Estas pequeñas diferencias las podemos visualizar en el gráfico de dispersión que mostramos a continuación.

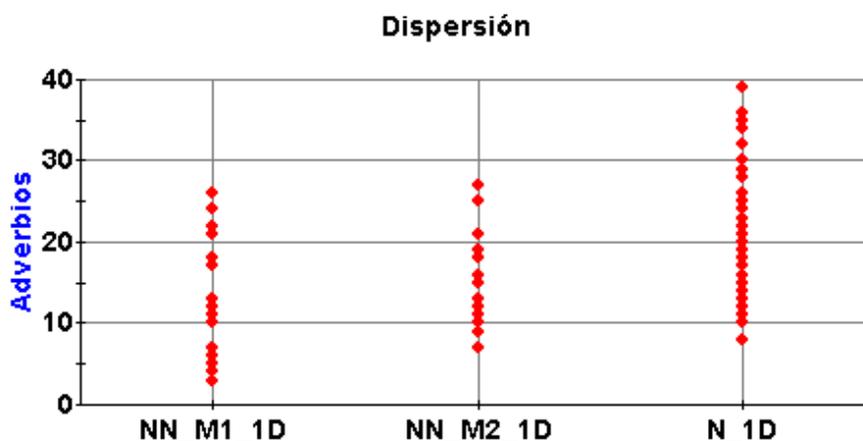


Gráfico 145. Gráfico de dispersión para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D

El otro descriptor que informa de la dispersión de los datos, el rango intercuartílico, presenta valores diferentes para cada muestra, aunque muy próximos entre sí, solo un punto los distancia. Podemos observar en la tabla de estadísticos que la muestra del grupo de no Nativos M1 presenta el valor más alto (IQR=10.0000). Mientras que el valor más bajo lo encontramos en la muestra del grupo de no Nativos M2 (IQR=8.0000). El valor intermedio lo presenta la muestra del grupo de Nativos (IQR=9.0000). Esto significa que la caja mayor es la de la muestra «NN-M1-1D», y, por tanto, el 50% de los valores de la muestra insertos en ella presentan una mayor dispersión en comparación con las otras dos muestras. En el diagrama de cajas (Gráfico 146) que mostramos a continuación podemos ver gráficamente esta información.

Cuando observamos la distancia entre la media y la mediana que presenta cada una de las muestras, comprobamos que la diferencia en las muestras «NN-M1-1D» y «N-1D» es mínima y así lo podemos constatar en el diagrama de cajas. En esta variable es en la muestra del grupo «NN-M2-1D» donde se aprecia una mayor distancia entre la media y la mediana, aunque también pequeña. Estas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

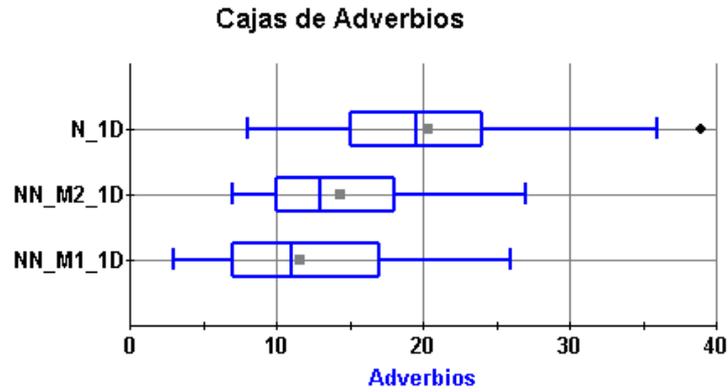


Gráfico 146. Diagrama de cajas para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D

Llama también la atención, por un lado, la casi coincidencia de valores en el descriptor «mínimo» entre la muestra «NN-M2-1D» y la muestra «N-1D». Sin embargo, vemos que el descriptor «máximo» de la muestra «N-1D» se distancia de la otra muestra en doce puntos como se observa en el diagrama. Por otro lado, tenemos el descriptor «máximo» de las muestras «NN-M2-1D» y «NN-M1-1D» muy cercanas la una de la otra.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es positivo en las muestras del grupo «NN-M2-1D»,  $curtosis=0.1527$ , y «N-1D»,  $curtosis=0.0054$ , y negativo en la muestra del grupo de no Nativos «NN-M1-1D»,  $curtosis=-0.3267$ . Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan los valores positivos («NN-M2-1D»; «N-1D») tienden a una curva más aguda, mientras que la muestra que presenta el valor negativo («NN-M1-1D») tiende a una curva más plana.

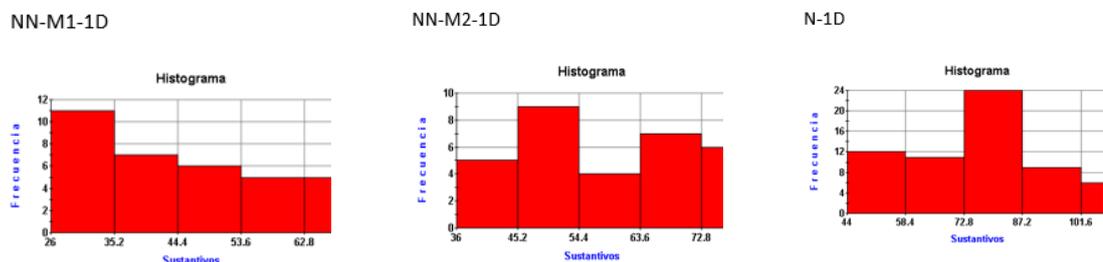


Gráfico 147. Histograma para la variable «Adverbios» NN-M1-1D / NN-M2-1D / N-1D

(7) Pronombres: Totales de pronombres.

<b>Muestras «ID»</b>			
	<b>NN-M1-1D</b> (n=34)	<b>NN-M2-1D</b> (n=31)	<b>N-1D</b> (n=62)
Pronombres	269	340	981

Observamos nuevamente la pequeña diferencia entre las dos muestras del grupo de no Nativos. Vemos que la muestra del M2 es algo superior a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos totales.

**Estadísticos para la variable Pronombres**

<b>Grupos</b>	<b>NN_M1_ID</b>	<b>NN_M2_ID</b>	<b>N_ID</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>16.6765</b>	<b>21.5806</b>	<b>31.8710</b>
<b>Mediana</b>	<b>17.0000</b>	<b>21.0000</b>	<b>30.0000</b>
<b>Varianza</b>	<b>27.2558</b>	<b>28.5849</b>	<b>72.9011</b>
<b>Desviación Típica</b>	<b>5.2207</b>	<b>5.3465</b>	<b>8.5382</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>13.0000</b>	<b>19.0000</b>
<b>Máximo</b>	<b>27.0000</b>	<b>32.0000</b>	<b>55.0000</b>
<b>Cuartil Inferior</b>	<b>12.0000</b>	<b>17.0000</b>	<b>25.0000</b>
<b>Cuartil Superior</b>	<b>20.0000</b>	<b>26.0000</b>	<b>37.0000</b>
<b>Rango Intercuartílico</b>	<b>8.0000</b>	<b>9.0000</b>	<b>12.0000</b>
<b>Curtosis</b>	<b>-0.5631</b>	<b>-0.9361</b>	<b>0.6445</b>

Tabla 70. Estadísticos para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D

Los valores de los descriptores que señalan las medidas de centralización, la media y la mediana, indican que es la muestra del grupo de Nativos la que presenta el valor más alto con una diferencia respecto a las otras dos muestras («NN-M1-1D»; «NN-M2-1D») de más de quince y diez puntos respectivamente, como podemos ver en la tabla de estadísticos. Observamos, asimismo, que la mayor diferencia entre media y mediana se da en la muestra del grupo «N-1D» con casi dos puntos; mientras que es en la muestra del grupo de no Nativos M2 donde encontramos la diferencia más baja, no llega a un punto.

La dispersión de las muestras la comprobamos a través de los valores del estadístico «desviación típica». Podemos observar que los valores de las dos muestras del grupo de Nativos, en el M1 y en el M2, están muy próximos entre sí (M1: s=5.2207; M2: s=5.3465). Sin embargo, el valor de la muestra del grupo «N-1D», s=8.5382, se distancia de las otras

dos en más de tres puntos. Estas pequeñas diferencias las podemos visualizar en el gráfico de dispersión que mostramos a continuación.

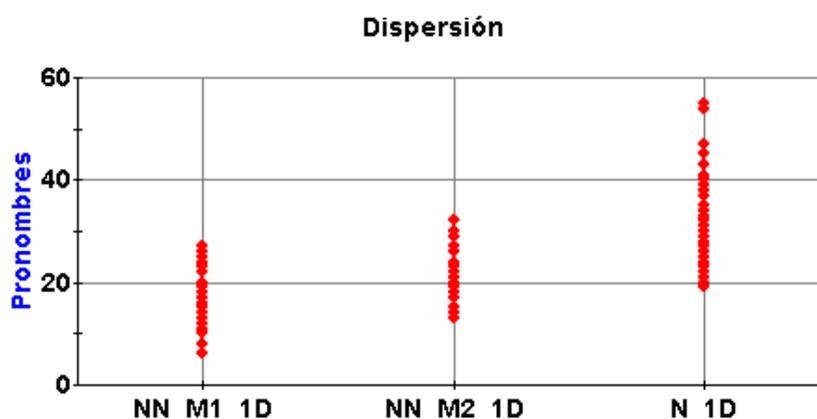


Gráfico 148. Gráfico de dispersión para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D

En cuanto al otro descriptor que mide la dispersión, el «rango intercuartílico», vemos que las tres muestras presentan un valor muy próximo entre sí. El valor más alto (IQR=12.0000) corresponde a la muestra del grupo N, mientras que el valor más bajo (IQR=8.0000) lo observamos en la muestra del M1 del grupo de no Nativos. La muestra del grupo de NN M2 presenta un valor IQR=9.0000. Podemos visualizar estos datos, con sus pequeñas diferencias, en los diagramas de cajas en el gráfico 149.

También podemos ver la distancia entre la media y la mediana que presenta cada una de las muestras. Como hemos señalado más arriba, la diferencia en las tres muestras era muy similar y en el diagrama de cajas lo podemos constatar. No obstante, observamos que es la muestra del grupo de Nativos la que presenta una distancia mayor entre las dos medidas de centralización. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

Al observar en el diagrama de cajas (Gráfico 149) el descriptor «máximo», llama la atención la gran diferencia que se puede ver entre la muestra del grupo de Nativos y las otras dos muestras del grupo de no Nativos. Las diferencias son de 28 puntos con el M1 y de 23 con el M2.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es negativo en las dos muestras del grupo de no Nativos: «NN-M1-1D», curtosis=-0.5631, y «NN-M2-1D», curtosis=-0.9361. Mientras que la muestra del grupo de Nativos presenta un valor positivo (curtosis=0.6445). Respecto a una curtosis cero, distribución

normal o campana de Gauss, las muestras que presentan valores negativos («NN-M1-1D»; «NN-M2-1D») tienden a una curva más plana, mientras que la muestra que presenta el valor positivo («N-1D») tiende a una curva más aguda. Podemos visualizar estos datos en los histogramas siguientes.

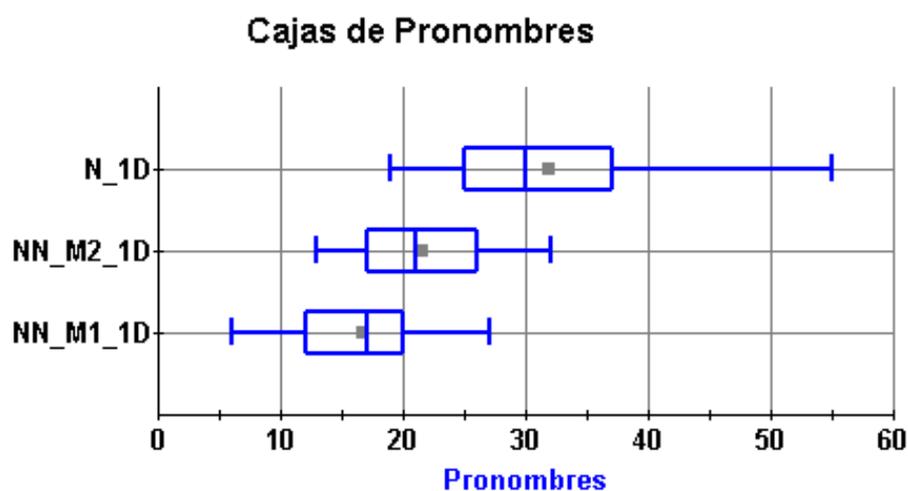


Gráfico 149. Diagrama de cajas para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D

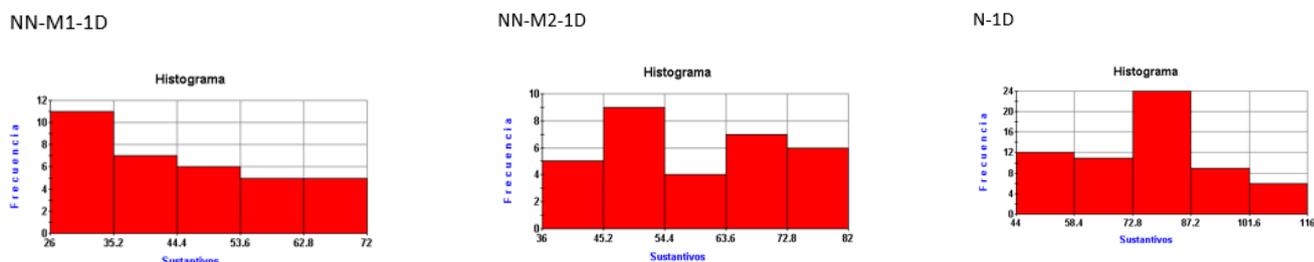


Gráfico 150. Histograma para la variable «Pronombres» NN-M1-1D / NN-M2-1D / N-1D

(8) Preposiciones: Totales de preposiciones.

	Muestras «1D»		
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Preposiciones	175	178	619

Observamos que la diferencia entre las dos muestras del grupo de no Nativos en esta variante es mínima, aunque sigue habiendo un número mayor en la muestra del M2 respecto a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos totales.

**Estadísticos para la variable Preposiciones**

<b>Grupos</b>	<b>NN_M1_1D</b>	<b>NN_M2_1D</b>	<b>N_1D</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>13.1176</b>	<b>15.0000</b>	<b>23.2581</b>
<b>Mediana</b>	<b>12.5000</b>	<b>15.0000</b>	<b>24.5000</b>
<b>Varianza</b>	<b>30.4706</b>	<b>22.1333</b>	<b>42.2274</b>
<b>Desviación Típica</b>	<b>5.5200</b>	<b>4.7046</b>	<b>6.4983</b>
<b>Mínimo</b>	<b>6.0000</b>	<b>4.0000</b>	<b>9.0000</b>
<b>Máximo</b>	<b>31.0000</b>	<b>22.0000</b>	<b>38.0000</b>
<b>Cuartil Inferior</b>	<b>9.0000</b>	<b>11.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>16.0000</b>	<b>19.0000</b>	<b>27.0000</b>
<b>Rango Intercuartilico</b>	<b>7.0000</b>	<b>8.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>2.2395</b>	<b>-0.4260</b>	<b>-0.4241</b>

Tabla 71. Estadísticos para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D

Al observar los valores de los descriptores que señalan las medidas de centralización, la media y la mediana, vemos que es la muestra del grupo de Nativos la que presenta el valor más alto con una diferencia respecto a las otras dos muestras («NN-M1-1D»; «NN-M2-1D») de más de diez y ocho puntos respectivamente, como podemos ver en la tabla de estadísticos. Observamos, asimismo, que la mayor diferencia entre media y mediana se da en la muestra del grupo «N-1D» con algo más de un punto; en la muestra del grupo de NN M1 la diferencia no llega a un punto, y llama la atención la coincidencia de valores en la muestra del M2 del grupo de no Nativos, como podemos comprobar en el gráfico de cajas (Gráfico 152).

La dispersión de las muestras la comprobamos a través de los valores del estadístico «desviación típica». Podemos observar que los valores de las tres muestras están próximos entre sí. El valor más alto lo encontramos en la muestra del grupo de Nativos ( $s=6.4983$ ). El valor más bajo corresponde a la muestra del grupo de no Nativos M2 ( $s=4.7046$ ). La muestra del grupo NN M1 presenta un valor  $s=5.5200$ . Al visualizar el gráfico de dispersión podemos constatar la menor dispersión de la muestra del grupo «NN-M2-1D» (Gráfico 151).

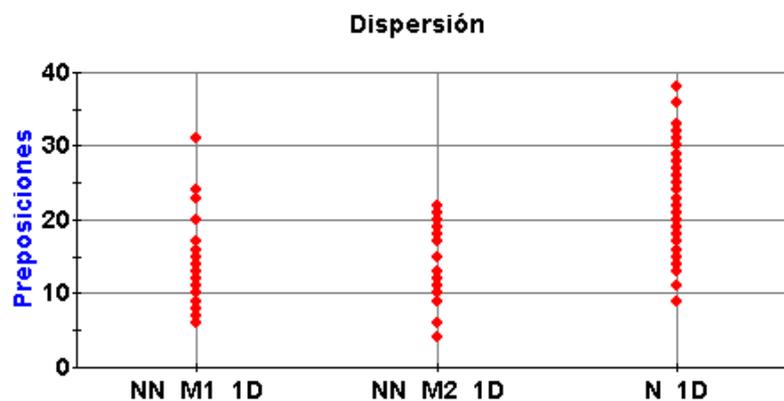


Gráfico 151. Gráfico de dispersión para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D

En cuanto al otro descriptor que mide la dispersión, el «rango intercuartílico», vemos que las tres muestras presentan un valor muy próximo entre sí. El valor más alto (IQR=9.0000) corresponde a la muestra del grupo Nivo, mientras que el valor más bajo (IQR=7.0000) lo observamos en la muestra del M1 del grupo de no Nativos. La muestra del grupo de no Nativos M2 presenta un valor IQR=8.0000. Podemos visualizar estos datos, con sus pequeñas diferencias, en los diagramas de cajas que mostramos a continuación y donde podemos ver con más claridad el tamaño de cada una de las cajas.

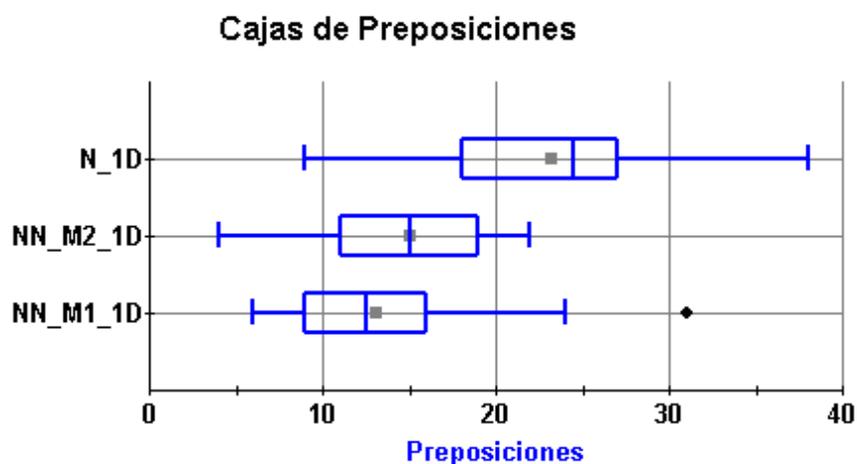


Gráfico 152. Diagrama de cajas para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D

Al observar los diagramas de cajas, constatamos, como hemos señalado más arriba, la coincidencia de las medidas de centralización, media y mediana, en la muestra «NN-M2-1D». Esto indica que la muestra tiene una distribución normal. También podemos que es la muestra del grupo de Nativos la que presenta una distancia mayor, aunque pequeña, entre

las dos medidas de centralización. Así como también al muestra del grupo de no Nativos M1. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

Al observar en el diagrama de cajas el descriptor «máximo», llama la atención la diferencia que se puede ver entre la muestra del grupo de Nativos y las otras dos muestras del grupo de no Nativos. Las diferencias son de dieciséis puntos con el M2 y de siete con el M1.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es positivo en la muestra: «NN-M1-1D», curtosis=2.2395. Mientras que las otras dos muestras presentan un valor negativo («NN-M2-1D», curtosis= -0.4260; «N-1D», curtosis= -0.4241). Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan valores negativos («NN-M2-1D»; «N-1D») tienden a una curva más plana, mientras que la muestra que presenta el valor positivo («NN-M1-1D») tiende a una curva más aguda. Podemos visualizar estos datos en los histogramas siguientes.

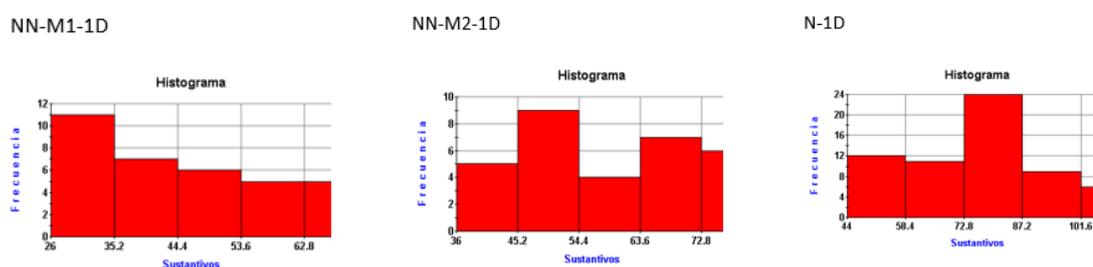


Gráfico 153. Histograma para la variable «Preposiciones» NN-M1-1D / NN-M2-1D / N-1D

(9) Artículos: Totales de artículos.

Muestras «ID»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Artículos	74	77	223

Observamos nuevamente la pequeña diferencia entre las dos muestras del grupo de no Nativos. Vemos que la muestra del M2 es algo superior a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos totales.

**Estadísticos para la variable Artículos**

<b>Grupos</b>	<b>NN_M1_1D</b>	<b>NN_M2_1D</b>	<b>N_1D</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>7.9118</b>	<b>10.9355</b>	<b>15.4032</b>
<b>Mediana</b>	<b>7.5000</b>	<b>11.0000</b>	<b>14.0000</b>
<b>Varianza</b>	<b>11.4162</b>	<b>13.5957</b>	<b>28.3757</b>
<b>Desviación Típica</b>	<b>3.3788</b>	<b>3.6872</b>	<b>5.3269</b>
<b>Mínimo</b>	<b>2.0000</b>	<b>5.0000</b>	<b>6.0000</b>
<b>Máximo</b>	<b>16.0000</b>	<b>18.0000</b>	<b>33.0000</b>
<b>Cuartil Inferior</b>	<b>5.0000</b>	<b>8.0000</b>	<b>12.0000</b>
<b>Cuartil Superior</b>	<b>11.0000</b>	<b>14.0000</b>	<b>19.0000</b>
<b>Rango Intercuartílico</b>	<b>6.0000</b>	<b>6.0000</b>	<b>7.0000</b>
<b>Curtosis</b>	<b>-0.4020</b>	<b>-0.8643</b>	<b>0.8238</b>

Tabla 72. Estadísticos para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D

Las medidas de centralización, la media y la mediana, son muy diferentes en cada una de las tres muestras. El valor más alto lo encontramos en la muestra del grupo de Nativos con una media=15.4032 y una mediana=14.0000. El valor más bajo lo presenta la muestra del grupo de no Nativos en el M1, con una diferencia respecto al grupo de Nativos de más de siete puntos para la media y de más de seis puntos para la mediana. No obstante, el dato que llama la atención aquí es la proximidad entre los valores de media y mediana de cada una de las muestras como se puede ver en la tabla de estadísticos.

El siguiente estadístico que nos ayuda a entender cómo se distribuyen las muestras, y que indica la dispersión, es la desviación típica. Como podemos ver en la tabla de estadísticos, las dos muestras del grupo de no Nativos tienen valores muy próximos: «NN-M1-1D»,  $s=3.3788$ ; «NN-M2-1D»,  $s=3.6872$ . Mientras que la muestra del grupo de Nativos presenta un valor más alto: «N-1D»,  $s=5.3269$ . Así, las dos muestras del grupo de no Nativos presentan menor dispersión que la muestra del grupo de Nativos. Estos datos los podemos visualizar en el gráfico de dispersión (Gráfico 154).

En cuanto al otro descriptor que mide la dispersión, el «rango intercuartílico», vemos que las dos muestras del grupo de no Nativos («NN-M1-1D»; «NN-M2-1D») presentan el mismo valor ( $IQR=6.0000$ ). Mientras que el grupo de Nativos presenta un valor de solo un punto más ( $IQR=7.0000$ ). Podemos visualizar estos datos en los diagramas de cajas (Gráfico 155).

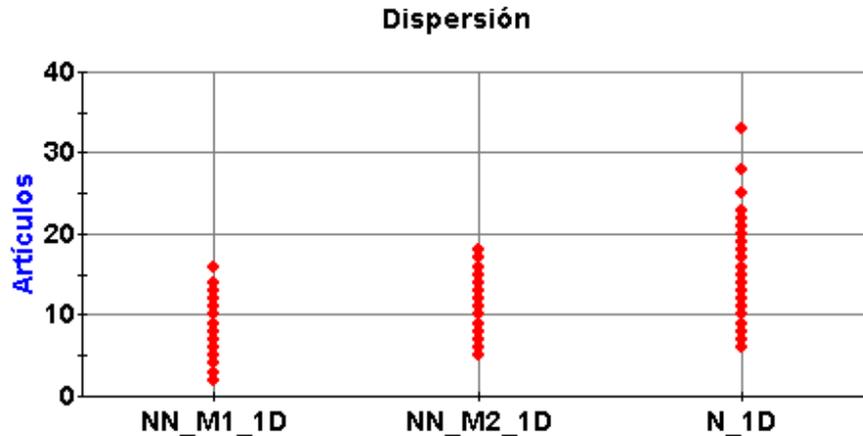


Gráfico 154. Gráfico de dispersión para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D

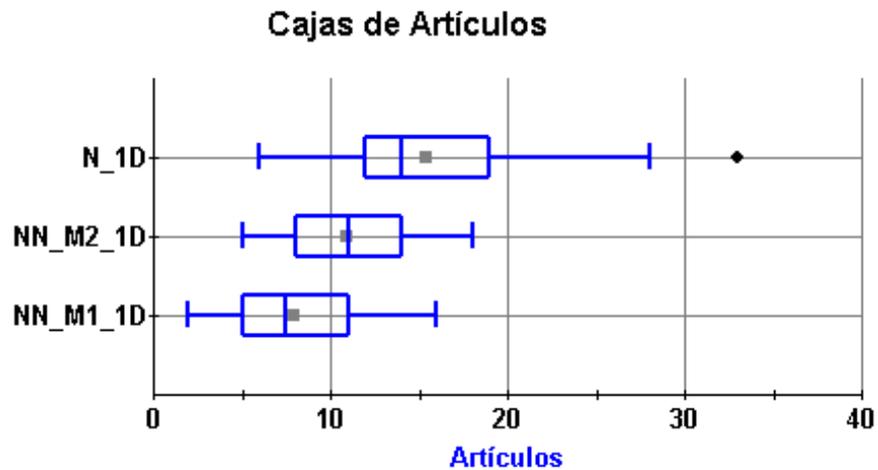


Gráfico 155. Diagrama de cajas para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D

Al observar los diagramas de cajas, constatamos, por un lado, como hemos señalado más arriba, la coincidencia de tamaño de las cajas, indicado por el rango intercuartílico, de las dos muestras del grupo de no Nativos. Por otro lado, la casi coincidencia entre las medidas de centralización, media y mediana, en la muestra «NN-M2-1D». También podemos ver que es la muestra del grupo de Nativos la que presenta una distancia mayor, aunque pequeña, entre las dos medidas de centralización. Así como también la muestra del grupo de no Nativos M1. Estas pequeñas diferencias indican que las muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

Al observar en el diagrama de cajas el descriptor «máximo», llama la atención la diferencia que se puede ver entre la muestra del grupo de Nativos y las otras dos muestras

del grupo de no Nativos. Las diferencias son de diecisiete puntos con el M1 y de quince con el M2.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» es positivo en la muestra: «N-1D»,  $\text{curtosis}=0.8238$ . Mientras que las otras dos muestras presentan un valor negativo («NN-M1-1D»,  $\text{curtosis}= -0.4020$ ; «NN-M2-1D»,  $\text{curtosis}= -0.8643$ ). Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan valores negativos («NN-M1-1D»; «NN-M2-1D») tienden a una curva más plana, mientras que la muestra que presenta el valor positivo («N-1D») tiende a una curva más aguda. Podemos visualizar estos datos en los histogramas siguientes.

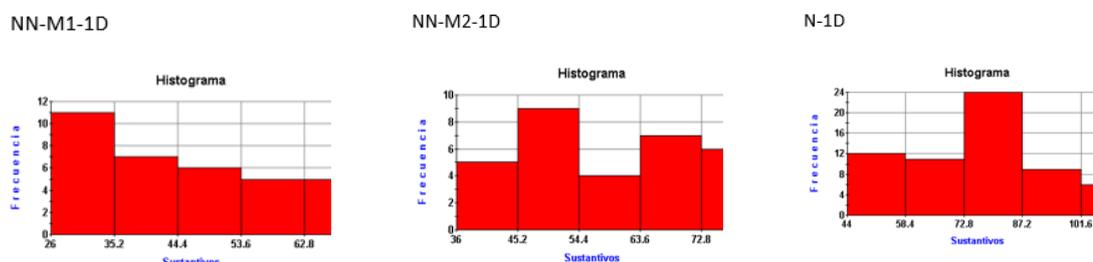


Gráfico 156. Histograma para la variable «Artículos» NN-M1-1D / NN-M2-1D / N-1D

(10) Conjunciones: Totales de conjunciones.

Muestras «ID»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
Conjunciones	121	137	620

Observamos que la diferencia entre las dos muestras del grupo de no Nativos en esta variante es pequeña, aunque sigue habiendo un número mayor en la muestra del M2 respecto a la del M1. A continuación vamos a ver qué dicen los estadísticos a partir de estos totales (Tabla 73).

Las medidas de centralización, la media y la mediana, son muy diferentes en cada una de las tres muestras. El valor más alto lo encontramos en la muestra del grupo de Nativos con una  $\text{media}=23.0323$  y una  $\text{mediana}=23.0000$ . El valor más bajo lo presenta la muestra del grupo de no Nativos en el M1, con una diferencia respecto al grupo de Nativos de más de doce puntos para la media y de trece puntos para la mediana. No obstante, el dato que llama la atención aquí es la proximidad entre los valores de media y mediana de cada una de

las muestras como se puede ver en la tabla de estadísticos, pero sobre todo de la muestra del grupo de Nativos.

**Estadísticos para la variable Conjunciones**

<b>Grupos</b>	<b>NN_M1_1D</b>	<b>NN_M2_1D</b>	<b>N_1D</b>
<b>N</b>	<b>34</b>	<b>31</b>	<b>62</b>
<b>Media</b>	<b>10.5000</b>	<b>13.5484</b>	<b>23.0323</b>
<b>Mediana</b>	<b>10.0000</b>	<b>12.0000</b>	<b>23.0000</b>
<b>Varianza</b>	<b>32.4394</b>	<b>42.4559</b>	<b>44.1629</b>
<b>Desviación Típica</b>	<b>5.6956</b>	<b>6.5158</b>	<b>6.6455</b>
<b>Mínimo</b>	<b>3.0000</b>	<b>4.0000</b>	<b>11.0000</b>
<b>Máximo</b>	<b>27.0000</b>	<b>33.0000</b>	<b>46.0000</b>
<b>Cuartil Inferior</b>	<b>6.0000</b>	<b>10.0000</b>	<b>18.0000</b>
<b>Cuartil Superior</b>	<b>13.0000</b>	<b>17.0000</b>	<b>27.0000</b>
<b>Rango Intercuartílico</b>	<b>7.0000</b>	<b>7.0000</b>	<b>9.0000</b>
<b>Curtosis</b>	<b>1.0109</b>	<b>1.9795</b>	<b>0.9065</b>

Tabla 73. Estadísticos para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D

La dispersión de las muestras la comprobamos a través de los valores del estadístico «desviación típica». Podemos observar que los valores de las muestras «NN-M2-1D» ( $s=6.5158$ ) y «N-1D» ( $s=6.6455$ ) están muy próximos entre sí. El valor más bajo corresponde a la muestra del grupo de no Nativos M1 ( $s=5.6956$ ). En el gráfico 157 podemos constatar la menor dispersión de la muestra del grupo «NN-M1-1D».

En cuanto al otro descriptor que mide la dispersión, el «rango intercuartílico», vemos que las dos muestras del grupo de no Nativos («NN-M1-1D»; «NN-M2-1D») presentan el mismo valor ( $IQR=7.0000$ ). Mientras que el grupo de Nativos presenta un valor de solo un punto más ( $IQR=9.0000$ ). Podemos visualizar estos datos en los diagramas de cajas que mostramos a continuación.

Al observar los diagramas de cajas, constatamos, por un lado, como hemos señalado más arriba, la coincidencia de tamaño de las cajas, indicado por el rango intercuartílico, de las dos muestras del grupo de no Nativos («NN-M1-1D»; «NN-M2-1D»). Por otro lado, la casi total coincidencia entre las medidas de centralización, media y mediana, en la muestra «N-1D». También podemos ver que es la muestra del grupo «NN-M2-1D» la que presenta una distancia mayor, aunque pequeña, entre las dos medidas de centralización. Así como también la muestra del grupo de no Nativos M1. Estas pequeñas diferencias indican que las

muestras presentan una distribución con tendencia a alejarse de la distribución normal o campana de Gauss.

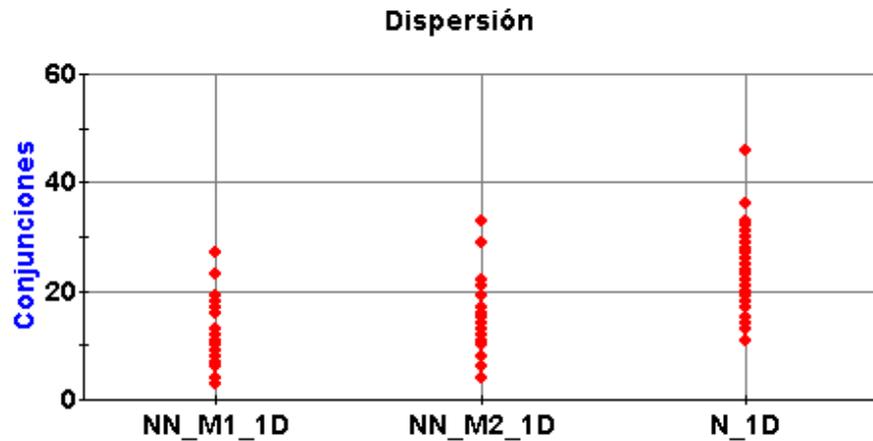


Gráfico 157. Gráfico de dispersión para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D

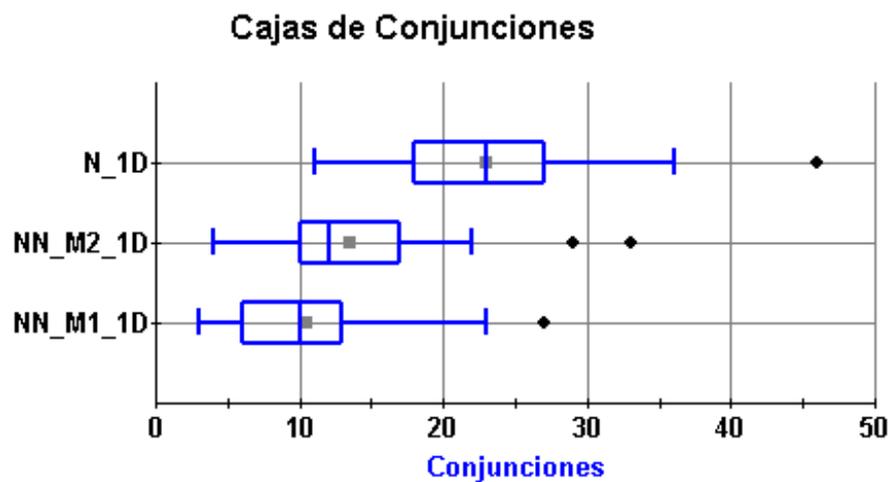


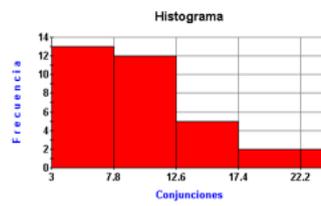
Gráfico 158. Diagrama de cajas para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D

Al observar en el diagrama de cajas (Gráfico 158) el descriptor «máximo», llama la atención la diferencia que se puede ver entre la muestra del grupo de Nativos y las otras dos muestras del grupo de no Nativos. Las diferencias son de dieciocho puntos con el M1 y de trece con el M2.

En cuanto a la forma de la distribución de la muestra, vemos que el valor del estadístico «curtosis» (Tabla 73) es positivo en las tres muestras: «NN-M1-1D», curtosis=1.0109; «NN-M2-1D», curtosis=1.9795; «N-1D», curtosis=0.9065. Respecto a una curtosis cero, distribución normal o campana de Gauss, las muestras que presentan valores positivos («N-

1D») tienden a una curva más aguda. Podemos visualizar estos datos en los histogramas siguientes.

NN-M1-1D



NN-M2-1D



N-1D



Gráfico 159. Histograma para la variable «Conjunciones» NN-M1-1D / NN-M2-1D / N-1D

## 6.2 Resultados y análisis de la «Estadística inferencial»

### 6.2.1 Resultados y análisis de la comparación del texto «F» y del texto «1D»

Para dar respuesta a las preguntas de investigación uno y dos, en este apartado nos vamos a centrar, por un lado, en las características léxicas y estructurales que presentan los dos tipos de muestras, la «F» (inventar una historia) y la «1D» (explicar una experiencia personal) y, por otro lado, en las diferencias o semejanzas que presentan las producciones de los sujetos no nativos (NN) y los sujetos nativos (N) entre sí.

En la pregunta uno, vamos a tomar todas las producciones de las muestras «F» y «1D» en conjunto, sin distinguir si las producciones escritas fueron realizadas por sujetos N o NN. La finalidad es comprobar si los dos tipos de texto se caracterizan por presentar un léxico y unas estructuras diferentes.

En la pregunta dos, vamos a comparar las producciones de los sujetos NN y los N en conjunto, es decir, sin distinguir entre tipo de texto. La finalidad es ver qué diferencias o semejanzas se dan entre ambos grupos de sujetos.

**Pregunta 1: ¿Hay diferencias léxicas o sintácticas que distingan los relatos en pasado en los que se exige más creatividad (inventar una historia) frente a los que exigen cierta introspección (relatar una experiencia personal)?**

Para responder a esta pregunta de investigación, vamos a utilizar la aplicación informática AntConc y las herramientas (a) *WordList*, para obtener un listado de las palabras más frecuentes; (b) *Cluster*, para identificar si dos o más palabras aparecen juntas en el corpus y con qué frecuencia, y (c) *Concordance*, para ver palabras concretas en contexto.

Así, el análisis que presentamos a continuación se estructura en tres fases:

- (i) Para empezar, vamos a ver cuáles son las 50 primeras palabras más frecuentes en general en las dos muestras, «F» y «1D» y también en el corpus CREA (Corpus de Referencia del Español Actual); después mostraremos las 50 primeras palabras más frecuentes, pero teniendo en cuenta una lista de *stopwords* (palabras que no deseamos cuantificar). De esta manera, podremos centrarnos en las palabras léxicas más frecuentes y también en la presencia de alguna palabra funcional de interés para nuestra investigación, como la

presencia del pronombre «yo» o de los posesivos de primera persona del singular. De esta manera podremos ver si hay diferencia o no entre una muestra y otra.

- (ii) En la segunda fase, y a partir de las palabras léxicas más frecuentes y las funcionales elegidas, veremos de qué manera se unen para comprobar si se produce el mismo tipo de agrupamientos en ambas muestras.
- (iii) Finalmente, también a partir de las palabras léxicas más frecuentes y las funcionales elegidas, observaremos en qué tipo de contexto suelen aparecer y si coincide o no en ambas muestras.

### Fase 1: Frecuencias.

Como ya hemos señalado anteriormente, aquí nos vamos a centrar en qué palabras son las más frecuentes en las muestras «F» y «1D» y vamos a presentar como punto de comparación el corpus CREA. Para realizar la comparativa solo vamos a considerar las 50 primeras posiciones de los tres corpus. En la siguiente tabla presentamos dichas palabras.

Muestra «F»			Muestra «1D»			Corpus CREA		
1	1578	a	1	1147	a	1.	de	9,999,518
2	1288	de	2	890	de	2.	la	6,277,560
3	1121	y	3	617	la	3.	que	4,681,839
4	1071	que	4	574	que	4.	el	4,569,652
5	902	la	5	570	y	5.	en	4,234,281
6	853	el	6	435	en	6.	y	4,180,279
7	660	mi	7	393	clase	7.	a	3,260,939
8	608	en	8	371	me	8.	los	2,618,657
9	579	me	9	347	mi	9.	se	2,022,514
10	561	no	10	343	el	10.	del	1,857,225
11	551	un	11	276	d	11.	las	1,686,741
12	462	con	12	230	no	12.	un	1,659,827
13	445	por	13	228	un	13.	por	1,561,904
14	419	una	14	224	muy	14.	con	1,481,607
15	401	estaba	15	214	primer	15.	no	1,465,503
16	325	cuando	16	178	fue	16.	una	1,347,603
17	317	se	17	171	una	17.	su	1,103,617
18	281	muy	18	170	n	18.	para	1,062,152
19	279	pero	19	160	con	19.	es	1,019,669
20	276	canelo	20	155	s	20.	al	951,054
21	266	para	21	143	era	21.	lo	866,955
22	248	yo	22	141	los	22.	como	773,465
23	231	perro	23	136	m	23.	más	661,696
24	225	al	24	129	profesora	24.	o	542,284
25	218	lo	25	126	se	25.	pero	450,512
26	210	era	26	122	para	26.	sus	449,870
27	202	los	27	119	pero	27.	le	413,241
28	183	hab	28	117	yo	28.	ha	380,339
29	179	su	29	108	al	29.	me	374,368
30	169	domingo	30	108	las	30.	si	327,480
31	157	casa	31	106	por	31.	sin	298,383
32	157	repente	32	95	hab	32.	sobre	289,704

33	155	s	33	94	nos	33.	este	285,461
34	146	fue	34	92	cuando	34.	ya	274,177
35	145	las	35	91	compa	35.	entre	267,493
36	145	paseando	36	91	estaba	36.	cuando	257,272
37	140	amiga	37	91	poco	37.	todo	247,340
38	134	n	38	87	o	38.	esta	238,841
39	132	le	39	86	lo	39.	ser	232,924
40	131	porque	40	85	es	40.	son	232,415
41	129	como	41	85	ya	41.	dos	228,439
42	118	es	42	81	conoce	42.	también	227,411
43	114	o	43	78	fué	43.	fue	223,791
44	113	d	44	76	porque	44.	había	223,430
45	113	ella	45	75	eros	45.	era	219,933
46	113	nos	46	75	ten	46.	muy	208,540
47	112	del	47	74	como	47.	años	203,027
48	111	mucho	48	74	mucho	48.	hasta	202,935
49	105	m	49	73	bien	49.	desde	198,647
50	88	pasado	50	65	espa	50.	está	194,168

Tabla 74. Las 50 palabras más frecuentes. M-F / M-1D

La información que aporta la herramienta *WordList* del AntConc se organiza, como vemos, por columnas: en la primera aparece el número que indica la posición de la palabra más frecuente dentro del corpus; en la segunda aparece el número de veces que aparece la palabra en el corpus y en la tercera aparece la palabra a la que hace referencia la información anterior. La columna dedicada al corpus CREA invierte el orden de la palabra y la frecuencia absoluta, pues es así el *output* que ofrece la página de la RAE (Real Academia Española).

Encontramos casos de palabras incompletas o una sola letra en las muestras «F» y «1D», esto es así porque en el caso de los textos de los sujetos NN es normal que haya errores en sus escritos.

Dentro de la lista de las 50 primeras palabras, podemos observar que la palabra más frecuente en las dos muestras que forman el corpus de esta investigación es la preposición «a», seguida de la preposición «de»; mientras que en el corpus CREA es la preposición «de» la primera, y la preposición «a» no la encontramos hasta la posición siete. Vemos, asimismo, que en la muestra «F», entre las diez palabras más frecuentes, no hay ninguna palabra léxica. Este hecho coincide con la investigación que hacen, por un lado, Cantos y Sánchez (2011) sobre el corpus del español (Cumbre) y el inglés (Lacell) y, por otro lado, Muñoz Lobo (2012) sobre el corpus CREA y el corpus de la Universidad de Brigham Young, en los dos estudios se señala que las palabras más frecuentes son las palabras funcionales. Hecho que confirmamos al observar la columna del corpus CREA, en el cual no es hasta la posición 39 que encontramos las primeras formas verbales, seis en total, que corresponden a los verbos «ser», «estar» y «haber», dos copulativos y un auxiliar; es decir, verbos que no aportan

contenido semántico. Los dos primeros, los copulativos, sirven simplemente de nexo de unión entre el sujeto y el predicado de la oración, y el verbo auxiliar se limita a aportar información gramatical.

Sí encontramos una PL, el sustantivo «clase», en la muestra «1D» en la posición siete. Pero no la debemos tener en consideración, pues forma parte del enunciado del texto que tenían que escribir. En la muestra «F», el primer sustantivo lo encontramos en la posición veinte, y en el corpus CREA, el primer sustantivo («años») aparece en la posición 47. Llama la atención que en el trabajo de Cantos y Sánchez (*op. cit.*) señalan que no es hasta la posición 60 cuando aparece el primer sustantivo.

Para comprobar si el tipo de léxico de las muestras «F» y «1D» se asemeja o no, vamos a presentar las primeras 50 palabras más frecuentes, pero sin tener en cuenta un listado de *stopwords* (palabras que no nos interesa cuantificar) como las palabras que formaban parte de los enunciados, artículos o preposiciones.

Muestra «F»			Muestra «1D»		
1	660	mi	1	371	me
2	579	me	2	347	mi
3	401	estaba	3	178	fue
4	248	yo	4	129	profesora
5	157	casa	5	117	yo
6	146	fue	6	94	nos
7	113	nos	7	91	estaba
8	88	pasado	8	91	poco
9	84	ana	9	64	curso
10	83	iba	10	62	mis
11	79	mis	11	61	escuela
12	76	hombre	12	49	idioma
13	74	dos	13	47	francés
14	65	amigos	14	46	gente
15	65	vi	15	46	todos
16	63	tiempo	16	45	nerviosa
17	62	dijo	17	43	día
18	62	fui	18	43	lengua
19	61	hacer	19	42	alumnos
20	55	entonces	20	41	aprender
21	55	gente	21	40	muchos
22	53	momento	22	39	clases
23	52	día	23	36	primera
24	52	todos	24	36	recuerdo
25	52	vez	25	34	inglés
26	51	ir	26	33	chica
27	50	después	27	33	nivel
28	49	miedo	28	31	estudiar
29	49	poco	29	31	también
30	47	donde	30	30	muchas
31	47	tranquilamente	31	30	ser
32	45	chico	32	29	cada
33	43	siempre	33	27	llegué
34	42	después	34	27	mucha

35	42	madre	35	27	nuevo
36	42	tarde	36	27	nunca
37	40	fiesta	37	27	tiempo
38	40	había	38	26	interesante
39	39	esperando	39	26	primero
40	39	lado	40	25	hablar
41	39	solo	41	24	amigos
42	39	tenía	42	24	dos
43	39	vida	43	24	durante
44	38	fuimos	44	23	aula
45	38	hora	45	23	casa
46	38	minutos	46	23	cosas
47	38	tambien	47	23	después
48	37	cosas	48	23	diferentes
49	36	coche	49	23	otros
50	36	ver	50	22	mejor

Tabla 75. Las 50 palabras más frecuentes con lista de *stopwords*. M-F / M-1D

Al comparar las dos listas de palabras más frecuentes, vemos que en la muestra «1D» (recordar el primer día de clase de un idioma) es fácil deducir que el tema del texto está relacionado con el campo semántico de la educación, ya que la mayor parte de los sustantivos que aparecen (profesora, curso, escuela, idioma, alumnos, clases, etc.), así como la presencia de los verbos «aprender» y «estudiar» hacen referencia a dicho campo semántico. Esto ratifica las palabras de Mar Cruz (2012:39) que dice que “conocer las palabras léxicas más frecuentes de un documento es una manera rápida y automatizable de obtener datos sobre el contenido general de un texto”. En cambio, en la muestra «F», como los sujetos tenían que inventar la historia a partir de una frase, el tema era más abierto, por tanto, no es posible encontrar un hilo conductor en el léxico que utilizan pues este es más variado.

Llama la atención, no obstante, que dentro de las quince primeras posiciones, aparezcan las mismas cinco palabras en las dos muestras: los posesivos «mi», «mis»; los pronombres personales complemento «me» y «nos», y el pronombre personal sujeto «yo». Más adelante analizaremos cómo se agrupan estas palabras en una y otra muestra y dentro de qué contexto se desenvuelven.

En cuanto a los sustantivos que aparecen dentro de las 50 primeras posiciones, vemos que en ambas muestras hay un total de diecisiete y, de ellos, seis se repiten en las dos muestras. Estos sustantivos son: casa (posiciones 6 y 46), amigos (posiciones 15 y 42), tiempo (posiciones 17 y 38), gente (posiciones 22 y 15), día (posiciones 24 y 18), cosas (posiciones 49 y 47).

## Fase 2: Agrupamientos.

En esta parte del análisis, nos vamos a centrar en las palabras que hemos destacado en la fase anterior: *mi, mis, yo, me, nos, casa, amigos, tiempo, gente, día, cosas*. Vamos a ver si se agrupan de la misma manera en las dos muestras.

A continuación mostramos los agrupamientos de la forma átona del posesivo «mi».

«F»		«1D»	
<i>mi</i>		<i>mi</i>	
12	4	mi amiga y yo	3 7 mi primera clase de
13	3	a mi amiga que	4 4 se sentó a mi
14	3	a mi perro y	5 4 sentó a mi lado
15	3	al lado de mi	6 2 a mi primera clase
16	3	una amiga de mi	7 2 aumento y mi interés
17	2	a celebrar mi cumpleaños	8 2 cerca de mi casa
18	2	a lado de mi	9 2 de idiomas de mi
19	2	a mi amiga y	10 2 de mi elección
20	2	a mi novio para	11 2 de mi pueblo, pero
21	2	amiga de mi hermana	12 2 desconocida para mi. Aunque
22	2	Canelo a mi lado	13 2 Desde ese momento mi
23	2	cuando de repente mi	14 2 el habla. Por mi
24	2	de mi cumpleaños	15 2 en aumento y mi
25	2	de mi perro y	16 2 En mi primer d
26	2	de repente mi perro	17 2 ese momento mi pasión
27	2	di cuenta que mi	18 2 ha sido mi mejor
28	2	dicho que mi padre	19 2 habla. Por mi parte
29	2	en frente de mi	20 2 idiomas de mi pueblo
30	2	era mi cumpleaños, pero	21 2 lengua desconocida para mi
31	2	fiesta en mi casa	22 2 mi interés ha traspasado
32	2	lado de mi casa	23 2 mi parte fue una
33	2	mejor amiga y mi	24 2 mi pasión por
34	2	mi amiga llegó y	25 2 mi pueblo, pero en
35	2	mi cumpleaños y	26 2 mi. Aunque al llegar
36	2	mi madre y padre	27 2 momento mi pasión
37	2	mi perro estaba llamando	28 2 para mi. Aunque al
38	2	orvido de mi cumpleaños	29 2 Por mi parte fue
39	2	que mi padre esta	30 2 sido mi mejor amiga
40	2	repente vi a mi	31 2 y mi interés ha

Tabla 76. Lista de agrupamientos del posesivo «mi» M-F / M-1D

Recordemos, antes de nada, que la función de las formas átonas de los posesivos es vincular el sustantivo al que acompañan con la persona del discurso (*yo, tú, nosotros, vosotros*); y que sus formas se deben a la concordancia de número y género con dicho sustantivo. En este caso, como se trata de la primera persona del singular, la concordancia se limita al número (*mi, mis*).

Asimismo, el utilizar una forma átona de posesivo significa que entre el sustantivo y la persona del discurso, «yo» en nuestro caso, hay una relación que, dependiendo del

contexto en el que se desarrolle el discurso, puede interpretarse como posesión, parentesco, relación de pertenencia a un grupo, etc.

En la tabla 76 vemos los agrupamientos más frecuentes de la forma átona del posesivo «mi». Como podemos ver, hay formas que corresponden a la forma tónica y no a la átona, pues si los sujetos no acentúan la palabra, el programa no discrimina entre una forma u otra.

Como se puede ver también, hemos obviado las primeras posiciones, pues remitían a la instrucción de la tarea. Así, hemos seleccionado desde la posición en la que ya no era un calco exacto del enunciado.

Observamos en la tabla de agrupamientos que, mientras que en la muestra «F» predomina el uso de la forma átona del posesivo «mi» dentro de una función referencial para expresar posesión («mi perro», posición 14-«F») o parentesco («mi hermana», posición 21-«F»); en la muestra «1D», llama la atención la combinación que se realiza con los sustantivos «interés» (posiciones 7, 22 y 31-«1D»); «pasión» (posiciones 17, 24 y 27-«1D»), y «elección» (posición 10-«1D»). Agrupamientos que se inscriben dentro de la función expresiva del lenguaje al denotar implicación personal y emocional por el significado semántico de los sustantivos elegidos; se está haciendo uso, claramente, de un léxico valorativo. Aquí el sujeto-emisor es el punto de referencia y como tal expresa su postura ante algo. No obstante, encontramos también usos de la forma átona «mi» con función referencial para expresar posesión («mi casa», posición 8-«1D»), parentesco («mi mejor amiga», posición 30-«1D») o expresar relación con respecto a un grupo («mi pueblo», posición 11, 20 y 25-«1D»).

En cuanto a la forma átona del posesivo «mis», vemos en la tabla 77 que en ambas muestras es la función referencial del lenguaje la que predomina en los agrupamientos presentes en las veinticinco primeras posiciones. La relación que se establece entre el sustantivo y la persona del discurso, «yo», mediante la forma átona del posesivo «mis», es principalmente de parentesco («mis amigos»; «mis padres»; «mis compañeros»). Sin embargo, volvemos a encontrar en la muestra «1D» un caso de agrupamiento más propio de la función expresiva del lenguaje («mis sensaciones», posición 25-«1D»). Nuevamente, en esta muestra se pone de manifiesto el posicionamiento del sujeto-emisor ante una situación; no se limita a informar, sino que está expresando un sentimiento que no encontramos en las combinaciones presentes en la muestra «F».

«F»			«1D»		
<i>mis</i>			<i>mis</i>		
1	2	bailar con <b>mis amigos</b>	1	5	a <b>mis compañeros</b>
2	2	<b>mis amigos</b> de infancia	2	5	<b>mis compañeros</b> y
3	2	que todos <b>mis amigos</b>	3	4	de <b>mis compañeros</b>
4	2	uno de <b>mis amigos</b>	4	3	con <b>mis compañeros</b>
5	1	a bailar con <b>mis</b>	5	3	<b>mis compañeros</b> de
6	1	a celebrar <b>mis cumpleaños</b>	6	3	<b>mis nuevos compañeros</b>
7	1	a decir a <b>mis</b>	7	2	a nadie, <b>mis</b>
8	1	a hablado de <b>mis</b>	8	2	a nadie, <b>mis compañeros</b>
9	1	a llenar <b>mis ojos</b>	9	2	<b>mis compañeros</b> eran
10	1	a <b>mis amigos</b> de	10	2	<b>mis compañeros</b> me
11	1	a <b>mis padres</b> la	11	2	nadie, <b>mis compañeros</b>
12	1	a <b>mis padres</b> . La	12	1	a Alemania con <b>mis</b>
13	1	a ocurrido. <b>Mis amigos</b>	13	1	a apuntado a <b>mis</b>
14	1	a retumbar en <b>mis</b>	14	1	a ayudar con <b>mis</b>
15	1	a todos <b>mis amigos</b>	15	1	a de todos <b>mis</b>
16	1	a uno de <b>mis</b>	16	1	a demasiado a <b>mis</b>
17	1	a verlo con <b>mis</b>	17	1	a escrito francés. <b>Mis</b>
18	1	a. Todos <b>mis amigos</b>	18	1	a Grecia con <b>mis</b>
19	1	acompañar a <b>mis</b>	19	1	a gusto con <b>mis</b>
20	1	Además, estaba <b>mis venticincos</b>	20	1	a llegar <b>mis compañeros</b>
21	1	al trabajo, de <b>mis</b>	21	1	a <b>mis amigos</b> . Me
22	1	algun sitio con <b>mis</b>	22	1	a <b>mis compañeras</b>
23	1	alta. Entonces <b>mis cumpleaños</b>	23	1	a <b>mis nuevos compañeros</b>
24	1	an de <b>mis ojos</b>	24	1	a <b>mis primeras clases</b>
25	1	an dormido <b>mis ni</b>	25	1	a perfectamente, <b>mis sensaciones</b>

Tabla 77. Lista de agrupamientos del posesivo átono «mis» M-F / M-1D

Los agrupamientos con el pronombre «me» los presentamos a continuación.

«F»			«1D»		
<i>me</i>			<i>me</i>		
1	10	cuando de repente <b>me</b>	1	4	<b>me sentí</b> un poco
2	7	que <b>me había</b>	2	3	<b>me</b> di cuenta que
3	6	<b>me</b> di cuenta de	3	3	<b>me gusta</b> mucho la
4	5	de repente <b>me</b> di	4	3	pero <b>me sentía</b>
5	5	repente <b>me</b> di cuenta	5	3	primer día <b>me</b>
6	4	<b>me</b> di cuenta que	6	3	sabía que <b>me</b>
7	4	<b>me</b> ha dicho que	7	2	a la vez <b>me</b>
8	4	y <b>me</b> dijo que	8	2	a lo que <b>me</b>
9	3	<b>me</b> dijo que no	9	2	Ahora <b>me</b> interesa la
10	3	<b>me</b> dijo que tenía	10	2	alemán <b>me</b> encanta
11	3	<b>me</b> senté en un	11	2	de alemán <b>me</b>
12	3	no <b>me</b> dijo nada	12	2	de inglés, no <b>me</b>
13	3	y <b>me</b> puse a	13	2	día <b>me</b> sentía
14	2	amiga <b>me</b> dijo que	14	2	en 2013 <b>me</b> cambié. El
15	2	cuenta de que <b>me</b>	15	2	en la clase, <b>me</b>
16	2	de que <b>me</b> había	16	2	enseguida <b>me</b> aceptaron. Al
17	2	de repente <b>me</b> encontré	17	2	entré la escuela, <b>me</b>
18	2	El <b>me</b> dijo que	18	2	Este tema <b>me</b> asusta
19	2	ella <b>me</b> dijo que	19	2	ha pasado mucho. <b>Me</b>
20	2	En ese momento <b>me</b>	20	2	inglés, no <b>me</b> resulta
21	2	lo que <b>me</b> había	21	2	la escuela, <b>me</b> sentía
22	2	<b>Me</b> di cuenta que	22	2	la lengua. Ahora <b>me</b>
23	2	<b>me</b> había olvidado	23	2	lengua. Ahora <b>me</b> interesa
24	2	<b>me</b> había visto	24	2	lo que <b>me</b> decía
25	2	<b>me</b> hizo caso y	25	2	<b>me</b> aceptaron. Al finalizar

Tabla 78. Lista de agrupamientos del pronombre complemento «me» M-F / M-1D

El pronombre complemento «me» puede funcionar como pronombre reflexivo, pronombre complemento directo o indirecto. En los agrupamientos que vemos en la tabla 78, encontramos la función de complemento indirecto, como por ejemplo en «me aceptaron»

(posición 25-«1D»), y de complemento indirecto, como en «me interesa» (posición 9-«1D»). Vemos, asimismo, que se combina con verbos pronominales como «ponerse» (posición 13-«F») o «sentirse» (posición 1-«1D»); verbos transitivos como «decir» (posición 7-«F») o «aceptar» (posición 25-«1D»), y verbos pseudoimpersonales (Campos, 2000:1564,II) como «gustar» (posición 3-«1D») o «encantar» (posición 10-«1D»).

Pero lo que realmente llama la atención de estas combinaciones es la función del lenguaje que predomina en los agrupamientos de la muestra «F» y «1D». Vemos que en la muestra «1D» las combinaciones del pronombre complemento «me», en las primeras veinticinco posiciones, remiten a la función expresiva del lenguaje, ya que hay una clara manifestación de sentimientos, de posicionamiento ante algo. La elección de verbos como «gustar», «encantar» o «interesar» así lo indican. En cambio, en la muestra «F» destaca la función referencial, es decir, aquí la intención es simplemente informar («me había olvidado», posición 23-«F»). También es destacable el agrupamiento del pronombre complemento «me» con el verbo «decir» para introducir el discurso referido. Encontramos ocho casos entre las veinticinco primeras posiciones (7, 8, 9, 10, 12, 14, 18 y 19); mientras que en la muestra «1D» no aparece ninguno.

En cuanto al pronombre complemento «nos» (tabla 79), igual que con el pronombre complemento «me», puede funcionar como pronombre reflexivo, pronombre complemento directo o indirecto. Pero, además, vemos en la tabla 79 ejemplos de su uso para expresar la reciprocidad en los casos «nos besamos», posición 20-«F»; «nos hablamos mucho», posición 15-«1D»; «nos miramos», posición 21-«1D».

Vemos nuevamente en la muestra «F» que los agrupamientos del pronombre complemento «nos» señalan un predominio de la función referencial del lenguaje. Son combinaciones que se limitan a informar de algo: «nos fuimos», posiciones 5, 8, 21-«F», o «nos preguntaron», posición 25-«F». Mientras que en la muestra «1D» los agrupamientos se inscriben dentro de la función conativa o apelativa del lenguaje. Así lo vemos en los ejemplos de las posiciones 3, 4, 7 y 10 en las que el pronombre complemento se combina con verbos de influencia como «animar» y «obligar», para expresar exhortación y obligación, respectivamente; o en las posiciones 16 y 19 en las que se combina con el verbo causativo «hacer» (Campos, 2000:1543,II). Estas combinaciones denotan una intención clara de que alguien actúe; es decir, hay un sujeto que intenta influir para que el otro haga algo. Así, la elección, por un lado, de los dos verbos de influencia, «animar» y «obligar», el primero con

un matiz positivo y el segundo con matiz negativo; y, por otro lado, del verbo causativo «hacer», con la idea también de obligar a realizar algo, no es algo vano.

«F»			«1D»		
NOS			NOS		
1	2	amiga y yo nos	1	2	adelante, la profesora nos
2	2	Canelo y yo nos	2	2	Alemania, personas que nos
3	2	cuando de repente nos	3	2	la profesora nos animó
4	2	estaba muy nervioso. Nos	4	2	nos animó a realizar
5	2	nos fuimos a un	5	2	nos transmiten sus vivencias
6	2	os que no nos	6	2	personas que nos transmiten
7	2	que nos habíamos	7	2	profesora nos animó a
8	2	y nos fuimos a	8	2	que nos transmiten sus
9	1	a abandonada, pero nos	9	2	y la profesora nos
10	1	a buscarnos y nos	10	1	a imaginado. Nos obligó
11	1	a el nos veremos	11	1	a la clase, nos
12	1	a hablar y nos	12	1	a muy agradable. Nos
13	1	a hasta llegaba, nos	13	1	a nos explicó conceptos
14	1	a medida que nos	14	1	a, la profesore, nos
15	1	a mi abrazo. Nos	15	1	abierta. Nos hablamos mucho
16	1	a y nos d	16	1	agradable. Nos hizo practicar
17	1	a y yo, nos	17	1	al alemán nos
18	1	a kala Ok Nos	18	1	al mismo tiempo nos
19	1	abandonada, pero nos servir	19	1	alemán nos hizo
20	1	abrazamos y nos besamos	20	1	algo años. Nos
21	1	abrazo. Nos fuimos a	21	1	aliviados y nos miramos
22	1	acera de enfrente. Nos	22	1	amable, nos dice se
23	1	ahora nos vemos muy	23	1	amos mucho y nos
24	1	al final nos paseamos	24	1	así que nos teníamos
25	1	alemanes y nos preguntaron	25	1	aula, nos divertimos. Esta

Tabla 79. Lista de agrupamientos del pronombre «nos» M-F / M-1D

En la tabla 80 mostramos los agrupamientos que la herramienta informática *Clusters* ha encontrado del pronombre sujeto de primera persona del singular «yo».

Lo primero que llama la atención cuando observamos los agrupamientos de las dos muestras es que todos ellos cumplen con la función referencial del lenguaje, es decir, la función de informar. En este caso no hay intención de posicionarse, de mostrar una actitud determinada ante lo que se está explicando, ni de influir en el otro interlocutor.

Ahora bien, si tenemos en cuenta que el sistema verbal del español se caracteriza por poseer en sus formas verbales las marcas de persona y número, la presencia del pronombre sujeto en el discurso no es algo imprescindible como sucede en otras lenguas, como en el inglés, por ejemplo. Así, cuando el emisor de un mensaje utiliza de forma explícita el pronombre sujeto lo hace con una intencionalidad, ayudar a que el mensaje sea lo más claro posible y evitar ambigüedades. Por ejemplo, en los casos de «mi amiga y yo», posición 1-«F», «Canelo y yo», posición 4-«F», o «compañeros y yo», posición 3-«1D», la finalidad de la presencia del pronombre sujeto es la de identificar a alguien de entre un grupo para despejar dudas acerca de quién se trata. Otro ejemplo similar lo encontramos en el

agrupamiento «yo + verbo», uno de los más comunes en las primeras veinticinco posiciones de las dos muestras. También en estos casos la finalidad del uso del pronombre sujeto es la de distinguir la figura del sujeto-emisor respecto de otros posibles sujetos.

En los agrupamientos de las posiciones 14 («yo también») y 21 («Yo, no») de la muestra «F», lo que se desea transmitir es la idea de continuidad del tema del discurso mostrando acuerdo o desacuerdo sobre lo ya dicho.

«F»			«ID»		
yo			yo		
1	4	mi amiga y yo	1	2	años. Yo cursaba
2	2	a correr y yo	2	2	clase llegué, yo no
3	2	amiga y yo nos	3	2	compañeros y yo
4	2	Canelo y yo estuvimos	4	2	de clase llegué, yo
5	2	Canelo y yo nos	5	2	llegué, yo no conocía
6	2	María y yo	6	2	yo no conocía
7	2	que yo estaba esperando	7	2	yo no entendía
8	2	Yo no podía celebrar	8	1	a 4 alumnos, yo era
9	2	Yo no sabía que	9	1	a 9, pero yo no
10	2	Yo no sabía	10	1	a calor. Yo senti
11	1	a algún sitio. Yo	11	1	a casa yo siempre
12	1	a correr, y yo	12	1	a como yo me
13	1	a explicar que yo	13	1	a como yo. La
14	1	a hospital. Yo también	14	1	a decir eran :Yo
15	1	a la ambulancia. Yo	15	1	a la escuela yo
16	1	a llegar. Yo conseguí	16	1	a llegó y yo
17	1	a llover y yo	17	1	a lo que yo
18	1	a nosotros. Entonces yo	18	1	a matricular. Pero yo
19	1	a pasear juntas yo	19	1	a muy divertido. Yo
20	1	a percibido que yo	20	1	a pero cuando yo
21	1	a pobre! Yo, no	21	1	a planearlo todo. Yo
22	1	a recogerla. Yo estaba	22	1	a yo tomé el
23	1	a un chico. Yo	23	1	a yo un poco
24	1	a unos chicos. Yo	24	1	a, yo les dije
25	1	a viajar con yo	25	1	Ahora yo vengo a

Tabla 80. Lista de agrupamientos del pronombre sujeto «yo» M-F / M-ID

En la tabla 81 presentamos los agrupamientos del sustantivo «día». Lo primero que observamos es que en ambas muestras hay un predominio de los agrupamientos que remiten a la función referencial del lenguaje, en la que simplemente se informa de algo («día de carnaval», posición 10-«F»; «día era mi cumpleaños», posición 14-«F»; «día de clase», posición 5-«ID»; «cada día teníamos», posición 18-«ID»). A pesar de ese dominio de la función referencial, también encontramos agrupamientos con una función expresiva del lenguaje, como en las posiciones 1, 17, 18, 21, 22 y 23 de la muestra «F» en los que se hace uso de adjetivos claramente valorativos (maravilloso, tranquilo, triste, etc.). En cambio, en la muestra «ID» solo aparecen dos agrupamientos con valor expresivo («un día fantástico», posición 8-«ID»; «día muy interesante», posición 9-«ID»). En esta muestra hay que

destacar los agrupamientos para la expresión de localización temporal («*fue el día 8 de*», posición 7-«1D»; «*fue día 06*», posición 22, 23-«1D»; «*aquel día*», posición 11, 12-«1D»).

«F»			«1D»		
<i>día</i>			<i>día</i>		
1	2	Fue un <b>día maravilloso</b>	1	5	primer día de clase
2	1	a cual día era	2	4	clase fue un día
3	1	a ser un día	3	3	clase fue el día
4	1	algun día me sentiraba	4	2	de clase fue día
5	1	ayudarnos. Fue un día	5	2	<b>día de clase fue</b>
6	1	cada día. Fue un	6	2	el primer día de
7	1	calendario a cual día	7	2	fue el día 8 de
8	1	cual día era mi	8	2	fue <b>un día fantástico</b>
9	1	Cumpleaños es día y	9	2	<b>un día muy interesante</b>
10	1	<b>día de carnaval</b> . Pero	10	1	ana, este día he
11	1	<b>día de cumpleaños</b> había	11	1	aquel día, como teníamos
12	1	<b>día de deporte</b> del	12	1	<b>aquel día</b> , les conocí
13	1	día dentro de este	13	1	<b>aquel día</b> XXX fue
14	1	día era mi cumpleaños	14	1	arme. La primer día
15	1	día es cumpleaños de	15	1	Bueno este día me
16	1	día estaban juntos. Fue	16	1	cada de un día
17	1	<b>día fue muy importante</b>	17	1	Cada día aprendo algo
18	1	<b>día maravilloso, soleado</b> y	18	1	Cada día teníamos
19	1	día me sentiraba (¿) así	19	1	casa. El primer día
20	1	día nació dos hijos	20	1	China. En aquel día
21	1	<b>día normal</b> , pero pienso	21	1	clase fue un día
22	1	<b>día tranquilo</b> , todo el	22	1	clase fue <b>día 06 de</b>
23	1	<b>día triste</b> y que	23	1	clase fue <b>día 06 del</b>
24	1	día y divertida y	24	1	clase fue Un día
25	1	día yo tenía algunas	25	1	con primero día. Cuando

Tabla 81. Lista de agrupamientos del sustantivo «día» M-F / M-1D

En la tabla 82 presentamos los agrupamientos del sustantivo «casa». Como vemos, los agrupamientos que destacan en ambas muestras son: «a + casa», «en + (su, mi) casa». Dos preposiciones que se usan para expresar lugar, pero con matices diferentes. Mientras que la preposición «a» señala destino, dirección, por tanto, suele utilizarse con verbos de movimiento; la preposición «en» se utiliza para localizar en el espacio. Así, vemos que en la muestra «F» predomina la primera combinación, es decir, que la idea de movimiento con dirección, destino a la «casa» prevalece sobre la de localización en el espacio.

También llama la atención la presencia de dos locuciones preposicionales («*al lado de mi casa*», posición 6-«F»; «*cerca de mi casa*», posición 1-«1D»). Las dos expresan proximidad entre los elementos que unen.

Como observamos, todos estos agrupamientos se inscriben dentro de la función referencial del lenguaje, pues su finalidad es la de informar de forma objetiva. No obstante, encontramos un agrupamiento en la muestra «1D» («*como en casa*», posición 7) en el que

la función expresiva del lenguaje queda clara; el sujeto-emisor manifiesta de esta manera su actitud ante algo, expresando la idea de comodidad, bienestar.

«F»			«1D»		
<i>casa</i>			<i>casa</i>		
1	2	a casa muy triste	1	2	cerca de mi casa
2	2	comer en su casa	2	1	a casa las ejercicios
3	2	de volver a casa	3	1	a casa y estaba
4	2	estaba en casa y	4	1	a casa yo siempre
5	2	fiesta en mi casa	5	1	a casa, estaba muy
6	2	lado de mi casa	6	1	a casa. Pero antes
7	1	a casa Canelo estaba	7	1	a como en casa
8	1	a casa con montón	8	1	a en casa, durante
9	1	a casa de ella	9	1	a escuchado en casa
10	1	a casa de Mar	10	1	a la casa que
11	1	a casa de un	11	1	a me casa porque
12	1	a casa después de	12	1	adjetivos en mi casa
13	1	a casa en el	13	1	casa las ejercicios audio
14	1	a casa en la	14	1	casa muy cerca de
15	1	a casa encontré a	15	1	casa más lejos
16	1	a casa enfurruñada	16	1	casa porque tenia que
17	1	a casa meva! Y	17	1	casa que, durante las
18	1	a casa mi marido	18	1	casa y estaba muy
19	1	a casa mucho m	19	1	casa yo siempre estaba
20	1	a casa muy tarde	20	1	casa, a veinte minutos
21	1	a casa o un	21	1	casa, cuando mi madre
22	1	a casa para ver	22	1	casa, durante la cena
23	1	a casa pero habia	23	1	casa, estaba muy cansada
24	1	a casa porque estaba	24	1	casa. Llegué a las
25	1	a casa rápidamente	25	1	casa. Al salir, todos

Tabla 82. Lista de agrupamientos del sustantivo «casa» M-F / M-1D

Comprobamos en la tabla 83 que la función referencial del lenguaje es la que domina en los agrupamientos del sustantivo «cosas». Vemos, así, combinaciones para expresar cantidad, como «*muchas cosas*», posición 3-«F», «*tantas cosas*», posición 12-«1D» o la forma más coloquial «*montón de cosas*», posición 15-«F». Pero también encontramos la función expresiva del lenguaje en los agrupamientos «*muchas cosas preciosas* », posición 2-«F»; «*cosas horribles*», posición 24-«F»; «*cosas interesantes*», posición 25-«F»; «*cosas importante sobre mi*», posición 17-«1D»). Como observamos en los ejemplos mencionados, hay tres agrupamientos en la muestra «F» con adjetivos valorativos, mientras que en la muestra «1D» solo hay uno.

«F»			«1D»		
<i>cosas</i>			<i>cosas</i>		
1	1	a guardar sus cosas	1	2	de las cosas que
2	1	a <b>muchas cosas preciosas</b>	2	2	<b>muchas cosas</b> que no
3	1	a Olga <b>muchas cosas</b>	3	2	una de las cosas
4	1	a preguntarme cosas como	4	1	a de cosas, algo
5	1	a que explicarme cosas	5	1	ado unas cosas/reglas
6	1	aeropuerto. Las cosas han	6	1	Ahora las cosas han
7	1	algunas cosas para comer	7	1	aprendiendo cosas nuevas, y
8	1	amiga tuvo <b>muchas cosas</b>	8	1	aprendimos <b>muchas cosas</b> nuevas
9	1	apreabé unas cosas porque	9	1	as la cosas fue
10	1	carretera. Las cosas no	10	1	asimilando las <b>muchas cosas</b>
11	1	comida, apreabé unas cosas	11	1	avanzado, pero repasando cosas
12	1	como las cosas han	12	1	conocer <b>tantas cosas</b> . Ah
13	1	comprado <b>muchas cosas</b> para	13	1	cosas de su mismo
14	1	conmigo diciendo cosas	14	1	cosas en común contigo
15	1	con <b>montón de cosas</b>	15	1	cosas fue arreglando. Yo
16	1	con otros cosas y	16	1	cosas han cambiado, me
17	1	<b>cosas chinas</b> . Al cabo	17	1	<b>cosas importante sobre mi</b>
18	1	cosas como ¿que tal	18	1	cosas nuevas, y gracias
19	1	<b>cosas de la tienda</b>	19	1	cosas que me esperaban
20	1	<b>cosas de oriente</b> . Dos	20	1	cosas que menos me
21	1	cosas han cambiado después	21	1	cosas que necesitamos para
22	1	cosas han cambiado mucho	22	1	cosas que no entendaba
23	1	cosas han cambiando sobre	23	1	cosas que no podía
24	1	<b>cosas horribles</b> . Al final	24	1	cosas que no sabia
25	1	<b>cosas interesantes</b> en	25	1	cosas que realmente me

Tabla 83. Lista de agrupamientos del sustantivo «cosas» M-F / M-1D

En cuanto al sustantivo «amigos», en la tabla 84, vemos que dos de los agrupamientos que llaman la atención son: (i) «amigos + de + sustantivo», para señalar procedencia («*amigos de infancia*», posición 2-«F»; «*amigos de Brasil*», posición 16-«F»; «*amigos de diferentes países*», posición 7-«1D»); (ii) «amigos + verbo», para indicar estado o acción («*amigos estaban muy entretenidos*», posición 22-«F»; «*amigos gritaban*», posición 25-«F»; «*amigos tomamos un café*», posición 17-«1D»). Podemos comprobar que estos agrupamientos, como los demás que aparecen en la tabla, cumplen con la función referencial del lenguaje. En este caso no encontramos ejemplos de función expresiva ni conativa, como en los agrupamientos anteriores.

«F»			«1D»		
<i>amigos</i>			<i>amigos</i>		
1	2	bailar con mis amigos	1	2	amigos en la clase
2	2	mis <b>amigos de infancia</b>	2	1	a algunos amigos de
3	2	que todos mis amigos	3	1	a amigos y no
4	2	uno de mis amigos	4	1	a mis amigos. Me
5	1	a mis <b>amigos de</b>	5	1	a muchos nuevos amigos
6	1	a ocurrido. Mis amigos	6	1	acabé haciendo nuevos amigos
7	1	a que íbamos amigos	7	1	algunos <b>amigos de diferentes</b>
8	1	a sus amigos u	8	1	<b>amigos de diferentes países</b>
9	1	a todos mis amigos	9	1	<b>amigos de diferentes países</b>
10	1	a. Todos mis amigos	10	1	amigos hablábamos, re
11	1	abrí, entraron muchos amigos	11	1	amigos hasta el momento
12	1	algunos amigos para tomar	12	1	amigos me invitara a
13	1	amigos a cenar y	13	1	amigos porque me resultaba
14	1	amigos a tomar una	14	1	amigos porque todos hablamos
15	1	amigos allí. En su	15	1	amigos que de otros
16	1	<b>amigos de Brasil</b> empezaron	16	1	amigos razón y
17	1	<b>amigos de infancia</b> , y	17	1	amigos tomamos un café
18	1	<b>amigos de infancia</b> . Después	18	1	amigos viejos. Me sentí
19	1	<b>amigos de Roma</b> para	19	1	amigos y me encontré
20	1	amigos en mi país	20	1	amigos y no conocía
21	1	<b>amigos estaban</b> allí, pero	21	1	amigos, como Debra es
22	1	<b>amigos estaban</b> muy entretenidos	22	1	amigos. En esa clase
23	1	<b>amigos estuvieron</b> esperando para	23	1	amigos. Escribamos las informaciones
24	1	<b>amigos fueron</b> preocupado con	24	1	amigos. Hablamos sobre cultura
25	1	<b>amigos gritaban</b> Feliz cumpleaños	25	1	amigos. Me recuerdo bien

Tabla 84. Lista de agrupamientos del sustantivo «amigos» M-F / M-1D

«F»			«1D»		
<i>tiempo</i>			<i>tiempo</i>		
1	3	<b>mucho tiempo</b> que no	1	1	a pasé el tiempo
2	2	que no tuve tiempo	2	1	a tiempo y deber
3	2	visto hace <b>mucho tiempo</b>	3	1	a, el tiempo no
4	1	<b>a buen tiempo</b> . Mi	4	1	a. El tiempo pasó
5	1	a llover. El tiempo	5	1	ahora tengo <b>mas tiempo</b>
6	1	a <b>mucho tiempo</b> . Desde	6	1	<b>al mismo tiempo</b> estaba
7	1	<b>a muy mal tiempo</b>	7	1	<b>al mismo tiempo</b> nos
8	1	<b>a tiempo</b> mi madre	8	1	<b>bastante tiempo</b> , así que
9	1	<b>a tiempo</b> para embarcar	9	1	<b>buen tiempo!</b> Me encanta
10	1	<b>a tiempo</b> para llegar	10	1	<b>buen tiempo</b> . Yo fui
11	1	<b>a tiempo</b> . Llamé a	11	1	cabo de un tiempo
12	1	a ya un tiempo	12	1	clase <b>con tiempo</b> . El
13	1	absolutamente del tiempo. A	13	1	comida, etc. el tiempo
14	1	<b>al mismo tiempo</b> que	14	1	con el tiempo, empecé
15	1	<b>al mismo tiempo</b> , Ápero	15	1	<b>con tiempo</b> algunos alumnos
16	1	allí <b>mucho tiempo</b> . Estábamos	16	1	<b>con tiempo</b> para poder
17	1	allí. En su tiempo	17	1	<b>con tiempo</b> . El edificio
18	1	allí por <b>mucho tiempo</b>	18	1	conozco, fui con tiempo
19	1	ar <b>al mismo tiempo</b>	19	1	Contar <b>mucho tiempo</b> . En
20	1	<b>buen tiempo</b> , había mucha	20	1	creía. El tiempo
21	1	<b>buen tiempo</b> . Mi amiga	21	1	cuando el tiempo llegué
22	1	calle. Normalmente ese tiempo	22	1	de 2008. Había <b>buen tiempo</b>
23	1	casa estaba <b>mal tiempo</b>	23	1	de un tiempo y
24	1	casi todos tiempo acompañaba	24	1	dedica su tiempo libre
25	1	causa del <b>mal tiempo</b>	25	1	día, el tiempo

Tabla 85. Lista de agrupamientos del sustantivo «tiempo» M-F / M-1D

En relación al sustantivo «tiempo», Muñoz Lobo (2012) también selecciona este sustantivo para analizarlo en su trabajo. El criterio que sigue el autor es la lista de frecuencias del corpus CREA. Muñoz Lobo (*op. cit.*) indica que «tiempo» ocupa la posición setenta y, aunque no es el primero en la lista, lo eligió por presentar mayor variedad de colocaciones. En nuestro trabajo también podemos comprobar, en la tabla 85, la diversidad de agrupamientos que presenta este sustantivo en las dos muestras analizadas. Destacamos aquí algunos ejemplos de agrupamientos: (i) para expresar cuantificación: «*mucho tiempo*», posición 1-«F»; «*bastante tiempo*», posición 8-«1D»; (ii) para expresar simultaneidad: «*al mismo tiempo*», posición 14-«F», 6-«1D»; (iii) con adjetivos: «*buen tiempo*», posición 4-«F», posición 9-«1D»; «*mal tiempo*», posición 7-«F»; (iv) con las preposiciones «a» y «con», para formar las locuciones adverbiales: «*a tiempo para*», posición 9-«F» (indica momento justo), «*con tiempo para*», posición 16-«1D» (indica antelación). Como vemos, en todos los agrupamientos percibimos la función referencial del lenguaje.

«F»			«1D»		
<i>gente</i>			<i>gente</i>		
1	4	había <b>mucha gente</b>	1	3	<b>mucha gente</b> de diferentes
2	2	a <b>muchísima gente</b>	2	2	a otra gente de
3	2	Toda la gente estaba	3	2	con gente que no
4	1	a ayudar. La gente	4	2	gente de diferentes países
5	1	a gente tan temprano	5	2	gente de todo el
6	1	a la gente. La	6	2	gente en la clase
7	1	a <b>mucha gente</b> capaz	7	2	Había gente de
8	1	a <b>mucha gente</b> en	8	2	había <b>mucha gente</b>
9	1	a <b>mucha gente</b> porque	9	2	<b>mucha gente</b> en la
10	1	a <b>mucha gente</b> , ped	10	1	a la gente
11	1	a otra gente. Por	11	1	a conocido gente muy
12	1	acercé a la gente	12	1	a gente de muchos
13	1	algo raro. La gente	13	1	a gente de todo
14	1	alta. Había <b>bastante gente</b>	14	1	a <b>gente genail</b> y
15	1	an <b>mucha gente</b> por	15	1	a la gente nueva
16	1	aquí hay <b>mucha gente</b>	16	1	a la gente y
17	1	aquí son la gente	17	1	a la gente. Todos
18	1	Así la gente que	18	1	a llegado <b>mucha gente</b>
19	1	avions y la gente	19	1	a <b>mucha gente</b> de
20	1	ayudar. La gente que	20	1	a <b>mucha gente</b> estaba
21	1	año <b>mucha gente</b> de	21	1	a <b>mucha gente</b> y
22	1	Barcelona, con <b>mucha gente</b>	22	1	a <b>poca gente</b> . En
23	1	<b>bastante gente</b> en nuestra	23	1	a <b>poco la gente</b>
24	1	Cada año <b>mucha gente</b>	24	1	a y ver gente
25	1	calles y toda gente	25	1	academia y con gente

Tabla 86. Lista de agrupamientos del sustantivo «gente» M-F / M-1D

En la tabla superior, vemos los agrupamientos del sustantivo «gente». Podemos comprobar que una parte importante de ellos, un 36%, se combina con el adjetivo «mucho», con el cual se está indicando el exceso, lo que sobrepasa a lo normal, incluso aparece el superlativo («*muchísima gente*», posición 2-«F»). Pero también encontramos los adjetivos «bastante» y «poco» acompañando al sustantivo «gente» (posiciones 23-«F», 22-«1D»). Estos agrupamientos para expresar la idea de cantidad, por exceso o por falta, podemos decir que están más cerca de la función expresiva del lenguaje que de la referencial, ya que el sujeto-emisor toma posición a la hora de valorar la cantidad de algo, en este caso, de «gente». Un ejemplo de función expresiva con léxico valorativo lo encontramos en la posición 14-«1D» («*gente genail*» - entendemos que se trata de un fallo y que la forma correcta es «genial»).

## **Fase 2: Concordancias.**

En esta parte del análisis, vamos a ver las palabras en contexto para ver con más detalle qué tipo de vocabulario rodea a la palabra seleccionada y con qué función del lenguaje se adhieren. Como muchos de los fragmentos que ofrece el *output* de la herramienta *Concordance* coinciden con el enunciado de las tareas propuestas:

1        domingo por la mañana estaba paseando con Canelo, mi perro, cuando de repente ...

hemos decidido realizar una selección de diez fragmentos del total de cada una de las muestras.

A continuación presentamos la tabla con algunos ejemplos del contexto en el cual podemos encontrar el posesivo singular de primera persona «mi» en las muestra «F» y «1D».

«F»	
mi	
5	to era una probabilidad mínima. Así que escribí a <b>mi esposa</b> que cuidara de Canelo, que le lavara el
22	miedo. Pero de repente me giré y allí estaba él, <b>mi novio</b> . Me enfadé ya que me hizo pasar un mal
45	e casa, situado dos calles más abajo. Al llegar a <b>mi habitación</b> me quitó la camisa y los pantalones
88	un sueño, una pesadilla horrible. Seguí con <b>mi</b> rutina habitual, intentando todo aquello que
133	or si alguien había visto lo sucedido. Pero para <b>mi sorpresa</b> , nadie reaccionó a la llamada y seguí
147	y estos acudieron rápidamente. Una vez tomada <b>mi</b> declaración respeto al transcurrir de los hechos,
197	algun ha abierto la luz y en el salon estaban <b>mi toda la familia</b> . Toda gente me dice: “cumpleaños
216	El gato, muy nervioso, parecía que quiere comer a <b>mi perro</b> y mi pobrecito Canelo tenia mucho miedo de
241	pués tendría que irme a un fiesta de cumpleaños de <b>mi sobrino</b> . Había ido a peluquero y gastaba mucho,
268	Conocí uno de los niños que vivió con su madre en <b>mi finca</b> , la madre era una profesora de un colegio

«1D»	
mi	
13	Su pronunciación alemana parecía perfecta ante <b>mi completa ignorancia</b> de este idioma, no entendía
72	<b>mi pasión</b> por el idioma ha ido en aumento y <b>mi</b> interés ha traspasado la lengua. Ahora me interesa
77	a coger confianza y soltura en el habla. Por <b>mi</b> parte fue una experiencia <b>increíble</b> , porque nunca
78	es había estudiado esta lengua. Desde ese momento <b>mi pasión</b> por el idioma ha ido en aumento y mi inter
79	<b>mi pasión</b> por el idioma ha ido en aumento y <b>mi interés</b> ha traspasado la lengua. Ahora me interés
94	o ya he dicho, fue un d'a horrible. Eso se debe a <b>mi carácter</b> : soy una persona muy <b>tímida</b> y <b>reservada</b> ,
146	ocurrió ya que era muy peque-a. Recuerdo que en <b>mi</b> primera clase de inglés estaba muy <b>ilusionada</b>
162	y era una locura empezar alemán Puede ser. Para <b>mi sorpresa</b> , la clase empezó totalmente en alemán.
231	entro ellos. Me sentí bien, pero fui preocupada a <b>mi</b> español porque mucho gente hablan mejor que yo.
252	os. Ellos son simpáticos y trabajadores. Recuerdo <b>mi</b> primer día de clase de español fue <b>interesante</b>

Tabla 87. Output de la herramienta *Concordance* para el adjetivo «mi» M-F / M-1D

Al observar el contexto en el que encontramos el posesivo «mi», se corrobora lo ya observado en la fase anterior de agrupamientos; es decir, que la muestra «1D» presenta un léxico explícito más valorativo que denota una implicación personal y emocional mayor que en la muestra «F». En ésta observamos que el vocabulario que rodea al adjetivo apenas expresa sentimientos, emoción o valoración. Los agrupamientos con el posesivo son para expresar relación de parentesco («*mi esposa*», posición 5-«F»; «*mi novio*», posición 22-«F»), o expresar posesión («*mi habitación*», posición 45-«F», «*mi finca*», posición 268-«F»). No obstante, en la posición 216 de la muestra «F» se refiere al perro como «*mi pobrecito Canelo tenía mucho miedo*»; es una expresión de sentimiento pero hacia un tercero, no hacia el sujeto-emisor del mensaje; como sí ocurre en la muestra «1D» («*mi completa ignorancia*», posición 13-«1D»; «*mi pasión*», posición 72«1D»). Así, vemos que en la muestra «1D» predomina la función expresiva del lenguaje frente al dominio de la función referencial en la muestra «F».

«F»	
mis	
1	en un acto desesperado, traté de levantarlo en mis brazos soportando su peso el cual, por el miedo
5	quedaban mirando anonadados por aquella situación. Mis mejillas podían sentir las lágrimas que caían
9	eran reducidas. Además, era viernes, por lo tanto mis amigas estaban en la universidad y no podían
17	ser muy amables conmigo y me llevaron con mis amigos. Finalmente, pude ir a cenar con ellos,
19	Primero de todo, no supe como reaccionar. Mis piernas no me respondían hasta que Canelo empezó
22	donde i como es mi casa. He presentado mi novio i mis nuevos amigos. Luego fuimos a cenar en casa de mi
37	mi cumpleaños. Y desde hace este momento, mis amigos me ofrecen un camion (juego) para cada
43	mayoría de amigos no vinieron. Solo vinieron tres mis mejor amigos. Pero todo estaban mojado entonces,
55	no tener mas alguien que te guste mucho. Creo que mis cumpleaños no más sean los mismos de
71	demasiado tarde: con todo lo que había hablado de mis problemas al trabajo, de mis problemas con la

«1D»	
mis	
1	disfruté mucho y es una anécdota más que contar a mis amigos.
4	alemán. Recuerdo aquel primer día perfectamente, mis sensaciones antes, durante y después. Llegué a
5	darse cuenta, cambié al catalán. Fueron llegando mis compañeros y nos fuimos presentando. Era una
7	amiga para ir a la primera clase. Me acuerdo de mis nervios constantes, pero a la vez, de estar contenta
11	in ningún amigo o conocido y, muy a mi pesar, sin mis padres. La idea de ir a Irlanda a aprender inglés
17	profesora con gestos y no entendía lo que me decían mis compañeros. Días antes, aprendí cómo saludar,
26	alemana y en qué países es hablada. Respecto a mis nuevos compañeros, no conocía a ninguno y me
31	en la puerta vi que había escrito francés. Mis nervios aumentaban a medida que se acercaba la
38	donde me había puesto. Pero, al girarme y mirar a mis compañeros vi la misma cara de no comprender
49	di mucha información sobre el programa del curso. Mis compañeros se presenten, yo comencé hablar con

Tabla 88. Output de la herramienta *Concordance* para el adjetivo «mis» M-F / M-1D

En cuanto al posesivo plural de primera persona «mis», comprobamos en el *output* de la herramienta *Concordance* (tabla 88) que principalmente se establece una relación de parentesco en ambas muestras (F»: posiciones 9, 17, 22, 37, 43; «1D»: posiciones 1, 5, 11, 17, 26, 38, 49), igual que en el análisis de los agrupamientos, y, por tanto, estas concordancias cumplen con la función referencial del lenguaje, informar de una realidad. No obstante, observamos también la presencia de ejemplos de concordancias, por un lado, con una clara función poética del lenguaje («*mis mejillas podían sentir las lágrimas que caían*», posición 5-«F») con la finalidad de crear belleza recurriendo a recursos retóricos, en este ejemplo a la prosopopeya. Por otro lado, una función expresiva («*Recuerdo aquel primer día perfectamente, mis sensaciones antes, durante y después*», posición 4-«1D»; «*me acuerdo de mis nervios constantes, pero a la vez, de estar contenta*», posición 7-«1D»). En estas concordancias, el sujeto-emisor está expresando sus sentimientos, sus emociones, en definitiva, su actitud ante lo que explica.

«F»	
me	
1	cuando de repente le empezó a sangrar la nariz. Me asusté y fui al veterinario para saber qué le
2	para saber qué le ocurría. El veterinario me dijo que Canelo tenía una enfermedad llamada
36	contra la pared. No podía estar más asustada, me moría de miedo! Pero de repente me giré y allí es
57	allí. Le pregunté si no le había pasado nada. El me aseguró que todo estaba bien y que no ha escuchad
63	más tarde, un agente de policía se acercó y me acusó de haber robado ese cuadro. Yo, desesperada
65	que caían de mis ojos. Me llevaron a comisaría y me interrogaron. Yo seguía en shock. No tenía
118	y me puse a llorar entre sus brazos. Al instante me preguntó qué sucedía, as' que se lo expliqué
295	todavía estaba con su novia y si ya estaba casado. Me responde que se casaba hace dos años con su novia
301	muy contento y le dije que muchas gracias y ella me dijo que me quiere mucho. Y luego nosotros fuimos
375	Nuestra madre va a moriré mi hermana mayor me dijo con una voz triste. No podía hacer nada que

«1D»	
me	
1	a la ESO. Como el instituto era nuevo para mí, no me conocía la localización de las aulas y me
22	compañeros, especialmente de María con quien ahora me llevo genial. En definitiva, mi introducción a
52	nos llevamos muy bien. Ahora, puedo decir que me siento bien integrada en mi nueva aula. Respeto a
73	A medida que iban entrando más alumnos me tranquilizaba, ya que todos estábamos en la misma
147	capaces de entender lo que había dicho y más. Me encantó hacer francés, la profesora nos motivaba
163	yo empezaba 2º de la ESO. La elegí porque siempre me han gustado los idiomas y el francés me parecía
282	alumnos parecían que conocer algunos en clase, y me puse más nerviosa. Primer hemos hecho introducir
304	pañeros nuevos. Y me encontré unos amigos viejos. Me sentí muy bien. El primer d'a de clase hicimos un
353	muy simpática. Nos hizo muchas bromas. Así que me sentía un poco mejor. Conocí una chica.
363	característica, cultura y costumbre de cada país. Me encanta de conocer tantas cosas.

Tabla 89. Output de la herramienta Concordance para el pronombre «me» M-F / M-1D

Si en la fase de agrupamientos, veíamos que el pronombre «me» en la muestra «F» su uso era mayoritariamente para introducir el discurso indirecto («me dijo...»), aquí (tabla 89) se confirma esa tendencia («me interrogaron», posición 65-«F»; «me preguntó», posición 118-«F»; «me dijo que...», posición 2, 301, 375-«F»). Pero, además, al ampliar el contexto donde se inscriben estos agrupamientos, vemos que estos verbos de comunicación proporcionan más información. Por ejemplo, al leer los fragmentos de las posiciones 301-«F» y 375-«F», comprobamos que estamos claramente ante la función expresiva del lenguaje. En el primer fragmento, el sujeto-emisor está haciendo una declaración de sus sentimientos, de su estado de ánimo, mientras que en el segundo, esa expresión de sentimientos se refiere a una tercera persona. Pero en ambos casos se está expresando un posicionamiento, una actitud ante la situación que se ha generado.

En la elección de los verbos en los ejemplos de las posiciones 65-«F» («interrogar») y 118-«F» («preguntar»), vemos implícita la función conativa del lenguaje. Presumimos que la semántica de dichos verbos en el discurso directo apela a que el interlocutor actúe, haga algo ante una demanda, en este caso, a responder.

Asimismo, también encontramos ejemplos en los que, igual que en la muestra «1D», se expresa de forma explícita emoción o sentimientos, función expresiva del lenguaje, a través de un léxico valorativo: «me asusté», posición 1-«F»; «me moría de miedo», posición 36-«F».

En la muestra «1D» se verifica, asimismo, que predomina la expresión de sentimientos y emociones, es decir, la función expresiva del lenguaje con una presencia llamativa de léxico valorativo: «*me llevo genial*», posición 22-«1D»; «*me tranquilizaba*», posición 73-«1D»; «*me sentia un poco mejor*», posición 353-«1D»). Como vemos en los ejemplos, en todos los casos esta expresión de sentimientos la «padece» el sujeto-emisor, es decir, la primera persona; mientras que en la muestra «F», como acabamos de ver, también hay referencia a la tercera persona.

«F»	
nos	
1	los días paseábamos por el mismo lugar y siempre nos encontrábamos con la misma gente, día tras día,
8	con el que siempre jugaba mi Canelo, pero no nos vieron ya que tomaron otra calle para ir al parque
18	paseo por la montaña, parecía abandonada, pero nos serviría para escondernos de la lluvia. Canelo y
24	a en dirección al pueblo. Finalmente, Canelo y yo nos separamos de ellos y empezamos a correr, era la
31	He preguntado a Olga muchas cosas y al final nos paseamos hablando mucho tiempo. Después de escuela
46	Muy bueno! Y entonces para postres, el camarero nos trajo una tarta con una vela para mi. Fue una
61	Estaba una chica guapa como de pendiente cuadro nos miró, nos dijo: pasa, pasa, voy a coger toalla
68	En aquel momento devolvio aquella chica que nos dio dos animales y los llevó sin decirnos nada.
84	el perrito blanco era suyo. Estaba muy contenido, nos invitó a tomar un copa. Que día tengamos! Estaba
102	país diferente. Y ahora, en ese momento, por fin nos vimos. Estábamos muy contentas. Ellas las dos

«1D»	
nos	
1	que muy básica: la profesora se llamaba Cristina- nos dio unas hojas con el alfabeto en francés y los
10	que no sería fácil ni mucho menos. Curiosamente, nos centramos en este mismo aspecto, la fonética,
18	ya que además de romper el hielo en el aula, nos divertimos. Esta clase fue hace tres años, en 4º
22	que además somos toda una clase de cinco mujeres y nos llevamos muy bien. Ahora, puedo decir que me
32	en un recinto escolar distinto al de primaria, no nos conocíamos mucho y nos sentíamos algo nerviosos
35	aunque muy sistemático. Más adelante, la profesora nos animó a realizar preguntas básicas para poder con
54	dicho y más. Me encantó hacer francés, la profesora nos motivaba i animaba siempre.
60	la asignatura en catalán. Todos en la clase nos sentimos aliviados y nos miramos unos a otros son
65	hora y cuando el tiempo llegué a entrar en clase nos sentimos mucho mejor. La clase empezó a las 12 y
90	de la clase. Pero la profesora fue muy simpática. Nos hizo muchas bromas. Así que me sentia un poco

Tabla 90. Output de la herramienta *Concordance* para el pronombre «nos» M-F / M-1D

En la tabla 90, las concordancias del pronombre complemento «nos» ayudan a confirmar lo observado en la fase anterior de análisis, en los agrupamientos. Es decir, en la muestra «F» predomina la función referencial del lenguaje, informar de lo que sucede («*y siempre nos encontrábamos con la misma gente*», posición 1-«F»; «*Finalmente, Canelo y yo nos separamos de ellos y empezamos a correr*», posición 24-«F»; «*el camarero nos trajo una tarta con una vela para mi*», posición 46-«F»). No obstante, también encontramos ejemplos en los que al ampliar el contexto vemos que el sujeto-emisor expresa su actitud ante lo que está explicando, ajustándose así a la función expresiva del lenguaje («*estaba muy contenido, nos invitó a tomar una copa*», posición 84-«F»; «*por fin nos vimos. Estábamos muy contentas*», posición 102-«F»).

Sin embargo, en la muestra «1D», en las concordancias, al igual que en los agrupamientos, observamos un predominio de la función expresiva del lenguaje con la presencia de verbos para expresar sentimientos o emociones («1D»: «*ya que además de romper el hielo en el aula, nos divertimos*», posición 18-«1D»; «*Todos en la clase nos sentimos aliviados y nos miramos unos a otros*», posición 60-«1D»). Pero también observamos una concordancia con función conativa del lenguaje («*Más adelante la profesora nos animó a hacer preguntas básicas para*», posición 35-«1D»). Vemos en este ejemplo una clara intención de que alguien actúe, responda a la acción del verbo.

«F»	
yo	
1	conflicto con las tropas del este de marte. Yo había participado como militante el año pasado
15	ces el ladrón huyó con las pertenencias, mientras yo estaba en el suelo un poco confusa. Al levantarme
23	de repente mi perro vio a un gato y le persiguió. Yo le llamaba muchas veces pero no me hizo caso y si
40	en dirección al pueblo. Finalmente, Canelo y yo nos separamos de ellos y empezamos a correr, era
56	pudiera estar quieto. De hecho, ni mi familia ni yo notamos nada hasta que lo saqué a pasear. Seguimos
69	años escolares. Ella estaba paseando con su hijo. Yo no sabía que ya tiene niños. He preguntado a Olga
82	al piso de Susanna ella vive en la planta bajo y yo en el septimo. Con la suerte que yo tenía el
129	detrás de mi perro. Mi perro sigué corriendo y yo tambien. Después de 20 minutos no pudo mirar a mi
145	tenía nunca tanto tiempo en la calle. En esta día yo conoció su mejor amiga y mi perro tambien. Ahora
214	su mujer. Resultó que él se caso hace 1 año. Así yo conocí a su mujer, ví sus piso nuevo en Sitges

«1D»	
yo	
1	empezé a estudiar alemán, en septiembre de 2012. Yo conocía a tres o cuatro personas de la clase, ya
12	de clase fue cuando yo tenía apenas siete años. Yo ya sabía que el segundo curso de primaria se
30	ya que es el que he elegido como segunda lengua. Yo ya había hecho un curso intensivo en Julio, y por
38	llevé ha hacer mi primera clase, de matemáticas. Yo estaba bastante perdida y me costaba seguir la
50	yo fue con me amigo para tomieste un café. Luego yo fue a me casa porque tenía que planchar me las
58	oso. En septiembre 2008 llegamos en Barcelona, y yo hablé ni español, ni catalan, ninguna palabra
73	estresado. Fue mucha gente de diferentes países. Yo no conocí a nadie. La profesora fue simpática y
85	a muy suerte que pude entrar esta escuela famosa. Yo tomé el autobús B19. Cuando llegué, ya había
101	siempre me ponía cara roja y cuando regresé a casa yo siempre estaba de malhumor. Cada día teníamos
114	Entonces, expliqué la clase la historia de ellas. Yo pensaba que eran divertidas y muy guapas. Después

Tabla 91. *Output* de la herramienta *Concordance* para el pronombre «yo» M-F / M-1D

En la tabla superior presentamos el contexto en el que encontramos el pronombre sujeto «yo». Como ya hemos señalado en el apartado de agrupamientos, la presencia del pronombre sujeto en un mensaje es para evitar ambigüedades o resaltar el hecho de que se trata de ese sujeto y no de otro posible o indicar la conformidad o disconformidad con el tema que se trata. Así, el *output* de la herramienta *Concordance* viene a confirmar lo visto en el apartado de agrupamientos. Por ejemplo, en las posiciones 1 de la muestra «F» («...*con los tropas del este de marte. Yo había participado como militante el año pasado*») y en la 73 de la muestra «1D» («*Fue mucha gente de diferentes países. Yo no conocí a nadie.*»), vemos que la presencia del pronombre sujeto tiene la finalidad de determinar que se trata de ese sujeto y no de otro. Comprobamos, asimismo, que en una de las concordancias la

presencia del pronombre sujeto es imprescindible para comprender el enunciado: «*Mi perro sigué corriendo y yo también*», posición 129-«F». Vemos que en este caso, la ausencia del pronombre sujeto haría incomprensible el mensaje. Igualmente, evidenciamos que la mayor parte de las concordancias cumple con la función referencial del lenguaje al informar de lo que lo que acontece.

«F»	
día	
1	Y estaba bebiendo aguas o refrescas. Fue el día de deporte del pueblo. Pensé que yo también que
5	bonito con quién estoy paseando cada día. Fue un día maravilloso, soleado y alegre para mí. Estuve muy
6	ó con su amiga mientras vigilando su hijo. Era un día tranquilo, todo el mundo eran feliz. Y mi perrito
7	iraba con cariño y curiosita. Ayer para mí fue un día normal, pero pienso qué para ellos este día fue m
8	ue un día normal, pero pienso qué para ellos este día fue muy importante.
10	ilia y de mis padres, yo penséis que iba a ser un día triste y que yo iría quedarme no sofá por todo el
11	o, para vivir en BCN. Miraba mi calendario a cual día era mi cumpleaños, pero pensaba esta día es cumpl
13	os. ¡Olvidé que era mi cumpleaños, pero pasaba un día fantástico!
15	tenia algunas llamadas de Ucrania. Cumpleaños es día y divertida y no divertida porque es fiesta, te v
22	ués de un rato Rafael vino para ayudarnos. Fue un día maravilloso!

«1D»	
día	
1	cosas que necesitamos para aprender español. Esta día yo conoci con los compañeros de me clase.
7	Mi primer día de clase fue un día maravilloso. En septiembre 2008 llegamos en Barcelona
13	Mi primer día de clase fue día 06 del mes de febrero en un lunes y empezava
18	otras culturas y muchas idiomas. el primer día ha sido muy tranquilo pero yo estaba muy nervioso
22	Mi primer día de clase fue el día 8 de Febrero el lunes di no me equivoco. He estado
34	Mi primer día de clase fue Un día en octubre pero no se la fecha exacta. Tuve que e
37	otra escuela y aquí estoy estudiando y el primer día de clase fue genial, conoci mucha gente había mas
38	Mi primer día de clase fue el día 8 de febrero de 2009. Fue muy interesante y me
39	pero todavía no hablamos bien español. En aquel día, les conoci a mis compañeros. En general a mi me
40	Mi primer día de clase fue un día fantástico. Tenia miedo porque no conocia nadie d

Tabla 92. Output de la herramienta Concordance para el sustantivo «día» M-F / M-1D

Como vemos en la tabla 92, las concordancias del sustantivo «día» corroboran el predominio, comentado en el apartado de agrupamientos, de la función expresiva del lenguaje en la muestra «F» frente a la muestra «1D». Advertimos que en la muestra «1D» el sustantivo «día» va seguido, en la mayoría de los casos, por la fecha («*fue dia 08 de febrero de 2009*», posición 38-«1D»), manera de situarnos en el espacio temporal (función referencial del lenguaje). Aun así, al ampliar el contexto, comprobamos que a la función referencial de ese agrupamiento se añade la función expresiva al pasar a valorar cómo fue el día: («*fue dia 08 de febrero de 2009. Fue muy interesante y...*», posición 38-«1D»). Una presencia más explícita de la función expresiva en la muestra «1D» la encontramos en las posiciones 7, 37 y 40, en las que aparecen adjetivos valorativos como «maravilloso, genial, fantástico».

En la muestra «F», sin embargo, vemos que la valoración (función expresiva del lenguaje) está presente en ocho de los diez ejemplos de concordancias extraídos

aleatoriamente del *output* de la herramienta *Concordance*, como en la posición 5 («*fue un día maravilloso, soleado y alegre para mí*», posición 5-«F»).

«F»	
casa	
1	a través de la picadura de un mosquito. <b>Me fui a casa muy triste</b> y empezaron los días de agonía por mi
5	acción un tanto extraña: me había olvidado <b>de ir a casa</b> de María a buscar ese libro que tanto necesitaba
20	el vecino estaba en casa y me dejó colarme <b>en mi casa</b> a través de su jardín. Sin embargo, antes de
31	Pasó media hora y le dije que debía <b>volver a casa</b> y decidió acompañarme. No obstante, antes,
54	con Canelo. De repente, empezamos a correr <b>hacia casa</b> , deseaba que no hubiera ocurrido nada. Cuando
67	Después de pasear yo he encene donde i como es mi casa. He presento mi novio i mis nuevos amigos.
82	toda familia soya. Muchas veses fui a comer <b>en su casa</b> . El tambien venia en mi casa. Yo cuando fui en
96	i al gimnasio como todos los dias. Cuando <b>volví a casa</b> , mi novio me preguntó si me gustaria ir a un
119	laves de casa. Es que vivía sola no pude <b>entrar a casa</b> sin llaves, por eso tenía que volver al camino
150	podriamos coger el metro a la casa. <b>Lleguemos a casa</b> y ella se fue a guardar sus cosas pero la vi muy

«1D»	
casa	
1	nos enseñó algunos para que practicáramos <b>en casa</b> . Al salir, todos estábamos muy animados y
5	a que vinieran a recogerme para <b>llevarme a la casa</b> que, durante las próximas tres semanas, sería mi
7	la chica nueva, y sin embargo, <b>me sentía como en casa</b> . Cada día era una nueva experiencia; a veces
9	hecho que agradecí porque tenía ganas de <b>llegar a casa</b> . Pero antes de irme, la chica que estaba a mi lado
12	a clase. Nos ha dicho que tenía el ploblama <b>en la casa</b> . Después de salir de la clase, fui a la biblioteca
15	a mis amigos. Me recuerdo bien que cuando <b>volví a casa</b> , estaba muy cansada. Cogí a metro desde
17	habian muchas cosas que no podía estudiar <b>en casa</b> . Mi profesora fue XXXX y era muy simpatica.
19	siempre me ponía cara roja y cuando <b>regresé a casa</b> yo siempre estaba de malhumor. Cada día
20	que ayudarme si no yo no iba a cocinar ni limpiar <b>casa</b> . Estudié solo un mes en esa escuela fue muy duro
22	la fuemos a la escuela, ella casa muy <b>cerca de mi casa</b> . Cuanto cojemos en el metro estamos muy alegres.

Tabla 93. *Output* de la herramienta *Concordance* para el sustantivo «casa» M-F / M-1D

En la tabla superior mostramos algunos ejemplos de las concordancias del sustantivo «casa» que ofrece la herramienta *Concordance*. Vemos que en ambas muestras hay un predominio de la combinación «verbo de movimiento + a + casa»: «*me había olvidado de ir a casa de María*», posición 5-«F»; «*cuando volví a casa*», posición 96-«F»; «*tenía ganas de llegar a casa*», posición 9-«1D»; «*cuando regresé a casa*», posición 19-«1D». Estas concordancias ratifican lo ya observado en el apartado de agrupamientos. Recordemos que este agrupamiento señala dirección, destino. Otra concordancia que también presenta el mismo significado de dirección es «*De repente, empezamos a correr hacia casa*», posición 54-«F». Aquí, también con verbo de movimiento («correr»), la preposición utilizada es «hacia». Otra preposición presente en las concordancias, y que estaba igualmente en los agrupamientos, es «en», que se usa para indicar localización espacial («*me dejó colarme en mi casa*», posición 20-«F»; «*nos enseñó algunos para que practicáramos en casa*», posición 1-«1D»; «*no podía estudiar en casa*», posición 17-«1D»). En todas estas concordancias comprobamos que se cumple con la función referencial del lenguaje, es decir, informar sin mediar expresiones valorativas, de sentimientos o emocionales. No obstante, descubrimos dos concordancias en las que sí se expresa sentimiento: (i) «*me fui a casa muy triste y*

*empezaron los días de agonía para mí*», posición 1-«F». En este ejemplo, la función expresiva del lenguaje queda claramente manifiesta, el sujeto-emisor no se limita a informar, sino que expone su actitud acerca de lo que dice. (ii) «*y sin embargo me sentía como en casa*», posición 7-«1D». La presencia del verbo «sentirse» en esta concordancia, confirma lo que ya hemos comentado en el apartado de agrupamientos de la combinación «*como en casa*»; lo que el sujeto-emisor está expresando es que «está cómodo, a gusto».

«F»	
cosas	
1	que podía hacer, iba mentalizandome de que estas cosas sucedían en las películas y entonces llegué a la
6	pidiéndome vernos un día para que le entregara sus cosas, sobretodo porque estaba preocupada por su
7	s mi amiga y yo, pensé que se pudo suceder muchas cosas interesante. Pués le dije a mi amiga que su hija
13	Tomamos un café en una cafetería y hablamos sobre cosas interes antes en nuestras vidas. Ella vivió en
17	que me he pasado todo el día. Después de eso, las cosas se acalmaran y pude aprovechar mi cumpleaños.
21	como soy extranjera, ella esplicaba para mi muchas cosas muy interesantes. Nuestros perros todo tiempo
25	e tenemos Tambien mi amiga y yo nos gustaban las cosas de la tienda. final, compré una collar y mi
28	el hombre empezaba a pelearse con migo diciendo cosas horribles. Al final descubrimos que era una
31	quieres? - Como te había dicho, has olvidado dos cosas: ponerme azúcar en el café y dejarme las llaves
34	llegó y tomiamos café y estabamos hablando que cosas han cambiando sobre seguridad en el

«1D»	
cosas	
1	bastante entretenida y en la que aprendimos muchas cosas nuevas.
3	lección bien asumida aunque asimilando las muchas cosas que me esperaban en el curso. Fue el
6	dudas se disiparon y tuve claro que una de las cosas que realmente me apasionaban era el aprendizaje
7	distinta con otras personas, estas tienen cosas en común contigo, y eso hizo que pudiera relacion
9	o, y por eso el primer día ya sabía la mayoría de cosas, algo que me desconcertó y me hizo dudar de mi
12	interesante porque la profesora explieta todos las cosas que necesitamos para aprender español. Esta día
15	fue interesante, mi profesora nos enseñó unas cosas importante sobre mi curso.
17	untas. El también era muy simpatico y me dijo las cosas que no entendaba. Eran las 11, descasamos,
19	antes de la pausa, XX había enseñado unas cosas/reglas sobre como usar el subjuntivo de español.
21	miga sientamamos juntas. La profesora explico la cosas, me senté me primer día de clase de español feliz

Tabla 94. Output de la herramienta *Concordance* para el sustantivo «cosas» M-F / M-1D

Lo primero que vemos de las concordancias del sustantivo «cosas» (tabla 94) es el predominio de la función referencial en la mayoría de los ejemplos. Verificamos lo visto en el apartado de agrupamientos, la presencia de formas para expresar la cantidad: «*pudo suceder muchas cosas interesante*», posición 7-«F»; «*por eso el primer día ya sabía la mayoría de cosas*», posición 9-«1D»; «*la profesora explieta todas las cosas que necesitamos*», posición 12-«1D». Pero también encontramos ejemplos de concordancias donde la función expresiva del lenguaje destaca: «*tuve claro que una de las cosas que realmente me apasionaban era el aprendizaje*», posición 6-«1D». Este ejemplo no hubiera sido posible verlo sin la herramienta *Concordance*, pues en agrupamientos lo normal es mostrar combinaciones de tres o cuatro palabras. Otra concordancia que cumple con la función expresiva y que rescatamos gracias a la herramienta *Concordance* es «*Tambien mi*

*amiga y yo nos gustaban las cosas de la tienda*», posición 25-«F». Vemos que la expresión de sentimientos implica a la primera persona del plural, sujeto-emisor y una tercera persona.

«F»	
amigos	
1	comenté via Facebook lo que me había ocurrido. Mis amigos no creyeron una palabra.
4	fueron muy amables conmigo y me llevaron con mis amigos. Finalmente, pude ir a cenar con ellos, pero
8	pareja. Canelo había pensado que ellos eran solo amigos hasta que vió este aspecto. Juan, el dueño de
13	elo, mi perro, cuando de repente... ve a uno de mis amigos que no había visto desde hace cinco años.
16	El día del mi cumpleaños me llamó muchos amigos de Roma para hacerme felicidad. También si
24	en lugar de las 2 por la tarde. La fiesta que mis amigos hicieron para mi fue superfantastico.
34	nos decidimos a regresar a casa y uno de mis amigos me dijo que puede traerme en moto. No estaba
42	vio había vuelto de trabajar, salimos con algunos amigos para tomar una cerveza y comer una tarta.
51	cumpleaños, pero no hubo mucha fiesta, pues mis amigos estaban muy entretenidos con los partidos
56	ado fue mi cumpleaños yo quedé sola ni familia ni amigos, ni salir, me fui muy triste. El día como eso no
60	abamos un idioma extranjero me parecía que ibamos amigos hacía mucho tiempo. Desde entonces hemos

«1D»	
amigos	
2	para mí. No conocía a nadie, pero enseguida hice amigos porque todos hablamos catalan entre nosotros
5	o de la clase (los demás alumnos eras mis mejores amigos hasta el momento) no hubiera tenido la
9	esté nivel. por fin mi gusta estar aquí con mis amigos mis prefosoras.
11	o como da la clase. También he encontrado muchos amigos, como Debra es brasileña, Jaruan es
14	horas cada días de lunes a viernes. Hacia casi 30 amigos en la clase. Muchas personas eran de china
16	muchos compañeros nuevos. Y me encontré unos amigos viejos. Me sentí muy bien. El primer
18	en esta escuela, deseo que pueda encontrar muchos amigos. Oh, quizá un de los amigos me invitará a
21	muchos nuevos amigos y me encontré unos viejos amigos. En esa clase hacíamos un juegos, para
22	clase hacíamos un juegos, para buscar unos nuevos amigos. Escribamos las informaciones nuestras.
24	a poco estuve relajado. En clase conocí a algunos amigos de diferentes países. Me gustó era como volver

Tabla 95. Output de la herramienta *Concordance* para el sustantivo «amigos» M-F / M-1D

En cuanto al sustantivo «amigos», encontramos en el *output* de concordancias las combinaciones ya señaladas en el apartado de agrupamientos («amigos + de + sustantivo», para indicar la procedencia, como en las posiciones 16-«F», 24-«1D»; «amigos + verbo», para señalar estado o acción, vemos ejemplos en las posiciones 1-«F», 51-«F»). Comprobamos que estos agrupamientos se insertan en unos contextos que cumplen con la función referencial, pues simplemente informan sin entrar en expresión de sentimientos o emociones. Sin embargo, encontramos ejemplos de concordancias en los que sí hay expresión de deseos, sentimientos, emociones o valoración, es decir, en las que el sujeto-emisor expresa su actitud ante lo dicho. Vemos ejemplos en las dos muestras: «*La fiesta que mis amigos hicieron para mi fue superfantastico*», posición 24-«F»; «*fue mi cumpleaños yo quedé sola ni familia ni amigos, ni salir, me fui muy triste*», posición 56-«F»; «*por fin mi gusta estar aquí con mis amigos mis prefosoras.*», posición 9-«1D»; «*deseo que pueda encontrar muchos amigos*», posición 18-«F».

«F»	
tiempo	
1	mi novio. Me enfadé ya que me hizo pasar un mal tiempo, pero al final acabé perdonándole porque
5	hasta encontrarlo. Divagué sin rumbo durante largo tiempo, sin obtener resultados positivos. En un abrir
8	pelo, etc., me iba a retrasar mucho y me faltaría tiempo para llegar puntual. Mis desgracias no acababan
13	chas cosas y al final nos paseamos hablando mucho tiempo. Después de escuela mi amiga estudió en la
18	casa. A veces vino su abuela, pero no quedó mucho tiempo con él, y se marchó. Yo vi el niño, jugaba la
21	ente... visto una amiga, que no he visto hace mucho tiempo. Ella tuvo una bicicleta nueva, muy bonita.
24	ara mi fue superfantástico. Y para venir llegué a tiempo para embarquer y facturar mi maleta.
32	o, ha sido mi mejor fiesta de cumpleaños: el todo tiempo a top y cerca de las personas a las que quiero
40	optaba. Sabiaba junto, comiaba, jugaba casi todos tiempo acompañaba. No puedo olvidar Canelo.
59	reloj que eran a las cuatro menos media hora del tiempo que entraba, Menos mal, tenía tiempo.

«1D»	
tiempo	
1	antes, durante y después. Llegué a la clase con tiempo. El edificio estaba lleno de alumnos indiferentes
5	extranjera y por poder hablarla, pero eso requería tiempo y debería esperar para ponerla en práctica.
8	de Ciutadella, y como no lo conozco, fui con tiempo para poder encontrar la clase. Después de
11	y, sinceramente, no fue tan malo como creía. El tiempo pasó rápido, hecho que agradecí porque tenía
14	sentía como un pez fuera del agua. Al cabo de un tiempo y tras numerosas horas de clases particulares,
17	al nivel tercero i ahora tengo mas tiempo i tambien me dedico un poco más.
20	lo pero yo estaba muy nervioso por que hace mucho tiempo no he estudiado en grupos.
22	el autobus y el metro fue al EOI. Contai mucho tiempo. En primera clase, las personas explica
24	de clase fue en septiembre de 2008. Habia buen tiempo. Yo fui a la escuela en autobús.
27	hablamos sobre cultura, interesa, comida, etc. el tiempo pasa pronto, el lunes próxima será ultima clase.

Tabla 96. Output de la herramienta Concordance para el sustantivo «tiempo» M-F / M-1D

En la tabla superior mostramos las concordancias para el sustantivo «tiempo». En ella comprobamos que los agrupamientos que hemos señalado en el apartado anterior para cuantificar y que cumplían con la función referencial del lenguaje, aquí también cumplen con esta función al limitarse a informar de lo que sucede. Presentamos a continuación unos ejemplos para ilustrarlo: «*Divagué sin rumbo durante largo tiempo,*», posición 5-«F»; «*A veces vino su abuela, pero no quedó mucho tiempo con él,*», posición 18-«F»; «*yo estaba muy nervioso por que hace mucho tiempo no he estudiado en grupos.*», posición 20-«1D».

Para expresar en qué momento se realiza una acción, en *agrupamientos* hemos señalado la presencia de las locuciones adverbiales «a tiempo para» y «con tiempo para». En la tabla de concordancias, vemos estas locuciones en contexto y comprobamos que cumplen con la función referencial del lenguaje. Ejemplo de ello lo encontramos en las siguientes concordancias: «*Y para venir llegué a tiempo para embarquer y facturar mi maleta,*», posición 24-«F»; «*Llegué a la clase con tiempo,*», posición 1-«1D»; «*fui con tiempo para poder encontrar la clase.*», posición 27-«1D».

Asimismo, localizamos en la tabla de concordancias dos ejemplos en los que se expresa cómo avanza el «tiempo»: «*El tiempo pasó rápido, hecho que agradecí,*», posición 11-«1D», «*el tiempo pasa pronto, el lunes próxima será ultima clase.*», posición 11-«1D». En este último ejemplo, vemos que el agrupamiento que se produce («tiempo pasa pronto»),

no es un agrupamiento que podamos considerar habitual en la lengua española, pero que en contexto, se entiende que el sujeto quiere expresar que el tiempo ha pasado de una manera rápida. Para expresar el tiempo transcurrido, encontramos el uso de la locución adverbial «al cabo de» («*Al cabo de un tiempo y tras numerosas horas*», posición 14-«1D»).

En todas las concordancias que hemos ido mencionando, comprobamos que la función del lenguaje que domina es la referencial.

«F»	
<i>gente</i>	
1	lugar y siempre nos encontrábamos con la misma <b>gente</b> , día tras día, pero el domingo alguna cosa
5	sorprendió su habilidad y fuerza, <b>no había mucha gente</b> capaz de defenderse así, por no decir nadie.
7	a volar todas las hojas caídas de los árboles, la <b>gente</b> se tenía que esconder para no salir volando
11	y en el salón estaban mi toda la familia. <b>Toda gente</b> me dice: “cumpleaños feliz”. Yo sentí 1 sofá
15	pueblo un poco más. “Por que ahora aquí <b>hay mucha gente</b> ”, pensando acercé a la gente. La gente llevaba
19	hospital y por eso accidente hacia tráfico y la <b>gente</b> llegaba tarde a su trabajo. Aunque el gobierno
23	es un país que viajar muy bien. Cada año <b>mucha gente</b> de diferente país viene aquí, muchas se hablan
28	die con el perro desconocido. <b>Tampoco había mucha gente</b> porque los domingos suelen salir de casa mas
35	grandes y con miedo de el. Además, mas venia mas <b>gente</b> de policia para ayudar. Estaba una ambulancia.
41	cabina del tren por eso me parecía que <b>toda la gente</b> estaban escuchando nuestros dialogos.

«1D»	
<i>gente</i>	
1	alma. Era mi primer año allí y aunque <b>había mucha gente</b> de mi antiguo colegio, la mayoría eran chicos
5	me cuesta establecer una conversación con <b>gente</b> que no conozco y más, establecer una amistad.
9	Recuerdo entrar en el aula el primer día y ver <b>gente</b> de otras clases de mi curso a los que no había
14	llegué al salón principal, ya estaba lleno de <b>gente</b> . Como no conocía a nadie, estaba un poco nerviosa
23	nuestras vidas. Después de la clase conocí a otra <b>gente</b> de diferentes países del mundo. Fue un día
30	primero día. Cuando llegué a la escuela, <b>hay mucha gente</b> estuvieron esperando. Esperé un momento, una
33	os pidió a hacer un ejercicio para conocer a otra <b>gente</b> de la clase. Conocí a tres personas que se
37	y tenía mucha paciencia. En mi clase fue la <b>gente</b> de todo el mundo, de Europa, de EEUU,
41	escuela en autobús. Cuando llegué ya <b>había mucha gente</b> aunque fue muy temprano a la clase.
43	o nerviosa. Cuando llegué la escuela, <b>había mucha gente</b> estaba esperando. La mayoría son extranjeros,

Tabla 97. Output de la herramienta *Concordance* para el sustantivo «gente» M-F / M-1D

En la tabla de concordancias del sustantivo «gente», al igual que en el apartado de *agrupamientos*, comprobamos que en el 50% de los casos se hace referencia a la cantidad. Si en aquel aparecían los adjetivos «mucho», «poco» o «bastante», en los ejemplos extraídos del *output* de concordancias, encontramos «mucho» y «todo»: «*no había mucha gente capaz de defenderse así*», posición 5-«F»; «*Toda gente me dice: “cumpleaños feliz”.*», posición 11-«F»; «*Cada año mucha gente de diferente país viene aquí*», posición 23-«F»; «*por eso me parecía que toda la gente estaban escuchando nuestros dialogos.*», posición 41-«F»; «*y aunque había mucha gente de mi antiguo colegio, la mayoría eran chicos*», posición 1-«1D»; «*Cuando llegué ya había mucha gente aunque fue muy temprano a la clase.*», posición 41-«1D». Como hemos señalado en el apartado de *agrupamientos*, la función del lenguaje que cumple estas concordancias para expresar cuantificación está más próxima de

la expresiva que de la referencial, pues el sujeto-emisor valora desde su particular punto de vista la *cantidad* de gente.

### **Resumen de la pregunta 1:**

Para responder a esta primera pregunta, hemos recurrido a la aplicación informática AntConc y a sus herramientas *WordList* (frecuencias), *Cluster* (agrupamientos) y *Concordance* (palabras en contexto). Con ellas, hemos analizado los dos tipos de producciones escritas correspondientes a las muestras «F» y «1D».

En el análisis de frecuencias, hemos visto que las palabras más frecuentes en ambas muestras, dentro de las diez primeras posiciones, son las palabras funcionales, pronombres, preposiciones, artículos y conjunciones. También hemos visto que aparece el sustantivo «clase» en la muestra «1D», pero eso se debe a que formaba parte de la instrucción para escribir el texto. Al comparar las dos muestras del corpus de nuestra investigación con el corpus CREA, hemos comprobado que mientras en nuestro corpus las dos primeras posiciones están ocupadas por las preposiciones «a» y «de», respectivamente, en el corpus CREA es la preposición «de» la que ocupa la primera posición y la preposición «a» no aparece hasta la posición siete. Asimismo, hemos comprobado la coincidencia con otros estudios, el de Cantos y Sánchez (2011) y el de Muñoz Lobo (2012), y con el mismo corpus CREA, que las cincuenta primeras posiciones de frecuencias corresponden a palabras funcionales.

Para evitar que en las primeras posiciones del listado de frecuencias apareciera ese tipo de léxico, elaboramos un listado de palabras que no deseábamos cuantificar (*stopwords*), como las palabras que formaban parte del enunciando de las tareas y algunas palabras funcionales. Tras volver a aplicar la herramienta *WordList*, comprobamos que en la muestra «1D» era posible deducir el tema del texto pues el léxico más frecuente pertenecía al campo semántico de la educación (profesora, curso, escuela, alumnos, idioma, clase, estudiar, aprender, etc.). Esto enlaza con lo expuesto por Cruz (2012:39) para referirse a las frecuencias de un texto. La autora señala que el acceso a las palabras léxicas facilita conocer de una forma rápida el contenido del texto.

Sin embargo, en la muestra «F» no ha sido posible encontrar un hilo conductor a partir de las palabras léxicas. Esto es así por tratarse de un tema más libre y, por tanto, más abierto a la imaginación.

Pero al comprobar si había similitud entre ambas muestras, sí se identificaron un total de once palabras, dentro de las veinticinco primeras posiciones, que coincidían. A partir de ellas se realizaron los análisis de agrupamientos y concordancias.

Así, el análisis de agrupamientos y de concordancias permitió verificar que la muestra «1D» presenta mayoritariamente un léxico que denota deseos, sentimientos, emociones o valoraciones en primera persona (*mi pasión; mi interés; me gusta mucho; me asustó*), por tanto, es una muestra en la que la función expresiva del lenguaje se presenta de una forma explícita en la mayoría de los casos. Mientras que en la muestra «F» la función del lenguaje que predomina es la referencial, con agrupamientos y concordancias que simplemente *informan* de algo, sin mostrar un posicionamiento ante lo mencionado. No obstante, también encontramos agrupamientos y concordancias que cumplían con la función expresiva del lenguaje (*me asusté; me moría de miedo*).

**Pregunta 2: Entre los sujetos no nativos y los sujetos nativos que participaron en la investigación, ¿hay diferencias léxicas o sintácticas entre las producciones de la muestra «F» (inventar una historia) frente a la muestra «1D» (relatar una experiencia personal)?**

Para responder a la segunda pregunta de investigación, vamos a seguir los mismos pasos que en la pregunta uno. A partir de la aplicación informática AntConc y las herramientas (a) *WordList*, (b) *Cluster* y (c) *Concordance*, vamos a ver si entre las muestras del grupo de sujetos NN y N hay diferencias relevantes en el léxico que utilizan y cómo lo utilizan.

Así, el análisis que presentamos a continuación se estructura también en tres fases:

(i) Para empezar, vamos a ver cuáles son las 50 primeras palabras más frecuentes en el grupo de NN, el grupo de N y el corpus CREA (Corpus de Referencia del Español Actual). A continuación, mostraremos las 50 primeras palabras más frecuentes, pero a partir de la lista de *stopwords* confeccionada para evitar la presencia de aquellas palabras que no interesa cuantificar.

(ii) En la segunda fase, y a partir de las palabras léxicas y funcionales más frecuentes, veremos de qué manera se combinan para comprobar si se produce el mismo tipo de agrupamientos en ambas muestras.

(iii) Finalmente, también a partir de las palabras léxicas y funcionales más frecuentes, observaremos en qué tipo de contexto suelen aparecer y si coincide o no en ambas muestras.

### Fase 1: Frecuencias.

En la siguiente tabla presentamos las 50 primeras posiciones del listado de frecuencias de la muestra del grupo de NN, del grupo de N y del corpus CREA.

Grupo «NN»			Grupo «N»			Corpus CREA		
1	1171	a	1	1544	a	1.	de	9,999,518
2	1125	de	2	1044	de	2.	la	6,277,560
3	982	y	3	870	que	3.	que	4,681,839
4	830	la	4	707	y	4.	el	4,569,652
5	773	que	5	685	la	5.	en	4,234,281
6	669	el	6	524	el	6.	y	4,180,279
7	638	mi	7	491	me	7.	a	3,260,939
8	624	en	8	417	en	8.	los	2,618,657
9	465	un	9	366	mi	9.	se	2,022,514
10	464	no	10	323	no	10.	del	1,857,225
11	456	me	11	314	un	11.	las	1,686,741
12	383	muy	12	287	con	12.	un	1,659,827
13	361	una	13	256	se	13.	por	1,561,904
14	333	con	14	252	por	14.	con	1,481,607
15	302	estaba	15	236	n	15.	no	1,465,503
16	299	por	16	228	una	16.	una	1,347,603
17	261	pero	17	214	s	17.	su	1,103,617
18	258	yo	18	197	al	18.	para	1,062,152
19	255	para	19	196	hab	19.	es	1,019,669
20	248	cuando	20	189	estaba	20.	al	951,054
21	220	clase	21	187	lo	21.	lo	866,955
22	207	d	22	183	m	22.	como	773,465
23	199	era	23	178	d	23.	más	661,696
24	186	se	24	175	canelo	24.	o	542,284
25	171	fue	25	175	los	25.	pero	450,512
26	166	los	26	172	clase	26.	sus	449,870
27	157	es	27	169	cuando	27.	le	413,241
28	153	porque	28	153	era	28.	ha	380,339
29	147	amiga	29	152	fue	29.	me	374,368
30	143	mucho	30	136	pero	30.	si	327,480
31	135	ella	31	132	para	31.	sin	298,383
32	135	su	32	126	las	32.	sobre	289,704
33	133	al	33	121	muy	33.	este	285,461
34	126	las	34	108	o	34.	ya	274,177
35	124	perro	35	108	perro	35.	entre	267,493
36	119	primer	36	106	yo	36.	cuando	257,272
37	117	lo	37	101	ya	37.	todo	247,340
38	114	nos	38	96	primer	38.	esta	238,841
39	109	casa	39	95	como	39.	ser	232,924
40	107	como	40	93	nos	40.	son	232,415
41	101	canelo	41	90	ten	41.	dos	228,439
42	97	domingo	42	82	as	42.	también	227,411

43	95	día	43	81	del	43.	fue	223,791
44	95	s	44	80	ma	44.	había	223,430
45	93	o	45	79	ana	45.	era	219,933
46	87	del	46	77	repente	46.	muy	208,540
47	87	pasado	47	72	domingo	47.	años	203,027
48	87	poco	48	72	su	48.	hasta	202,935
49	82	le	49	71	casa	49.	desde	198,647
50	82	repente	50	71	os	50.	está	194,168

Tabla 98. *Output* de la herramienta *WordList*.. NN – N

Comprobamos que dentro de las diez primeras posiciones, las palabras que encontramos son todas ellas funcionales, es decir, sin contenido semántico. Como ya vimos en la pregunta uno, está dentro de lo normal que las palabras funcionales ocupen las primeras posiciones en un análisis de frecuencias (Cantos y Sánchez 2011). Si observamos las diez primeras palabras, comprobamos que siete (*a, de, y, la, que, el, en*) coinciden en los dos grupos de nuestro corpus y en el corpus CREA, aunque no siempre ocupando la misma posición. Así, vemos, por ejemplo que la preposición «a» ocupa la posición uno tanto en el grupo de NN como en el de N, mientras que en el corpus CREA la encontramos en el puesto siete. Otro elemento que llama la atención es la coincidencia de la posición de «que» en el grupo de N y en el corpus CREA, en la posición tres, mientras que en el grupo de NN aparece en la posición cinco. El caso más llamativo, no obstante, dentro de este grupo de diez palabras funcionales, es la presencia del posesivo «mi» en las posiciones siete y nueve en el grupo de NN y en el grupo de N, respectivamente, mientras que en el corpus CREA no aparece hasta la posición cincuenta y uno, fuera ya de la selección que hemos realizado.

En cuanto a las palabras léxicas de este listado de frecuencias, observamos que la primera forma verbal la encontramos en la posición quince y en la posición diecinueve en el grupo de NN y de N, respectivamente. La forma corresponde al pretérito imperfecto del modo indicativo del verbo «estar» en ambos casos, «estaba», verbo que carece de contenido semántico. En el corpus CREA, la primera forma verbal aparece en la posición treinta y nueve, como ya indicamos en la pregunta uno. En relación a los sustantivos, el corpus CREA solo registra uno en las primeras cincuenta posiciones y aparece en la cuarenta y siete. Se trata del sustantivo «años». En cambio, el grupo NN y N presentan un total de seis y cinco sustantivos, respectivamente. Pero se debe indicar que salvo uno, «casa», el resto corresponde al enunciado de la tarea.

Para centrarnos en las palabras léxicas y en las palabras funcionales de interés para nuestro estudio, aplicamos nuevamente la herramienta *WordList*, pero con la lista de

*stopwords* elaborada para eliminar aquellas palabras que no son de utilidad para la investigación.

A continuación presentamos la lista de frecuencias con la lista de *stopwords*.

Grupo «NN»			Grupo «N»		
1	638	mi	1	491	me
2	456	me	2	366	mi
3	302	estaba	3	189	estaba
4	258	yo	4	152	fue
5	171	fue	5	106	yo
6	114	nos	6	93	nos
7	109	casa	7	79	ana
8	95	día	8	71	casa
9	87	pasado	9	64	mis
10	87	poco	10	61	profesor
11	78	gente	11	53	poco
12	76	amigos	12	48	francés
13	76	mis	13	42	idioma
14	70	profesora	14	40	ser
15	68	dos	15	38	curso
16	66	tiempo	16	38	todos
17	65	iba	17	38	vez
18	64	dijo	18	38	vi
19	63	fui	19	37	nunca
20	60	todos	20	36	lengua
21	58	escuela	21	35	inglés
22	53	mas	22	34	clases
23	53	muchos	23	32	iba
24	53	tambien	24	32	momento
25	52	tenia	25	31	recuerdo
26	51	hacer	26	30	después
27	50	después	27	30	dos
28	48	entonces	28	30	hombre
29	47	chica	29	30	lado
30	47	hombre	30	30	él
31	47	tranquilamente	31	29	hacer
32	45	cada	32	29	miedo
33	45	mucha	33	29	tan
34	44	muchas	34	28	ir
35	43	cosas	35	28	tarde
36	41	momento	36	27	nerviosa
37	40	esperan	37	27	tanto
38	40	fiesta	38	26	durante
39	40	había	39	26	mismo
40	40	siempre	40	25	aprender
41	39	mejor	41	25	hacia
42	39	persona	42	25	hizo
43	39	vida	43	25	llegar
44	38	donde	44	25	solo
45	38	fuimos	45	25	ver
46	38	nosotros	46	24	cuenta
47	37	chico	47	24	nuevo
48	37	miedo	48	24	sus
49	36	ir	49	24	tiempo
50	36	luego	50	23	di

Tabla 99. Output de la herramienta *WordList* con *stopwords* NN – N

Comprobamos en la tabla 99 que las cinco palabras que ocupan las primeras posiciones son las mismas en los dos grupos (*mi, me, estaba, yo, fue*), aunque con un orden diferente, como se puede observar. En cuanto al número de sustantivos presentes dentro de las primeras 50 posiciones, vemos que el grupo NN (14) supera en cuatro al grupo N (10). De ese total de sustantivos, observamos que solo coinciden cinco (*casa, hombre, momento, miedo, tiempo*). Así, en el análisis de *agrupamientos y concordancias*, nos vamos a centrar, por un lado, en el posesivo singular de primera persona «mi», del pronombre complemento «me» y del complemento sujeto «yo». Por otro lado, focalizaremos en los sustantivos *casa, hombre, momento, miedo, tiempo*. Las formas verbales de «ser» y «estar» no las vamos a tener en consideración en este análisis, pues, como ya hemos señalado más arriba, son verbos sin contenido semántico.

## Fase 2: Agrupamientos.

Grupo «NN»			Grupo «N»		
<i>mi</i>			<i>mi</i>		
20	2	a lado de mi	20	2	habla. Por mi parte
21	2	a <b>mi amiga</b> y	21	2	<b>mi interés</b> ha traspasado
22	2	a mi novio para	22	2	mi parte fue una
23	2	a <b>mi perro</b> y	23	2	mi pasión por
24	2	<b>al mi lado</b> . Era	24	2	momento mi pasión
25	2	<b>cerca de mi casa</b>	25	2	Por mi parte fue
26	2	de <b>mi perro</b> y	26	2	repente vi a mi
27	2	dicho que <b>mi padre</b>	27	2	ver a mi perro
28	2	en frente de mi	28	2	y <b>mi interés</b> ha
29	2	En mi primer d	29	1	<b>a mi alrededor</b>
30	2	era <b>mi cumpleaños</b> , pero	30	1	<b>a mi izquierda</b>
31	2	fiesta en <b>mi casa</b>	31	1	<b>a mi lado</b>
32	2	mejor amiga y mi	32	1	a alguien de mi
33	2	<b>mi amiga</b> llegó y	33	1	a alguien. Para mi
34	2	<b>Mi amiga</b> me dijo	34	1	a Canelo a mi
35	2	<b>mi cumpleaños</b> y	35	1	a causa de mi
36	2	<b>mi madre</b> y padre	36	1	a comparando <b>mi edad</b>
37	2	<b>mi perro</b> estaba llamando	37	1	a conocer mi tuit
38	2	orvido de <b>mi cumpleaños</b>	38	1	a de <b>mi cumpleaños</b>
39	2	que <b>mi padre</b> esta	39	1	a deshacer <b>mi camino</b>
40	2	se orvido de mi	40	1	a disponer durante mi
41	2	sido <b>mi mejor amiga</b>	41	1	a en <b>mi cama</b>
42	2	tienda, <b>mi amiga</b> y	42	1	a era a mi
43	2	un amigo de mi	43	1	a era <b>mi prima</b>
44	2	una fiesta en mi	44	1	a imaginar <b>mi vida</b>

Tabla 100. Lista de agrupamientos del posesivo «mi» NN - N

En la tabla 100 mostramos los agrupamientos con el posesivo átono de primera persona singular «mi». Como ya hemos señalado en el apartado 6.2.1 en la sección de *agrupamientos*, la función de las formas átonas de los posesivos es vincular el sustantivo al que acompañan con la persona del discurso, estableciendo una relación que, dependiendo

del contexto en el que se desarrolle el discurso, puede interpretarse como posesión, parentesco, relación de pertenencia a un grupo, etc.

En esta sección, como en la del apartado 6.2.1, también hemos saltado las primeras posiciones, pues remitían a la instrucción de la tarea. Así, hemos seleccionado desde la posición en la que ya no era un calco exacto del enunciado.

Observamos en la tabla de agrupamientos que, mientras que en el grupo «NN» encontramos hasta catorce agrupamientos en los que predomina el uso de la forma átona del posesivo «mi + sustantivo» dentro de una función referencial para expresar posesión («*mi perro*», posición 23-«NN»; «*mi casa*», posición 31-«NN») o parentesco («*mi amiga*», posición 21-«F»; «*mi padre*», posición 27-«NN»); en el grupo «N», solo aparecen dos («*mi cama*», posición 41-«N»; «*mi prima*», posición 43-«N»).

Comprobamos que hay dos agrupamientos en la muestra del grupo N que no aparecen en la del grupo NN. Estas son, por un lado, «mi + interés» (posiciones 21-N y 28-N), agrupamiento que se inscribe dentro de la función expresiva del lenguaje al denotar implicación personal con la elección de léxico valorativo. Aquí el sujeto-emisor es el punto de referencia y como tal expresa su postura ante algo. Por otro lado, aparece el agrupamiento «por + mi + parte» (posiciones 20-N y 25-N), locución adverbial para expresar una opinión.

También llama la atención los agrupamientos para formar las locuciones adverbiales («*a mi alrededor*», posición 29-«N»; «*a mi izquierda*», posición 30-«N»; «*a mi lado*», posición 31-«N»). Comprobamos que este último también aparece en el grupo NN («*al mi lado*», posición 24-«NN»). Encontramos, sin embargo, un agrupamiento en el grupo NN que no aparece en el grupo N, la locución preposicional «*cerca de mi casa*», posición 25-«NN». Como vemos, todos estos agrupamientos para expresar localización espacial cumplen con la función referencial del lenguaje.

En la tabla 101 presentamos los agrupamientos del pronombre complemento «me». Como vemos, se combina con verbos pronominales como «sentirse» (posiciones 15, 16, 17-«NN»); verbos transitivos como «decir» (posición 1-«NN») o verbos pseudoimpersonales como «gustar, encantar, parecer, interesar, asustar, apasionar»: «*no me gusta mucho*», posición 20-«NN»; «*Este tema me asusta*», posición 18-«N»; «*Me apasiona la mezcla*», posición 24-«N». Pero lo llamativo entre las muestras del grupo NN y N es la elección de verbos pseudoimpersonales para expresar sentimientos. Mientras que el grupo NN solo recurre al verbo «gustar», por un lado, y, por otro, al pronominal «sentirse», vemos que el

grupo N presenta una gama más amplia («gustar, encantar, interesar, asustar, apasionar»). Este es un dato interesante que hay que tener presente a la hora de proporcionar *input* a los estudiantes de ELE. No obstante, a pesar de esta diferencia en el léxico utilizado, comprobamos que el número de agrupamientos que cumplen con la función expresiva del lenguaje es similar en ambas muestras, siete para el grupo NN y seis para el grupo N.

También es relevante el agrupamiento del pronombre complemento «me» con el verbo «decir» para introducir el discurso indirecto. Encontramos 10 casos entre las veinticinco primeras posiciones en la muestra de grupo NN (1, 2, 3, 5, 6, 8, 9, 10, 18 y 23); mientras que en la muestra del grupo N no aparece ninguno.

Finalmente, queremos mencionar la presencia del agrupamiento «*me di cuenta*», posiciones 4-«NN», 3,4, 6-«N», en el que el verbo pronominal «darse» se combina con el sustantivo «cuenta» para formar la locución verbal «darse cuenta de algo».

Grupo «NN»			Grupo «N»		
<i>me</i>			<i>me</i>		
1	4	me ha dicho que	1	9	cuando de repente me
2	4	y me dijo que	2	8	que me había
3	3	amiga me dijo que	3	6	me di cuenta de
4	3	Me di cuenta que	4	6	me di cuenta que
5	3	me dijo que no	5	5	de repente me di
6	3	me dijo que tenía	6	5	repente me di cuenta
7	3	me gusta mucho la	7	3	de que me había
8	3	no me dijo nada	8	3	sabía que me
9	2	El me dijo que	9	2	a la vez me
10	2	ella me dijo que	10	2	a lo que me
11	2	entré la escuela, me	11	2	Ahora me interesa la
12	2	la escuela, me sent	12	2	alemán me encanta
13	2	me han puesto al	13	2	cuenta de que me
14	2	me parecía que	14	2	de alemán me
15	2	me sentí un poco	15	2	de repente me encontré
16	2	Me sentía bien	16	2	En ese momento me
17	2	me siento muy bien	17	2	en la clase, me
18	2	Mi amiga me dijo	18	2	Este tema me asusta
19	2	muy mal y me	19	2	ha pasado mucho. Me
20	2	no me gusta mucho	20	2	la lengua. Ahora me
21	2	no me gusta nada	21	2	lengua. Ahora me interesa
22	2	países. Me senté	22	2	lo que me dec
23	2	pero no me dijo	23	2	lo que me hab
24	2	primer día me	24	2	Me apasiona la mezcla
25	2	Y me gusta mucho	25	2	me asusta, tengo que

Tabla 101. Lista de agrupamientos del pronombre complemento «me» NN - N

Grupo «NN»			Grupo «N»		
yo			yo		
1	4	mi amiga y yo	1	2	a correr y yo
2	2	amiga y yo nos	2	2	años. Yo cursaba
3	2	María y yo	3	2	Canelo y yo estuvimos
4	2	que yo estaba esperando	4	2	Canelo y yo nos
5	2	yo con mi amiga	5	2	Yo conocía a
6	2	Yo no podía celebrar	6	1	a como yo me
7	2	Yo no sabía que	7	1	a como yo. La
8	1	a 4 alumnos, yo era	8	1	a correr, y yo
9	1	a 9, pero yo no	9	1	a decir eran : Yo
10	1	a algún sitio. Yo	10	1	a llegar. Yo conseguí
11	1	a calor. Yo sent	11	1	a llegó y yo
12	1	a casa yo siempre	12	1	a lo que yo
13	1	a explicar que yo	13	1	a planearlo todo. Yo
14	1	a hospital. Yo también	14	1	a un chico. Yo
15	1	a la ambulancia. Yo	15	1	a unos chicos. Yo
16	1	a la escuela yo	16	1	a, yo les dije
17	1	a llover y yo	17	1	a. Canelo y yo
18	1	a matricular. Pero yo	18	1	al catalán. Yo
19	1	a muy divertido. Yo	19	1	al parque hoy. Yo
20	1	a nosotros. Entonces yo	20	1	alemán y yo
21	1	a pasear juntas yo	21	1	alto del que yo
22	1	a percibido que yo	22	1	han desaparecido, y yo
23	1	a pero cuando yo	23	1	antes Canelo y yo
24	1	a pobre! Yo, no	24	1	Antes de que yo
25	1	a recogerla. Yo estaba	25	1	aquel entonces yo ten

Tabla 102. Lista de agrupamientos del pronombre sujeto «yo» NN - N

En la tabla superior mostramos los agrupamientos del pronombre sujeto de primera persona singular «yo». Como ya hemos señalado en el apartado 6.2.1, cuando el emisor de un mensaje utiliza de forma explícita el pronombre sujeto lo hace para evitar ambigüedades o resaltar que se trata de ese sujeto y no de otro. Vemos que los agrupamientos que aparecen aquí coinciden con los ya presentados en la sección correspondiente del apartado 6.2.1: (i) «sustantivo + y + yo» (posiciones 1, 2, 3-«NN»; 3, 4, 17, 23-«N»), con la finalidad de identificar a alguien de entre un grupo para despejar dudas acerca de quién se trata. (ii) «yo + verbo» (posiciones 4, 6, 7, 25-«NN»; 2, 5, 10-«N»), para identificar y destacar la figura del sujeto-emisor respecto de otros posibles sujetos.

Asimismo, encontramos el agrupamiento «yo + también» (posición 14-«NN») para transmitir la idea de continuidad del tema del discurso mostrando acuerdo sobre lo ya dicho.

Como vemos, todos estos agrupamientos cumplen con la función referencial del lenguaje, es decir, la función de informar, no hay intención de posicionarse, de mostrar una actitud determinada ante lo que se está explicando, ni de influir en el otro interlocutor.

Grupo «NN»			Grupo «N»		
<i>casa</i>			<i>casa</i>		
1	2	<b>cerca de mi casa</b>	1	2	<b>de vuelta</b> a casa
2	2	comer <b>en su casa</b>	2	1	a casa de Mar
3	2	fiesta <b>en mi casa</b>	3	1	a casa después de
4	1	a casa Canelo estaba	4	1	a casa encontré a
5	1	a casa con montón	5	1	a casa enfurruñada
6	1	a casa de ella	6	1	a casa mucho m
7	1	a casa de un	7	1	a casa muy triste
8	1	a casa en el	8	1	a casa oí un
9	1	a casa en la	9	1	a casa para ver
10	1	a casa las ejercicios	10	1	a casa porque estaba
11	1	a casa meva. Y	11	1	a casa rápidamente
12	1	a casa mi marido	12	1	a casa tras el
13	1	a casa muy tarde	13	1	a casa y comenté
14	1	a casa muy triste	14	1	a casa y decidí
15	1	a casa pero habia	15	1	a casa y me
16	1	a casa sin llaves	16	1	a casa y, al
17	1	a casa y dejé	17	1	a casa y, asustada
18	1	a casa y ella	18	1	a casa, dejé Canelo
19	1	a casa y estaba	19	1	a casa, me inund
20	1	a casa y miraba	20	1	a casa, porque sab
21	1	a casa y uno	21	1	a casa, pues el
22	1	a casa yo siempre	22	1	a casa, ya que
23	1	a casa, estaba muy	23	1	a casa. Misteriosamente, lo
24	1	a casa, he encontrado	24	1	a casa. Pero antes
25	1	a casa, lo dejé	25	1	a <b>como en casa</b>

Tabla 103. Lista de agrupamientos del sustantivo «casa» NN - N

Lo primero que llama la atención de los agrupamientos de «casa» en las primeras veinticinco posiciones tanto del grupo NN como del grupo N es la presencia abrumadora del agrupamiento «a + casa». Luego aparecen dos casos de agrupamiento con la preposición «en + casa» (posiciones 2, 3-«NN»). Como ya hemos señalado en el apartado 6.2.1, la preposición «a» señala destino, dirección, por tanto, suele utilizarse con verbos de movimiento; mientras que la preposición «en» se utiliza para localizar en el espacio, es decir, identificar un lugar concreto. En ambos casos, la función del lenguaje que prevalece en estos agrupamientos es la referencial, pues la finalidad es la de informar.

Aparece, igualmente, la locución preposicional «cerca de + sustantivo», posición 1-«NN» para expresar proximidad entre los elementos que unen, y la locución adverbial «de vuelta + sustantivo», posición 1-«N» para señalar que se regresa al punto de origen. También en estos agrupamientos vemos que la función del lenguaje que domina es la referencial.

Finalmente, destacamos el agrupamiento «*como en casa*», posición 25-«N», en el que la función expresiva del lenguaje queda clara; aquí el sujeto-emisor manifiesta de esta manera su actitud ante algo, expresando la idea de comodidad, bienestar.

Grupo «NN»			Grupo «N»		
<i>momento</i>			<i>momento</i>		
1	2	hijos. <b>En ese momento</b>	1	4	<b>partir de ese momento</b>
2	1	a un momento que	2	2	<b>Desde ese momento</b> mi
3	1	ahora, <b>en ese momento</b>	3	2	<b>En ese momento</b> me
4	1	alegrar con el momento	4	2	ese momento mi pasión
5	1	aquel momento devolvió aquella	5	2	lengua. Desde ese momento
6	1	Aquel momento mi perro	6	2	momento mi pasión
7	1	aquel momento s <sup>7</sup> que	7	2	partir de aquel momento
8	1	aquel momento, de repente	8	1	academia? Por un momento
9	1	aquel momento. Por un	9	1	al momento. Fui hacia
10	1	Aquell momento no tuvimos	10	1	algun momento determinado necesitaba
11	1	aquell momento. Claro, las	11	1	amigos hasta el momento
12	1	bolsa. <b>En ese momento</b>	12	1	aquel momento empecé a
13	1	cabo de un momento	13	1	aquel momento, la profesora
14	1	caliente. Fue un momento	14	1	ayuda. <b>En ese momento</b>
15	1	cansada. In quel momento	15	1	ayudaron en todo momento
16	1	casa. En el momento	16	1	búsqueda. <b>En ese momento</b>
17	1	chino, <b>en ese momento</b>	17	1	cachorro. <b>En ese momento</b>
18	1	ciudad. <b>En aquel momento</b>	18	1	comprar <b>en este momento</b>
19	1	con el momento que	19	1	conductor desapareció al momento
20	1	curar. Aquel momento mi	20	1	cuerpo. <b>En el momento</b>
21	1	de cado momento haora	21	1	<b>desde aquel momento</b> empecé
22	1	de momento. E hijos	22	1	<b>desde aquel momento</b> , la
23	1	de un momento me	23	1	<b>desde ese momento</b> Canelo
24	1	<b>desde hace este momento</b>	24	1	<b>desde ese momento</b> las
25	1	disfrutar de cado momento	25	1	<b>desde ese momento</b> nos

Tabla 104. Lista de agrupamientos del sustantivo «momento» NN - N

En la tabla superior, tabla de agrupamientos del sustantivo «momento», destaca el agrupamiento «preposición + momento» para formar las locuciones adverbiales (i) «*en ese/aquel momento*», posiciones 1, 3, 12, 17, 18-«NN»; 3, 14, 16, 17, 18, 20-«N», para ubicar en un punto concreto del espacio temporal aquello a lo que refiere el discurso. (ii) «*desde ese/aquel momento*», posiciones 24-«NN»; 2, 21, 22, 23, 24, 25-«N», para indicar el punto de origen en el espacio temporal. En ambos agrupamientos, la función del lenguaje que domina es la referencial.

En cuanto a los agrupamientos con el sustantivo «hombre», vemos en la tabla 105 que destacan las combinaciones para (i) describir («adjetivo + hombre»; «hombre + adjetivo») o (ii) identificar («hombre + de + sustantivo»). Ejemplos de descripción con adjetivo pospuesto los tenemos en: «*un hombre especial*», posición 21-«NN»; «*un hombre disfrazado*», posición 12-«N». Con adjetivo antepuesto aparece el mismo agrupamiento en ambos grupos: «*buen hombre*», posición 16-«NN», 18-«N». En estos casos, la anteposición del adjetivo añade un matiz de subjetividad a la descripción, por tanto, entra aquí la función expresiva del lenguaje, ya que el sujeto-emisor del mensaje se posiciona ante lo que dice, expresa su actitud. Mientras que con la posposición se sigue un criterio de objetividad,

simplemente se dice lo que se observa. Así, la función del lenguaje que domina es la referencial.

Para identificar, encontramos los siguientes agrupamientos: «*hombre del bigote*», posición 6-«N»; «*hombre del disfraz*», posición 7-«N». Como vemos en estas combinaciones, también es la función referencial del lenguaje la que domina, pues se trata simplemente de identificar a través de destacar un elemento del ente del que se habla. Hay que resaltar de este agrupamiento que entre las veinticinco primeras posiciones solo lo encontramos en la muestra del grupo N.

Grupo «NN»			Grupo «N»		
<i>hombre</i>			<i>hombre</i>		
1	2	que el hombre estaba	1	2	hombre que había
2	1	a conocido un hombre	2	1	a perseguir al hombre
3	1	a un hombre de	3	1	a un hombre escondido
4	1	ha visto un hombre	4	1	a <b>un misterioso hombre</b>
5	1	a Yang. Un hombre	5	1	a ver al hombre
6	1	acercó un hombre con	6	1	al <b>hombre del bigote</b>
7	1	acercé del hombre. Éste	7	1	al <b>hombre del disfraz</b>
8	1	al hombre que fuera	8	1	al hombre, lanzándolo
9	1	amiga cuando un hombre	9	1	alrededor. Un hombre cay
10	1	asiento. Era un hombre	10	1	amos. El hombre me
11	1	aunque el hombre estaba	11	1	Apareció otro hombre joven
12	1	avisó al hombre que	12	1	apareció <b>un hombre disfrazado</b>
13	1	ayudarle pero el hombre	13	1	Aquel hombre era un
14	1	ayudo. El hombre me	14	1	aquel hombre que hab
15	1	basura. El hombre adentro	15	1	aquel mismo hombre me
16	1	<b>buen hombre</b> queria pararlo	16	1	asesinatos cometidos. Aquel hombre
17	1	calle pero el hombre	17	1	atrapó al hombre, lanz
18	1	coche golpeó un hombre	18	1	<b>buen hombre</b> , había
19	1	comido un otro hombre	19	1	Canelo atrapó al hombre
20	1	con ese hombre puedo	20	1	chico del hombre fue
21	1	conocido <b>un hombre especial</b>	21	1	cometidos. Aquel hombre era
22	1	cuando el hombre me	22	1	como ese mismo hombre
23	1	cuando un hombre me	23	1	como un hombre de
24	1	de basura. El hombre	24	1	con el hombre hacia
25	1	de repente. Un hombre	25	1	conocíamos. El hombre

Tabla 105. Lista de agrupamientos del sustantivo «hombre» NN - N

En la tabla 106 presentamos los agrupamientos con el sustantivo «miedo». Vemos que las asociaciones que aparecen son principalmente para expresar cuantificación. Pero mientras en la muestra del grupo NN la combinación se realiza solo con el adjetivo «mucho» («*mucho miedo*», posiciones 4, 5, 12-«NN»; «*muchísimo miedo*», posición 6-«NN»); en la muestra del grupo N encontramos más variedad de agrupamientos: («*bastante miedo*», posiciones 20, 24-«N»; «*algo de miedo*», posiciones 2, 18-«N»; «*pasé un miedo*», posición 19-«N»). Vemos que son agrupamientos que cumplen con la función expresiva del lenguaje, por un lado, porque el significado semántico del sustantivo ya expresa por sí mismo un

sentimiento y, por otro lado, porque al realizar una valoración sobre la *cantidad* de miedo que se tiene o se siente se está siguiendo un criterio subjetivo, por tanto, el sujeto-emisor muestra su actitud ante algo. Otro agrupamiento que llama la atención en la muestra del grupo NN es «*miraba con miedo*», posiciones 15-«NN», también dentro de la función expresiva del lenguaje.

Grupo «NN»			Grupo «N»		
<i>miedo</i>			<i>miedo</i>		
1	2	<i>miedo de</i> ese gato	1	3	tenía miedo de
2	2	tenía mucho miedo de	2	1	a <b>algo de miedo</b>
3	1	a casa. Tuve miedo	3	1	a de miedo. Pero
4	1	a <b>mucho miedo</b> de	4	1	a <b>desconfianza y miedo</b>
5	1	a <b>mucho miedo</b> , porque	5	1	a ese miedo de
6	1	a <b>muchísimo miedo</b>	6	1	a miedo a no
7	1	a que tener miedo	7	1	a miedo de no
8	1	amiga ten'amos miedo	8	1	a miedo de pensar
9	1	amos miedo de la	9	1	a miedo de que
10	1	automaticamente, con miedo.	10	1	a miedo porque todo
11	1	cabeza. ¡Yo tenía miedo	11	1	a miedo. Antes cuando
12	1	Canelo tenía <b>mucho miedo</b>	12	1	a miedo. No me
13	1	Canelo, probrecito, sentía miedo	13	1	a mucho miedo ya
14	1	casa. Tuve miedo porque	14	1	a m'. El miedo
15	1	chica <i>miraba con miedo</i>	15	1	a tanto miedo que
16	1	chica tiene miedo con	16	1	a vencer el miedo
17	1	con miedo de el	17	1	a, muchas sensaciones: miedo
18	1	con miedo de perro	18	1	<b>algo de miedo</b> , pero
19	1	con miedo. Queríamos	19	1	ansiedad. <b>Pasé un miedo</b>
20	1	con miedo. ¡Fue sólo	20	1	<b>bastante miedo</b> pero creo
21	1	de miedo empezó a	21	1	Canelo y con miedo
22	1	de miedo. De repente	22	1	con miedo miré a
23	1	día fantástico. Tenía miedo	23	1	cual, por el miedo
24	1	dio miedo, entonces empezaba	24	1	daba <b>bastante miedo</b> pero
25	1	dio mucho miedo a	25	1	de miedo! Pero de

Tabla 106. Lista de agrupamientos del sustantivo «miedo» NN – N

Grupo «NN»			Grupo «N»		
<i>tiempo</i>			<i>tiempo</i>		
1	3	mucho tiempo que no	1	1	a llover. El tiempo
2	2	que no tuve tiempo	2	1	a tiempo mi madre
3	2	visto hace mucho tiempo	3	1	a tiempo para llegar
4	1	a buen tiempo. Mi	4	1	a tiempo y deber
5	1	a mucho tiempo. Desde	5	1	a ya un tiempo
6	1	a muy mal tiempo	6	1	a. El tiempo pas
7	1	a pasé el tiempo	7	1	al mismo tiempo estaba
8	1	a tiempo para embarquerer	8	1	al mismo tiempo nos
9	1	a tiempo. Llamé a	9	1	al mismo tiempo que
10	1	a, el tiempo no	10	1	bastante tiempo, as' que
11	1	absolutamente del tiempo. A	11	1	cabo de un tiempo
12	1	ahora tengo mas tiempo	12	1	clase con tiempo. El
13	1	al mismo tiempo, pero	13	1	con el tiempo, empezé
14	1	allí mucho tiempo. Estabamos	14	1	con tiempo para poder
15	1	allí. En su tiempo	15	1	con tiempo. El edificio
16	1	allí por mucho tiempo	16	1	conozco, fui con tiempo
17	1	al mismo tiempo	17	1	creía. El tiempo
18	1	buen tiempo. Me encantó	18	1	de tiempo en el
19	1	buen tiempo, habia mucha	19	1	de un tiempo y
20	1	buen tiempo. Mi amiga	20	1	durante largo tiempo, sin
21	1	buen tiempo. Yo fui	21	1	el tiempo así que
22	1	calle. Normalmente ese tiempo	22	1	el tiempo pasaba muy
23	1	casa estaba mal tiempo	23	1	El tiempo pasó
24	1	casi todos tiempo acompañaba	24	1	el tiempo que corre
25	1	causa del mal tiempo	25	1	el tiempo que llevaba

Tabla 107. Lista de agrupamientos del sustantivo «tiempo» NN - N

Vemos en la tabla 107 los agrupamientos con el sustantivo «tiempo». Ya hemos mencionado que Muñoz Lobo (2012) también analiza este sustantivo en su trabajo por la variedad de colocaciones que presenta, y en la sección de agrupamientos del apartado 6.2.1 hemos mostrado que también en nuestro corpus las combinaciones presentes en las muestras «F» y «1D» son variadas. Al observar las muestras por grupo (NN y N), comprobamos que esas asociaciones se repiten también aquí. Así, encontramos agrupamientos: (i) para expresar cuantificación: «*mucho tiempo*», posición 1, 3, 5, 14, 16-«NN»; «*bastante tiempo*», posición 10-«N»; (ii) para expresar simultaneidad: «*al mismo tiempo*», posición 13, 17-«NN», 7, 8, 9-«N»; (iii) con adjetivos: «*buen/mal tiempo*», posiciones 4, 6, 18, 19, 20, 21, 23-«NN»; (iv) con las preposiciones «a» y «con», para formar las locuciones adverbiales: «*a tiempo (para)*», posición 8, 9-«NN» (indica momento justo), «*con tiempo (para)*», posición 14, 15, 16-«N» (indica antelación). En todos estos agrupamientos comprobamos que la función del lenguaje que domina es la referencial, ya que el sujeto-emisor solo informa, no entra a valorar ni expresar sentimientos o emociones.

### Fase 3: Concordancias.

En esta sección vamos a ver las palabras en contexto para ver con más detalle qué tipo de léxico rodea a la palabra seleccionada y con qué función del lenguaje se asocian. Como ya hemos visto en el apartado 6.2.1 muchos de los fragmentos que ofrece el *output* de la herramienta *Concordance* coinciden con el enunciado de las tareas propuestas, de ahí que hayamos decidido realizar una selección de diez fragmentos del total de cada una de las muestras de cada grupo, NN y N.

En la tabla 108 presentamos algunos ejemplos del contexto en el cual podemos encontrar el posesivo singular de primera persona «mi» en las muestras de los grupos «NN» y «N».

Al observar dichos ejemplos, vemos que en la muestra del grupo NN los agrupamientos que hemos analizado anteriormente mantienen principalmente la función referencial del lenguaje limitándose a indicar qué tipo de relación se establece entre el posesivo y el sustantivo al que acompañan: (i) de parentesco («*mi amiga*», posición 414, 522, 601-«NN»; «*mi novio*», posición 137-«NN»); (ii) o de posesión («*mi perro*», posición 100, 157-«NN»). No obstante, aparece una concordancia que remite a la función expresiva al expresar un sentimiento ante lo sucedido: «*Entonces el día de mi cumpleaños fue aburrido*», posición 187-«NN». En cambio, en la muestra del grupo N, encontramos tres concordancias en las que el contexto influye en el agrupamiento para hacer que cumpla con la función expresiva del lenguaje: «*lo recuerdo como el peor primer día de clase de mi vida*», posición 38-«N»; «*durante el descanso me relajé y abandoné más mi timidez*», posición 116-«N»; «*No obstante, una mañana soleada me desperté en mi casa. Era lunes, y el sol brillaba con intensidad*», posición 259-«N». Como vemos, los agrupamientos por sí solos («mi vida» o «en mi casa») no remiten a una función expresiva del lenguaje, pero el contexto nos indica que la intención del sujeto-emisor no es solo informar, sino mostrar su actitud ante lo que se está explicando.

Grupo «NN»	
mi	
92	algun ha abrido la luz y en el salon estaban <b>mi toda la familia</b> . Toda gente me dice: “cumpleaños
100	Ellos se acercaron con caras irritadas. El <b>mi perro</b> escapó muy rapido e los gatos de tras de
114	migo viejo. Juan. Él estaba cruzando una plaza de <b>mi zona</b> . “¡Qué casualidad! Juan, ¿Qué tal?
137	peluquero y gastaba mucho, estaba segura que a <b>mi novio</b> la noticia de estos gastos no le gustaria
157	desspués el mi perro se chocar por el coche, y llevo <b>mi perro</b> con Canelo a hospital. Porque cuando la
187	deportes en lugar de mirando. Entonces el dia de <b>mi cumpleaños fue aburrido</b> y solidad por que
382	de Nepal, cuando hicimos la matricula y por eso <b>mi primer día</b> me sentí a lado del Santosh.
414	conocí alguien la gente. Cuanto empezamos yo con <b>mi amiga</b> sientamamos juntas. La profesora explico
522	as cosas de la tienda. final, compré una collar y <b>mi amiga</b> compró un cuadro chino. Fuimos un día
601	de mi amiga, estaba corrido muy pronto a puerta. <b>Mi amiga</b> no sabia qué hacer, en este momento,

Grupo «N»	
mi	
38	<b>lo recuerdo como el peor primer día de clase de mi vida</b> , aunque, mirándolo desde un punto de vista
62	clase fue hace años, cuando yo iba a parvulario. <b>Mi profesora</b> , la señorita Anabel, dedicó una de las
81	todo momento. Así, cuando fueron pasando los días <b>mi</b> confianza aumentó e incluso me decidí a
107	mera clase de catalán porque ya no me recuerdo de <b>mi</b> primera clase de francés o castellano. Estaba
116	pero durante el descanso me relajé y <b>abandoné más mi timidez</b> . Cuando volvimos a entrar, la profesora
187	Recordando los cursos de primeros auxilios que <b>mi abuelo</b> me insistió en asistir, llamé a una
223	que se moviera. Él puso su patita encima de <b>mi</b> pierna y algo vi en su mirada que me hizo dudar
259	No obstante, <b>una mañana soleada me desperté en mi casa</b> . Era lunes, y <b>el sol brillaba con intensidad</b>
308	No veía a sus padres por mucho que mirase a <b>mi</b> alrededor. Le dije que mejor no acariciase al
356	era como si no pudiera estar quieto. De hecho, ni <b>mi familia</b> ni yo notamos nada hasta que lo saqué a

Tabla 108. Output de la herramienta *Concordance* para el adjetivo «mi» NN - N

En la tabla 109 sobre las concordancias del pronombre complemento «me», vemos que tanto en la muestra del grupo NN como en la del grupo N, aparecen cinco contextos que claramente se ajustan a la función expresiva del lenguaje al expresar sentimientos y emociones (posiciones 33, 44, 55, 68, 101-«NN»; 69, 99, 128, 192, 391-«N»). En todos estos ejemplos, el sujeto-emisor muestra su actitud ante lo que explica y lo hace mediante un léxico valorativo que denota subjetividad.

Hay que resaltar también las dos concordancias que encontramos en la muestra del grupo NN a partir del agrupamiento del pronombre complemento con verbos que introducen el discurso indirecto, *decir* y *explicar*. Si como hemos comprobado esta asociación predominaba en la sección de agrupamientos, aquí en concordancias solo encontramos dos casos de los ejemplos seleccionados del *output* total en las posiciones 129 y 148. El resto de concordancias se limitan a describir, por tanto, cumplen con la función referencial del lenguaje.

Grupo «NN»	
me	
10	ase fue Divertido. Fue el día 11 de febrero, no me acuerdo. En el primer día, primacia, explicamos
33	La profesora nos enseñó la gramática del pasado. Me sentí la clase bien.
44	en la aula. Me sentí tranquilo pero también me sentí un poco nervioso porque esta clase fue
55	el, es mi nivel, la profesora es muy simpática, y me gusta mucho como da la clase. También he
68	pero ella era un poco mayor que yo pero la verdad me da igual de la edad si seguimos siendo amigas
101	negros en las caras. Y las pistolas en las manos. Me ponía muy nervioso no sabía que tenía que hacer
129	inglés, porque Maya no habla español. También me dijo que aquí son la gente muy simpáticos
148	que cada persona llevaba comida, bebida y regalo, me explicó que quieren darme una sorpresa.
204	ron comer conmigo en mi casa. De la tarde, cuando me quedaba en casa ver la tele tranquilamente,
441	su bolsa. En ese momento me dolían los pies y me senté en un banco. En el banco había una bolsa,

Grupo «N»	
me	
41	Al no saber dónde se encontraba la escuela, me acompañaron en coche pero me llevaron al lugar
69	francés o intentaba hacerme entender con gestos. Me sentí muy apoyada por mi profesora y los
99	el profesorado y los compañeros eran fabulosos, y me ayudaron en todo momento. Así, cuando fueron
128	catalán todo el tiempo así que no entendía mucho y me daba vergüenza a hablar. Aquel día, aprendimos
159	o entender que se decía junto con la ayuda que se me proporcionó entonces comprendí que no todo sería
192	se y ese día es nerviosismo. No conocía a nadie y me sentía muy sola, la verdad. No me hacía mucha
234	medio del camino hacia el pueblo, y lo que menos me esperaba, es que quien conducía era mi prima,
280	era presente en la casa. Al cabo de un rato, clex me abrió la puerta y se sorprendió de verme allí.
304	s pantalones mojados y los cambié por unos secos. Me quité la ropa empapada y por fin entré en calor
391	con Dani Mateo, un famoso humorista catalán. A mí me gustaba mucho por aquel entonces ya que tenía

Tabla 109. Output de la herramienta *Concordance* para el pronombre «me» NN - N

Grupo «NN»	
yo	
1	que necesitamos para aprender español. Esta día yo conocí con los compañeros de me clase. Después de
27	mi español porque mucho gente hablan mejor que yo. La profesora nos explico cosas que vamos a hacer
40	Fue fin de verano, pero fue calor también. Cuando yo entré mi aula 36, he visto una señora muy bonita
52	Y aquel perro fue el pastor alemán. Después, yo y Canelo damos una vuelta por el parque que se en
64	todo los sitios pero no encontré ninguna cruceta. Yo iba a mi vesino Joan (el también tien un perro) y
112	de septiembre de 2009. Fue un día muy divertido. Yo llegué a las 10:20 y la clase había empezado
127	sabía que esta chica es de la misma ciudad como yo, pero cuando la profesora preguntó de donde
141	Yo abrí mi móvil pero tampoco funcionaba. Pues yo lo utilicé a la luz para buscar a mi perro.
174	para Medicina, Susanne estaba estudiando moda, y yo estudiando derecho y Paty no estaba estudiando
245	Barcelona tardaron 2 horas por el fuerte viento. Yo le esperé una hora más y llegó a las cinco de la

Grupo «N»	
yo	
4	e miraban mucho, con cara de no saber qué pintaba yo allí. Pasado un rato, me preguntaron, al fin,
15	este idioma en una prestigiosa universidad. Cuando yo y Raquel, mi amiga, entramos a la clase, nos
28	Todos mis compañeros me miraban con curiosidad. Yo tenía mucho miedo ya que no sabía hablar Alemán,
37	e preguntaron si yo era la chica nueva de España, yo les dije que sí y les pedí que me llevaran al
45	primer día de clase fue un tanto diferente a lo que yo estaba acostumbrado. Antes vivía en un pueblo
51	entendía por qué me miraba de aquella manera, pero yo tampoco podía dejar de contemplarla. Fue entonces
70	de repente mi perro vio a un gato y le persiguió. Yo le llamaba muchas veces pero no me hizo caso y
86	truenos se escondieron en la cabaña. Canelo y yo estuvimos sentados esperando que se nos secara
98	los domingos es el único momento en que Laura o yo podemos llevarle y él lo sabe. Estábamos a unos
105	pero me fue imposible, él era más rápido que yo. Me volví loca mientras buscaba a Canelo y no

Tabla 110. Output de la herramienta *Concordance* para el pronombre «yo» NN - N

En la tabla superior, tabla 110, presentamos el contexto en el que encontramos el pronombre sujeto «yo». Como ya hemos señalado con anterioridad, la presencia del pronombre sujeto en el discurso es para evitar ambigüedades o resaltar el hecho de que se trata de ese sujeto y no de otro. Así, el *output* de la herramienta *Concordance* viene a ratificar

lo visto que en el apartado de *agrupamientos*. Lo que interesa comprobar aquí, en *concordancias*, es identificar qué función del lenguaje predomina en los contextos en los que aparece el pronombre sujeto «yo», y, como vemos en la tabla, es la función referencial del lenguaje la que domina. En la muestra del grupo NN no encontramos ninguna concordancia en la que aparezca la función expresiva o conativa; mientras que en la muestra del grupo N encontramos dos ejemplos de función expresiva: «*miraban mucho, con cara de no saber qué pintaba yo allí*», posición 4-«N»; «*Yo tenía mucho miedo*», posición 28-«N». La elección de un léxico valorativo hace que estas concordancias adquieran función expresiva.

A continuación presentamos las concordancias del sustantivo «casa».

Grupo «NN»	
casa	
5	ver los libros del alumno y otros para leer en casa. El primer día de clase de español en la EOI fue
17	re a la puerta y sorpresa, mi padres viniendo a mi casa con sonreir. Estaba muy contento y podría mira
23	a noche, por eso, el niño siempre quetaba solo en casa. A veces vino su abuela, pero no quedó mucho
32	de 8 paradas cammino por 20 minutos y llego a una casa muy grande y bonita. Abre la puerta la mujer que
38	s en un restaurante. En este día, cuado estaba en casa y emperé dos amigos, oí el timbre de la puerta.
62	siempre me ponía cara roja y cuando regresé a casa yo siempre estaba de malhumor. Cada dia
71	llaves de casa. Es que vivía sola no pude entrar a casa sin llaves, por eso tenía que volver al camino
87	Entonces nosotros fuimos caminando hasta nuestra casa. Ese día fue horrible para mi porque hace dos sema
99	estaba en la casa y dijo que enseguido venía de la casa al aeropuerto. Entonces yo estaba un poco
108	tenía que esperar a mi amiga y llevarla a mi casa. De pronto aparecieron unos doctores que me lleva

Grupo «N»	
casa	
5	a que vinieran a recogerme para llevarme a la casa que, durante las próximas tres semanas, sería mi
9	hecho que agradecí porque tenía ganas de llegar a casa. Pero antes de irme, la chica que estaba a mi lado
12	en la otra punta del pueblo. Cuando salí de mi casa, el cielo estaba despejado y empezaba a salir el
16	vení de una casa abandonada, mi perro ladró a la casa, como si hubiera algo extraño en ella, y yo, que
33	paseaba por la acera de enfrente. Nos marchamos a casa y comenté via Facebook lo que me había ocurrido.
42	No obstante, una mañana soleada me desperté en mi casa. Era lunes, y el sol brillaba con intensidad.
48	excusa y dejé a la parejita feliz, marchándome a casa enfurruñada con Canelo, que se acababa de
54	que procedían de la montaña que hay al lado de mi casa. Me entró tanta curiosidad, que decidí ir a ver
61	Empecé a imaginar que podía entrar alguien en casa y robarnos. La verdad es que tenía ganas de
70	nos encontramos un par de coches de policia en casa de los vecinos. Les habían estado robando y,

Tabla 111. *Output* de la herramienta *Concordance* para el sustantivo «casa» NN - N

En la tabla superior mostramos algunos ejemplos de las concordancias del sustantivo «casa» que ofrece la herramienta *Concordance*. Nuevamente el *output* de concordancias verifica lo ya comentado en *agrupamientos*. Así, aquí vamos a comprobar la función del lenguaje que predomina en las concordancias. Vemos que en la muestra del grupo NN hay tres ejemplos de concordancias que cumplen con la función expresiva del lenguaje (posiciones 17, 32 y 62-«NN»), mientras que en la del grupo N encontramos dos («*No obstante, una mañana soleada me desperté en mi casa. Era lunes, y el sol brillaba con*

*intensidad*», posición 42-«N»); «*dejé a la parejita feliz, marchándome a casa enfurruñada con Canelo*», posición 48-«N»).

En la tabla siguiente mostramos las concordancias del sustantivo «momento».

Grupo «NN»	
<i>momento</i>	
2	hay mucha gente estuvieron esperando. Esperé un momento, una profesora entradó. Es delgada, guapa
5	paró en la carretera por semáforo rojo. <b>En aquel momento</b> , de repente, salieron dos turismos blancos.
10	dejado de ellos, cuándo podrían verme. <b>En aquel momento</b> sí que quería dejar todo para volver a China
13	e no conocía nadie en m' clase pero <b>al cabo de un momento</b> me sentí mas comodo. Por el primer d'a de
18	l medico vigila todos, pero no podía curar. <b>Aquel momento</b> mi perro solo lloraba y hamirando a mi.
21	el vino a buscarnos y nos acompañó a casa. <b>En el momento</b> en lo que cruzamos las escaleras en frente
26	se acercó una persona. la vi y me quedé un momento inmóvil. Era María, una amiga de mi infancia
29	vez lo vi que él también estaba nervioso. <b>En ese momento</b> no estaba segura que el hombre estaba nervioso
35	porque no sabía que tendría que hacer <b>en ese momento</b> . Los niños eran muy pobres, y sobre todo,
40	Algunas personas lloraron. Había un momento que no vi nada a causa del nieve. Después cinco

Grupo «N»	
<i>momento</i>	
3	se despidió con la frase que <b>a partir de ese momento</b> usaría siempre: Goodbye, class! Y antes de
6	compañeros eran fabulosos, y me ayudaron <b>en todo momento</b> . Así, cuando fueron pasando los días mi
9	allí. Lo que más me aterraba era que si en algun momento determinado necesitaba preguntar una É o si tenia
11	formó por parte de los mismos estudiantes. <b>En el momento</b> de entrar en clase me dio un poco de vergüenza
13	la idea de empezar <b>un idioma desconocido hasta el momento</b> , llegábamos al aula a través de los estrechos
17	hacia mí con la energía de un cachorro. <b>En ese momento</b> supe que estaba recuperado.
25	me acercó y se presentó Derek, me dijo. <b>En ese momento</b> , mi perro, se abalanzó sobr él y empezó
28	valor, decidí ir a hablar con ella. <b>Justo en ese momento</b> , salió el sol y dejó de llover. Aprovechamos
31	Las mañanas de los domingos es el único momento en que Laura o yo podemos llevarle y él lo s
32	y, adivinad quién era el ladrón. <b>A partir de ese momento</b> Canelo volvió a estar igual de tranquilo y

Tabla 112. Output de la herramienta *Concordance* para el sustantivo «momento» NN - N

Volvemos a encontrar en la tabla de *concordancias* los agrupamientos «preposición + momento» para formar las locuciones adverbiales (i) «*en ese/aquel + momento*», posiciones 5, 10, 21, 29, 35-«NN»; 11, 17, 25, 28-«N», para ubicar en un punto concreto del espacio temporal aquello a lo que refiere el discurso. (ii) «*hasta + momento*», posiciones 13-«N»; para indicar el punto límite en el espacio temporal. (iii) «*a partir de ese + momento*», posiciones 3, 32-«N», para indicar el punto de origen en el espacio temporal. (iv) «*al cabo de + momento*», posiciones 13-«NN», para indicar el tiempo transcurrido. Comprobamos que en todos los contextos la función del lenguaje que predomina es la referencial, pues simplemente se informa de lo que sucede, sin entrar en valoraciones o expresión de sentimientos o emociones. Hay un uso muy reducido de léxico valorativo.

En cuanto a las concordancias con el sustantivo «hombre», vemos en la tabla 113 que destacan por un dominio de la función referencial del lenguaje. Se describe o se identifica mediante un léxico objetivo, no valorativo: («*Un hombre de chian y de 20 algo años.*», posición 1-«NN»; «*Seguí el hombre a lo largo de la calle.* », posición 20-«NN»; «*El chico creyó que el hombre iba a pedirle algo*», posición 5-«N»; «*El hombre huyó en calzoncillos*

*gritando despavorido*», posición 21-«N»; «*nunca más volvimos a ver al hombre del bigote negro sentado en ese banco*», posición 29-«N». Aun así, encontramos ejemplos también de la función expresiva del lenguaje en ambos grupos: «*Sentí nerviosa cuando el hombre me dijo que el monstruo era gigante*», posición 10-«NN».

Grupo «NN»	
<i>hombre</i>	
1	facil. Cuando llegué me encontró a Yang. Un <b>hombre de chian</b> y de 20 algo años. Nos hablamos
4	había un accidente de trafico con coche golpeó un <b>hombre mayor de edad quizás 60 años</b> . Todo el
10	y empezó a llorar. <b>Sentí nerviosa cuando el hombre me dijo que el monstruo era gigante</b> con ojos tan
14	tenía suerte que había escapado. ¡Le avisó al <b>hombre</b> que fuera una idea mala dormir en la basura y
20	Decidí a verlo con mis propios ojos. Seguí el <b>hombre</b> a lo largo de la calle. Cuando llegamos a la
23	que no duerma en la basura. De verdad, el <b>hombre</b> estaba sucio porque había dormido la siesta en
29	para acabar con los problemas y me acercé del <b>hombre</b> . Este no había tomado una ducha desde mucho
35	te. Ya no me dijo nada de nada. Estaba segura ese <b>hombre era loco</b> . Pensé que era mejor si me levante y
40	con el grupo. Tres minutos después entré el <b>hombre</b> que quería controlar los bijettes de los
45	corrió a la tienda. La persona quien robó era un <b>hombre</b> parecía 30 años, no era alto y llevaba una ropa

Grupo «N»	
<i>hombre</i>	
2	más fuerte. No entendía nada. Apareció otro <b>hombre joven</b> allí que me empujó al suelo y pateó a
5	ven de unos dieciséis años. El chico creyó que el <b>hombre</b> iba a pedirle algo, pero resultó que éste
8	ayuda. En ese momento me giré y vi como un <b>hombre de compleción fuerte</b> y vestido de oscuro
13	y a ladrar con fuerza. Los días pasaban, y aquel <b>hombre</b> que había alterado a Canelo se mostraba cordial
15	estrategias de los asesinatos cometidos. Aquel <b>hombre era un criminal</b> y yo podía demostrarlo. Cuando
17	n Canelo, mi perro, cuando de repente apareció un <b>hombre disfrazado de salchicha</b> . Estaba promocionando
21	lanzándolo al suelo y arrancándole el traje. El <b>hombre</b> huyó en calzoncillos gritando despavorido,
24	con Canelo, mi perro, cuando de repente un <b>hombre de mediana edad</b> pasó por nuestro lado a gran
27	ni Canelo (que se puso a ladrar) conocíamos. El <b>hombre</b> me empezó a preguntar por mi vida
29	perezoso de siempre, y nunca más volvimos a ver al <b>hombre del bigote negro</b> sentado en ese banco.

Tabla 113. *Output* de la herramienta *Concordance* para el sustantivo «hombre»

En la tabla 114 presentamos las concordancias del sustantivo «miedo». Como ya hemos señalado anteriormente, el significado semántico del sustantivo ya expresa por sí mismo un sentimiento; así, los contextos en los que va a aparecer se ajustan a la función expresiva del lenguaje. Las siguientes concordancias dan muestra de ello: «*tiene noticias muy malas. Tenía miedo de escucharlas, pero en el corazon sabía*», posición 10-«NN»; «*No podía estar más asustada, me moría de miedo*», posición 17-«N»; «*Pasé un miedo terrible*», posición 19-«N». Vemos que estos ejemplos se centran en la primera persona, sujeto-emisor; sin embargo, en la muestra del grupo NN encontramos ejemplos en los que la expresión de sentimiento recae en otras personas: «*el hombre estaba gritando y saltando con miedo*», posición 16-«NN»; «*toda gente estaba mirando con ojos grandes y con miedo de el*», posición 29-«NN».

Grupo «NN»	
<i>miedo</i>	
2	pero cuando he llamado no me dice nada. Tuve más miedo. Y cuando he ido a salon, alguien ha abierto la
5	a mi perro y mi pobrecito Canelo tenía mucho miedo de ese gato gigante, feo y sucio. Sinceramente,
10	de su casa y tiene noticias muy malas. Tenía miedo de escucharlas, pero en el corazon sabía que
13	é a gritar para distraer el monstruo. Tenía mucho miedo porque no había visto nadie con el perro
16	o aunque el hombre estaba gritando y saltando con miedo. ¡Fue sólo una camioneta de basura! El hombre
23	era grave pero estaba sangrando. Las dos teníamos miedo de la enfermedad que los animales salvajes
25	sobre nosotras. Respondimos automaticamente, con miedo. Queríamos escapar, pero no supemos donde
29	toda gente estaba mirando con ojos grandes y con miedo de él. Ademas, mas venia mas gente de policia
33	con mucho gritos. Las pasgeros tenían mucho miedo, empezaron a llamar a la policia y a sus
37	legar al aeropuerto. Vilo en el aire. Tenía mucho miedo, porque no podría ayudar. La gente que estaba

Grupo «N»	
<i>miedo</i>	
2	la, dejándonos medio atemorizados. Para vencer el miedo al alemán nos hizo leer. Yo nunca había leído
4	pausa de los nervios, de la emoción, y también del miedo. Al no saber dónde se encontraba la escuela,
7	compañeras por el interés por aprender y por el miedo que supone empezar una lengua desde cero.
10	el curso yo estaba un poco nerviosa porque tenía miedo de no saber desenvolverme bien con el alemán.
14	verdad. No me hacía mucha ilusión empezar, tenía miedo. Antes cuando era más pequeña, no me gustaba
17	a pared. No podía estar más asustada, me moría de miedo. Pero de repente me giré y allí estaba él, mi
19	la señora sufrió un ataque de ansiedad. Pasé un miedo terrible. Cuando lo pensé con frialdad,
22	había algo en él que me transmitía desconfianza y miedo. Un día, decidida a investigar sobre él,
26	chaqueta que abrigara más o que simplemente tenía miedo. No me considero una persona miedosa, pero el
29	aumentaba a cada minuto que pasaba, y tenía miedo de pensar que le podría haber pasado cualquier

Tabla 114. Output de la herramienta *Concordance* para el sustantivo «miedo» NN - N

En la tabla siguiente mostramos las concordancias del sustantivo «tiempo».

Grupo «NN»	
<i>tiempo</i>	
3	novamente al nivel tercero i ahora tengo mas tiempo i tambien me dedico un poco más.
5	de crutura y como la vida. El primer día pasé el tiempo muy rapido, y fue muy divertido.
7	yo soy un poco tímida y sentí nerviosa pero con tiempo algunos aluminos han cambiado a mis amigos.
11	encontrado una amiga, que no he visto hace mucho tiempo. Se llama Alexca. Nos conocemos des de
15	casa. A veces vino su abuela, pero no quedó mucho tiempo con él, y se marchó. Yo ví el niño, jugaba la
22	porque no me recordólo. ¡Que fatal! Es que no tuve tiempo para recordarlo, hice muchas cosas para ganar
31	el autobus y el metro fue al EOI. Contai mucho tiempo. En primera clase, las personas explica las
35	experiencia, sus ocupaciones y a qué dedica su tiempo libre. Despues pasamos a estudiar los adjetivos
39	Despues nos fuimos a un bar. Estabamos allí mucho tiempo. Estabamos hablando mucho y nuestros perros
56	libros para estudiar. Ese momento hacia buen tiempo, habia mucha gente, estaba contenta.

Grupo «N»	
<i>tiempo</i>	
2	utas para que formulásemos preguntas que al mismo tiempo nos permitieran saber cuales eran nuestros
4	e la clase, yo deseaba que se acabara ya, pero el tiempo pasaba muy despacio. Tuvimos que hacer unos
7	Mi primer d'a de clase fue hace bastante tiempo, así que hablaré de mi primer día de clase de
8	de Ciutadella, y como no lo conozco, fui con tiempo para poder encontrar la clase. Después de un rato
10	porque me gusta aprender idiomas. Pero al mismo tiempo estaba muy nerviosa. En la clase eramos
11	y, sinceramente, no fue tan malo como creía. El tiempo pasó rápido, hecho que agradecí porque tenía
15	mi novio. Me enfadé ya que me hizo pasar un mal tiempo, pero al final acabé perdonándole porque
17	tormentas. En unos minutos empezó a llover. El tiempo está loco- pensé- Hace unas horas lucía un
21	donde voy a pasear con Canelo, pensé que hacía tiempo mi madre me contó que María tenía un
24	me y no me lo permitía. De pronto, en un ápice de tiempo en el que conseguir sacar la cabeza del agua,

Tabla 115. Output de la herramienta *Concordance* para el sustantivo «tiempo» NN - N

Vemos en la tabla de *concordancias* que los agrupamientos que hemos señalado en el apartado anterior para cuantificar y que cumplían con la función referencial del lenguaje, aquí también cumplen con esta función al limitarse a informar de lo que sucede. Ejemplo de

ello lo encontramos en: «*encontrado una amiga, que no he visto hace mucho tiempo*», posición 11-«NN»; «*Despues nos fuimos a un bar. Estabamos allí mucho tiempo*», posición 39-«NN»; «*Mi primer día de clase fue hace bastante tiempo*», posición 7-«N».

Si pasamos a ver qué concordancias cumplen con la función expresiva del lenguaje, comprobamos que en la muestra del grupo NN solo hay una («*yo soy un poco timida y sentí nerviosa pero con tiempo algunos aluminos han cambiado a mis amigos*», posición 7-«NN»). En esta concordancia, el sujeto-emisor no solo da a conocer cómo es, su personalidad, sino que también expresa su estado ante una situación determinada. En cambio, en la muestra del grupo N encontramos cuatro ejemplos de concordancias que cumplen con la expresión expresiva (posiciones 10, 11, 15 y 17-«N»). La de la posición 10, por ejemplo, presenta sentimientos opuestos, por un lado, un sentimiento positivo, pero por el otro lado, negativo porque «se siente nerviosa».

### **Resumen de la pregunta 2:**

Para responder a la segunda pregunta de investigación, hemos recurrido, igual que para la pregunta uno, a la aplicación informática AntConc y a sus herramientas *WordList* (frecuencias), *Cluster* (agrupamientos) y *Concordance* (palabras en contexto). Con ellas, hemos analizado los dos tipos de producciones escritas correspondientes a las muestras de los grupos «NN» y «N».

En el análisis de frecuencias, hemos visto que las palabras más frecuentes en las muestras de los dos grupos, dentro de las diez primeras posiciones, son las palabras funcionales, pronombres, preposiciones, artículos y conjunciones. De ellas, siete aparecen en los dos listados de frecuencias (*a, de, y, la, que, el, en*). Al comparar las dos muestras del corpus de nuestra investigación con el corpus CREA, hemos comprobado que mientras en nuestro corpus las dos primeras posiciones están ocupadas por las preposiciones «a» y «de», respectivamente, en el corpus CREA es la preposición «de» la que ocupa la primera posición y la preposición «a» no aparece hasta la posición siete.

También hemos comprobado que en el corpus CREA, la primera forma verbal aparece en la posición treinta y nueve, como ya indicamos en la pregunta uno. En relación a los sustantivos, el corpus CREA solo registra uno en las primeras cincuenta posiciones y aparece en la cuarenta y siete. Se trata del sustantivo «años». En cambio, el grupo NN y N presentan

un total de seis y cinco sustantivos, respectivamente. Pero se debe indicar que salvo uno, «casa», el resto corresponde al enunciado de la tarea.

Para evitar que en las primeras posiciones del listado de frecuencias apareciera ese tipo de léxico, elaboramos un listado de palabras que no deseábamos cuantificar (*stopwords*), como las palabras que formaban parte del enunciado de las tareas y algunas palabras funcionales. Tras volver a aplicar la herramienta *WordList*, y al contrario de lo sucedido en la muestra «1D» de la pregunta uno, aquí no ha sido posible encontrar un hilo conductor a partir de las palabras léxicas. Esto es así por analizar los dos tipos de textos, «F» y «1D», juntos, siguiendo el criterio de grupo y no de género textual.

Al comprobar si había similitud entre los dos grupos, se identificaron una serie de palabras que coincidían en ambos (*mi, me, yo, casa, hombre, momento, miedo, tiempo*). A partir de ellas se realizaron los análisis de agrupamientos y concordancias.

Así, el análisis de agrupamientos y de concordancias permitió verificar que hay bastantes similitudes entre ambos grupos, aunque sí encontramos algunos casos de agrupamientos y concordancias en el grupo N que se alejaban del grupo NN, como un léxico más variado, por ejemplo, para expresar sentimientos (*gustar, encantar, asustar, etc.*) o para cuantificar (*mucho, bastante, algo, etc.*). También observamos un ligero predominio de la función expresiva en la muestra del grupo N frente a la del grupo NN.

### ***6.2.2 Resultados y análisis de la comparación de los dos momentos de recogida, M1 y M2, del grupo de NN***

En este apartado vamos a centrar nuestro análisis en el grupo de NN exclusivamente y vamos a dar respuesta a las preguntas tres y cuatro de la investigación.

En la pregunta tres, vamos a tomar en conjunto todas las producciones escritas (muestras «F» y «1D») que realizaron los sujetos en los dos momentos de recogida de datos, al inicio del curso (M1) y al finalizar el mismo (M2), para comprobar si hay mejoría en sus producciones escritas en fluidez, diversidad léxica y densidad léxica tras una instrucción de 120 horas siguiendo el enfoque comunicativo.

En la pregunta cuatro seguimos analizando y comparando los M1 y M2, pero centrándonos en las muestras «F» y «1D» por separado, para comprobar si hay diferencias

en fluidez, diversidad léxica y densidad léxica entre los dos tipos de texto que conforman cada una de las muestras.

**Pregunta 3: Tras un periodo de instrucción de 120 horas y siguiendo una enseñanza mediante tareas ¿se observa progreso en cuanto a la fluidez, diversidad y densidad léxica en las producciones escritas de los estudiantes no nativos entre el M1 (inicio del curso, nivel B1) y el M2 (final del curso, nivel B1)?**

Para poder responder a esta tercera pregunta de investigación, vamos a realizar un análisis general de los resultados que ya hemos presentado en el apartado anterior de «estadística descriptiva».

Este análisis lo vamos a llevar a cabo en dos fases:

(i) En primer lugar vamos a comparar qué grado de fluidez y diversidad léxica hay en cada una de las muestras del grupo de no nativos, es decir, entre el M1 (inicio del curso, nivel B1) y el M2 (final del curso, nivel B1). Para verificar si se produce aumento en la fluidez entre la muestra del M1 y del M2, vamos a tener en cuenta el número total de palabras (*tokens*) escritas en cada una de las muestras. En cuanto a la diversidad léxica, vamos a tener en consideración la ratio entre palabras totales y palabras totales diferentes presentes en cada una de las muestras (Wolfe-Quintero et *al.*, 1998; Jarvis, 2003).

(ii) A continuación vamos a comprobar la densidad léxica presente en cada una de las muestras y compararla también entre la muestra del M1 y del M2. Para realizar este análisis partimos de la media de PL, palabras léxicas (verbos, sustantivos, adjetivos y adverbios) frente a la media de PF, palabras funcionales (pronombres, preposiciones, artículos, conjunciones) presente en las muestras del M1 y del M2. Se considera que un texto es más «denso» cuando la proporción de PL es superior a las PF. Eso es así porque las palabras léxicas tienen la función de suministrar información al texto, mientras que las palabras funcionales se caracterizan por su uso exclusivamente gramatical (López-Mezquita, 2007; Reyes Díaz, 2010).

**Fase 1:** Fluidez y diversidad léxica de las muestras del M1 y del M2.

Como vimos en el apartado de «estadística descriptiva» en la tabla 41 se producía un aumento en todos los valores absolutos de las variables seleccionadas.

A partir de esos totales, los diferentes descriptores de la «estadística descriptiva» de cada una de las variables nos mostraron de qué manera las dos muestras estaban próximas o se alejaban la una de la otra. Uno de esos descriptores, la media para *tokens* (tabla 42) o para *types* (tabla 43), también nos mostró que la muestra del M2 presentaba valores superiores a los del M1.

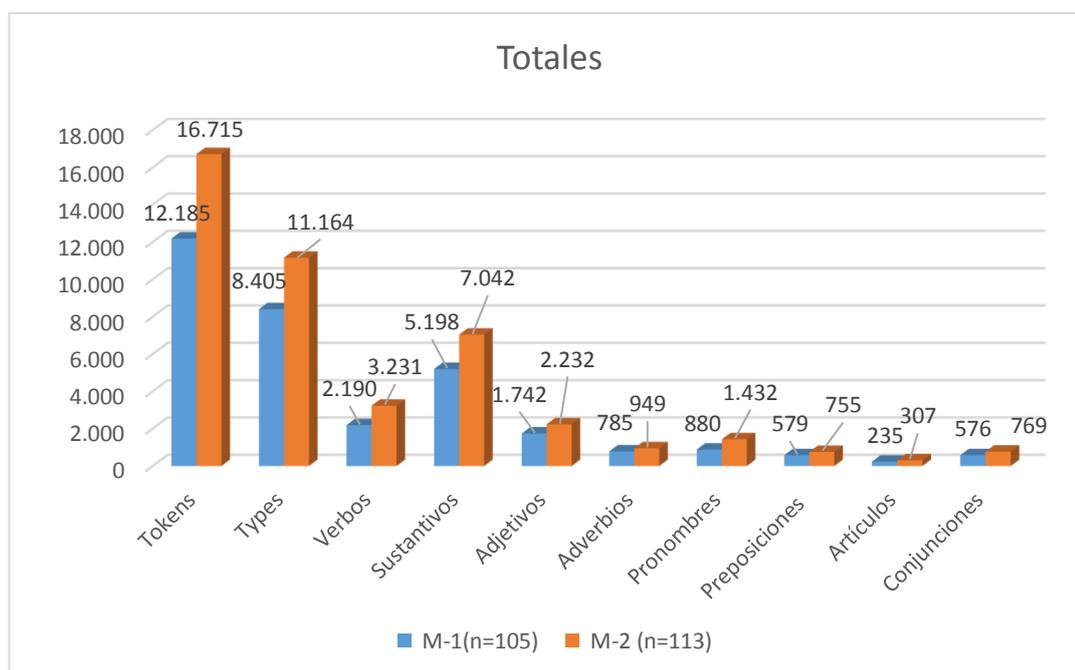


Gráfico 160. Totales del M1 y del M2

Como en esta parte de la investigación nuestro primer objetivo es comprobar si en la muestra del M2 se produce un aumento en la fluidez léxica de las producciones escritas, vamos a centrarnos en los valores de *tokens* que son los que proporcionan dicha información. Así, en el gráfico siguiente podemos ver más claramente las diferencias entre la media de *Tokens* de cada muestra según los valores obtenidos en la estadística descriptiva del apartado 6.1.3.

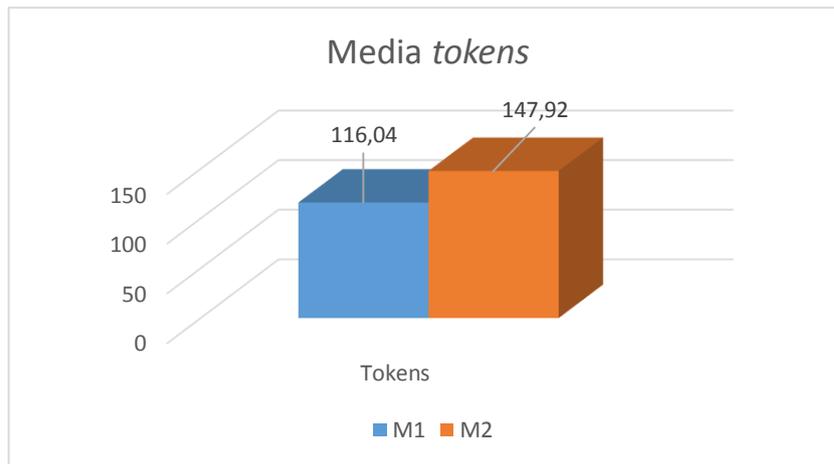


Gráfico 161. Media de *Tokens* y *Types*. M1 M2

Vemos que hay una diferencia de casi treinta y dos puntos en la media de *tokens* entre la muestra del M1 y la muestra del M2. Pero para comprobar si esa diferencia es significativa estadísticamente, recurrimos a pruebas estadísticas que explicamos a continuación.

En primer lugar comprobamos la bondad de ajuste a la normalidad de la media de la variable *tokens* de ambas muestras para decidir qué prueba estadística es la más apropiada para llevar a cabo el tratamiento estadístico. Como vemos en la tabla 116, en el M1 debemos rechazar la  $H_0$  de ajuste a la normalidad pues el p-valor es inferior al valor de significación de 0.05, mientras que en el M2 no podemos rechazar la  $H_0$  de ajuste a la normalidad, ya que el p-valor es superior al valor de significación de 0.05.

Cálculo de la bondad de ajuste a la normalidad de la variable <i>tokens</i>	
M-1	M-2
p-valor = <b>0.0002E-2</b> <0.05	p-valor = 0.0877 >0.05

Tabla 116. Prueba de Shapiro-Wilk de *tokens* para el M1 y el M2

Como una de las muestras, la del M1, no sigue la bondad de ajuste a la normalidad, la prueba estadística que vamos a utilizar para comprobar si hay o no diferencia significativa estadísticamente es el test U de Mann-Whitney (Wilcoxon), una prueba no paramétrica.

#### Estadístico W de Wilcoxon

-----  
**Hipótesis Nula:** igualdad de distribuciones  
**Hipótesis Alternativa:** distribución M1 no igual distribución M2

**W1 = 12729.5000; E[W1] = 11497.5000; V[W1] = 216490.3511**  
**W2 = 11141.5000; E[W2] = 12373.5000; V[W2] = 216490.3511**

**Estadístico de contraste de W: 2.6478**  
**p-valor de W: 0.0081**

Figura 26. Test de Mann-Whitney (Wilcoxon)

Comprobamos en la figura 26 que debemos rechazar la  $H_0$  de igualdad, pues el resultado del p-valor es inferior al valor de significación de 0.05 ( $p\text{-valor}=0.0081 < 0.05$ ). Por tanto, confirmamos que la diferencia existente en el descriptor «media» de la variable *tokens* es significativa estadísticamente. Así, podemos afirmar que se produce un aumento significativo en la fluidez léxica en la muestra del M2. Es decir, tras una instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, los sujetos del M2 son capaces de escribir textos más largos en las mismas condiciones que en el M1, dentro del aula y con un tiempo máximo de realización de treinta minutos.

El segundo objetivo de esta fase es averiguar si también ha aumentado la diversidad léxica de las producciones de la muestra del M2 respecto al M1. La forma habitual de realizar este cálculo es con la ratio de *Types-Tokens* (RTT), es decir, dividir el número total de palabras diferentes entre el número total de palabras. Pero si el tamaño de los textos a analizar es muy diferente, los resultados de la RTT pueden verse afectados (Wolfe-Quintero et al., 1998; Vermeer, 2000; Jarvis, 2002). Ya que los textos de nuestra investigación se caracterizan, por un lado, por ser pequeños, ya que no superan las 250 palabras, y, por otro lado, por presentar gran diferencia de tamaño, pues oscilan entre las 50 y las 250 palabras; se aconseja utilizar la medida de diversidad léxica Uber U (Dugast, 1978, 1979, 1980; Vermeer, 2000; Jarvis, 2002). La fórmula para calcular el índice Uber U es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Recordemos cuáles son los totales de *Tokens* y *Types* de los cuales partimos.

Grupo NN Muestras «M1» y «M-2»		
	M-1(n=105)	M-2 (n=113)
<i>Tokens</i>	12.185	16.715
<i>Types</i>	8.405	11.164

Tabla 117. Totales de *Tokens* y *Types* M-1 /M-2

Así, el índice Uber U que obtenemos a partir de esos totales son  $U = 0.16$  para la muestra del M1 y  $U = 0.18$  para la muestra del M2. Para indagar si esa pequeña diferencia que se aprecia es estadísticamente significativa, seguimos los mismos pasos que utilizamos para calcular la significación en la variable anterior. En primer lugar comprobamos la bondad de ajuste a la normalidad del índice Uber U para decidir qué prueba estadística es la más apropiada. Los resultados de la prueba de Shapiro-Wilk indican que no se puede rechazar la  $H_0$ , bondad de ajuste a la normalidad, ya que los valores son superiores al valor de significación 0.05.

Cálculo de la bondad de ajuste a la normalidad del índice Uber U	
M-1	M-2
p-valor Shapiro-Wilk: 0.1547 > 0.05	p-valor Shapiro-Wilk: 0.1557 > 0.05

Tabla 118. Prueba de Shapiro-Wilk del índice Uber U para el M1 y el M2

Al haber ajuste a la normalidad, el siguiente paso es comprobar la igualdad de varianzas para decidir si recurrimos a la prueba estadística t-Student de varianzas iguales o diferentes. Para realizar dicha comprobación, el contraste de varianzas, aplicamos el test F-Snedecor. El resultado que obtenemos, como observamos en la figura siguiente, no nos permite rechazar la  $H_0$  de igualdad de varianzas pues el valor ( $p\text{-valor} = 0.06811 > 0.05$ ) es superior al valor de significación de 0.05.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:       no igual
Estadístico de contraste F:   0.9233
p-valor:                       0.6811
    
```

Figura 27. Prueba F-Snedecor

Así, el siguiente paso es aplicar la prueba t-Student de varianzas iguales para averiguar si la diferencia observada en el índice Uber U es significativa estadísticamente.

<b>t-Student</b>	
-----	
<b>Hipótesis Nula:</b>	<b>diferencia de medias = 0.0000</b>
<b>Hipótesis Alternativa:</b>	<b>no igual</b>
<b>t-Student:</b>	<b>-2.1978</b>
<b>p-valor:</b>	<b>0.0290</b>

Figura 28. Prueba t-Student de varianzas iguales

Vemos en el resultado de la figura 28 que el p-valor es inferior al valor de significación 0.05 ( $p\text{-valor}=0.0290 < 0.05$ ). Eso significa que debemos rechazar la  $H_0$  de igualdad y, por tanto, podemos afirmar que, al igual que con la fluidez léxica, se produce un aumento significativo estadísticamente de la diversidad léxica en la muestra del M2.

Es decir, que tras una instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, los sujetos del M2 son capaces no solo de escribir textos más largos, más fluidos, sino también con mayor diversidad léxica.

## **Fase 2:** Densidad léxica de las muestras del M1 y del M2.

Como hemos indicado más arriba, en esta fase vamos a comprobar qué grado de densidad léxica presenta cada una de las muestras del grupo de no nativos. Para calcularla, vamos a llevar a cabo el tratamiento estadístico a partir de las medias de palabras léxicas y funcionales de ambas muestras (Pallotti, 2015).

A continuación mostramos el gráfico 162 con la media tanto de las palabras léxicas como de las palabras funcionales por separado y observamos que en todas las variables se produce un aumento en el valor de la media en el M2 respecto al M1.

Como para nuestro análisis de la densidad léxica necesitamos comparar la media del conjunto de palabras léxicas y funcionales, pasamos a mostrar en la tabla siguiente los resultados de totales y de media de todas las PL y PF de las muestras del M1 y del M2 (Tabla 119).

Comprobamos en la tabla 119 que los valores totales del M2 son superiores a los del M1. Respecto a los valores de la media, llama la atención la proximidad de los valores de ambas muestras. Podemos ver que el valor de la media de PL es algo más alto en la M1, mientras que en la PF es a la inversa, es algo mayor el valor de la M2.

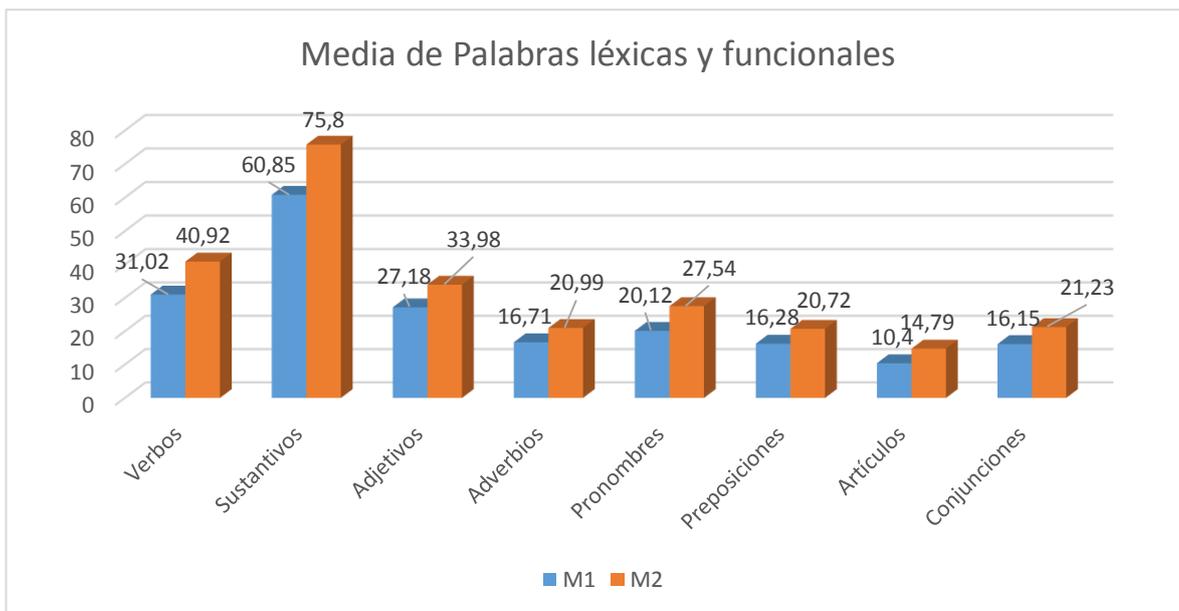


Gráfico 162. Media de palabras léxicas y palabras funcionales

TOTALES			MEDIA		
	M1	M2		M1	M2
<b>PL</b>	9.915	13.454	<b>PL</b>	0.8137	0.8049
<b>PF</b>	2.270	3.261	<b>PF</b>	<b>0.1863</b>	<b>0.1951</b>

Tabla 119. Totales y media de palabras léxicas y palabras funcionales

En los gráficos siguientes se pueden visualizar esas pequeñas diferencias.

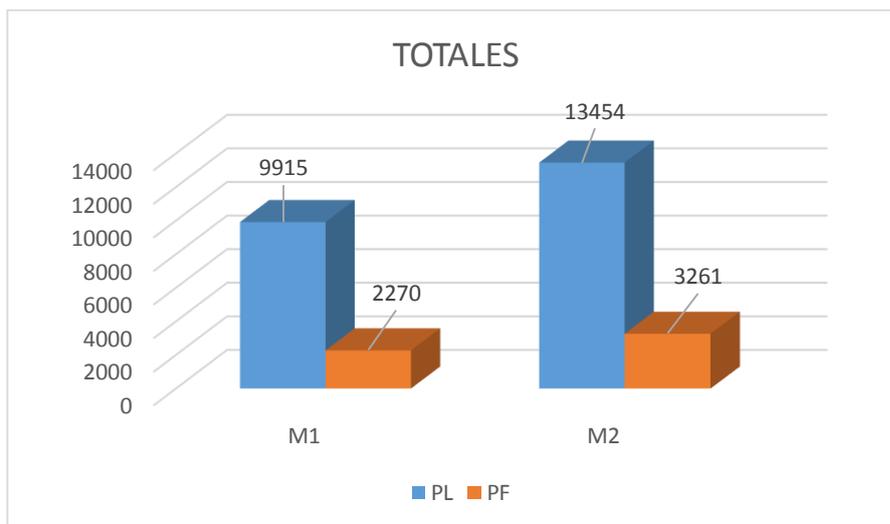


Gráfico 163. Totales de palabras léxicas y palabras funcionales

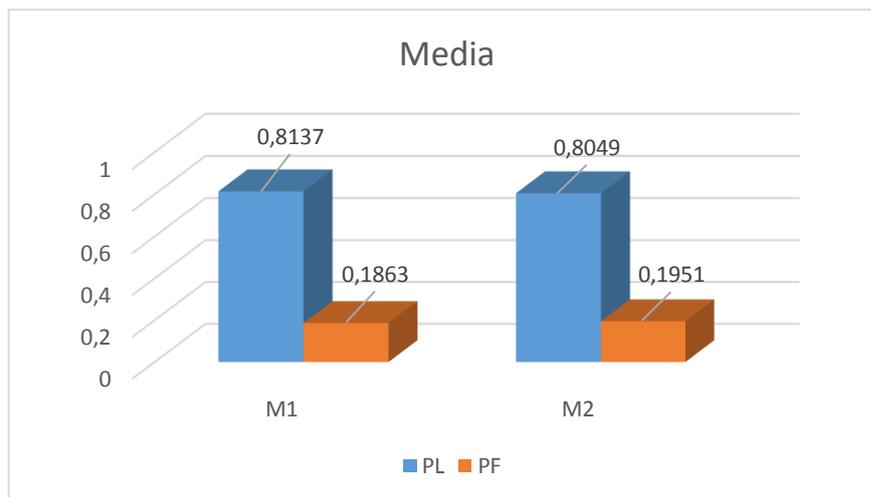


Gráfico 164. Media de palabras léxicas y palabras funcionales

Como verificamos en los gráficos, (i) hay una diferencia importante entre la media de palabras léxicas y funcionales dentro de una misma muestra, es decir, la media más alta corresponde a las PL tanto en la muestra del M1 (PL=9915) como del M2 (PL=13454); (ii) sin embargo, observamos que la diferencia de los valores entre la muestra del M1 y del M2 son mínimos. Para confirmar si esas diferencias son significativas estadísticamente, recurrimos a pruebas estadísticas que explicamos a continuación.

En primer lugar comprobamos la bondad de ajuste a la normalidad de la media de cada una de las variables de ambas muestras para decidir qué prueba estadística es la más apropiada para realizar el tratamiento estadístico.

Cálculo de la bondad de ajuste a la normalidad		
	M1	M2
<b>PL</b>	p-valor = 0.0002E-13 < 0.05	p-valor = 0.0002E-12 < 0.05
<b>PF</b>	p-valor = <b>0.0002E-11 &lt; 0.05</b>	p-valor = <b>0.0003E-13 &lt; 0.05</b>

Tabla 120. Prueba de Shapiro-Wilk para PL y PF en el M1 y el M2

Como podemos apreciar en la tabla, debemos rechazar la  $H_0$  de ajuste a la normalidad, ya que los valores del p-valor son inferiores al valor de significación de 0.05. Por tanto, para realizar el tratamiento estadístico de las muestras, hemos de recurrir a la prueba estadística no paramétrica U de Mann-Whitney (Wilcoxon).

Una vez decidido el test estadístico, (i) primero vamos a comprobar si la diferencia que hay entre palabras léxicas y funcionales dentro de cada una de las muestras del M1 y del M2 es significativa estadísticamente; (ii) y a continuación, analizaremos si la pequeña diferencia que se aprecia en las PL entre el M1 y el M2 es o no estadísticamente significativa.

Tras realizar el tratamiento estadístico para verificar si la diferencia entre la media de PL (M=0,8137) y PF (M=0,1863) en la muestra del M1 es significativa estadísticamente, debemos rechazar la  $H_0$  de igualdad, ya que el p-valor es inferior al valor de significación de 0.05 (M1: p-valor = 0.0000 < 0.05), como podemos observar en la siguiente figura.

```

Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa: distribución PL no igual distribución PF

Uxy = 19576.0000; E[Uxy] = 88200.0000; V[Uxy] = 1.2329E7
Uyx = 156824.0000; E[Uyx] = 88200.0000; V[Uyx] = 1.2329E7

Estadístico de contraste de U: 19.5436
p-valor de U: 0.0000

```

Figura 29. Densidad léxica en la muestra del M1

Por tanto, ante este resultado, podemos afirmar que los estudiantes ya al inicio del curso de nivel B1 son capaces de producir textos con cierto grado de densidad léxica.

Al analizar la diferencia entre la media de PL (M=0.8049) y PF (M=0.1951) en la muestra del M2, al igual que con el resultado de la muestra del M1, debemos rechazar la  $H_0$  de igualdad, ya que el p-valor es inferior al valor de significación de 0.05 (M2: p-valor = 0.0000 < 0.05), como podemos observar en la figura 30. Es decir, estamos ante una diferencia estadísticamente significativa.

```

Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa: distribución PL no igual distribución PF

Uxy = 28475.5000; E[Uxy] = 102152.0000; V[Uxy] = 1.5382E7
Uyx = 175828.5000; E[Uyx] = 102152.0000; V[Uyx] = 1.5382E7

Estadístico de contraste de U: 18.7853
p-valor de U: 0.0000

```

Figura 30. Densidad léxica en la muestra del M2

Una vez comprobado que podemos considerar las producciones de ambas muestras como textos densos léxicamente, pasamos a compararlas entre sí, ya que el objetivo final en esta fase es comprobar si el aumento en el valor de la media de PL, (M1: M=0,81; M2:

M=0,80) entre el M1 y el M2 también es significativo estadísticamente. Así, aplicamos nuevamente el test U de Mann-Whitney ahora a las dos muestras para comprobarlo.

```
Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa:  distribución M1 no igual distribución M2

Uxy = 88653.5000;  E[Uxy] = 94920.0000;  V[Uxy] = 1.3808E7
Uyx = 101186.5000;  E[Uyx] = 94920.0000;  V[Uyx] = 1.3808E7

Estadístico de contraste de U:    1.6864
p-valor de U:                    0.0917
```

Figura 31. Densidad léxica entre la muestra del M1 y del M2

Los datos que arrojan el test de Mann-Whitney (figura 31) hacen que no podamos rechazar la  $H_0$  de igualdad, ya que el p-valor es mayor al valor de significación de 0.05 (p-valor=0.0917 < 0.05). Por tanto, aquí vemos que la diferencia en la media de palabras léxicas entre el M1 y el M2 no es significativa estadísticamente. Eso significa que tras una instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, los sujetos del M2 no mejoran su rendimiento en relación a escribir textos más densos léxicamente.

### Resumen de la pregunta 3:

Como acabamos de comprobar, el aumento del número total de palabras (*tokens*) que se produce en el grupo de no nativos entre la muestra del M1 (inicio del curso de nivel B1) y del M2 (final del curso de nivel B1) es significativo estadísticamente y esto nos permite afirmar que los sujetos del M2 presentan una mayor **fluidez** en sus producciones al ser capaces de escribir textos más extensos en el mismo tiempo (30 minutos) que en la muestra del M1 y en el mismo contexto, dentro del aula. Asimismo, hemos verificado que las producciones de la muestra del M2 también ganan en **diversidad léxica** respecto a las del M1, y que ese aumento es igualmente significativo estadísticamente. En cuanto a la **densidad léxica**, hemos comprobado, por un lado, que la diferencia entre la media de PL y PF dentro de cada una de las muestras es estadísticamente significativa, mostrando, de esta manera, que estamos ante producciones densas léxicamente desde el inicio de curso. Por otro lado, al comparar la diferencia de la densidad léxica entre las muestras del M1 y del M2,

constatamos que no hay diferencia significativa estadísticamente. Es decir, que la densidad léxica no mejora en la muestra del M2. Así, estos datos permiten concluir que tras una instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, la competencia escrita de los estudiantes mejora de forma significativa a nivel de fluidez y diversidad léxica, pero no en densidad léxica.

**Pregunta 4: ¿Hay diferencia en los resultados de fluidez, diversidad léxica y densidad léxica en aquellas producciones que se trabajaron a través de la enseñanza mediante tareas (relato en pasado creativo) frente a las que no (experiencia personal) entre las muestras del M1 y del M2?**

Para responder a la cuarta pregunta de investigación:

(i) En primer lugar vamos a comparar las muestras «F» y «1D» en el M1 (fase 1a/b) y en el M2 (fase 2a/b) para ver qué diferencias hay entre dichas muestras en cada uno de los dos momentos respecto a fluidez, diversidad léxica y densidad léxica.

(ii) A continuación (fase 3a/b/c) vamos a comprobar si las diferencias que hemos observado en la estadística descriptiva en cuanto a números totales entre las muestras «F» y «1D» del M1 y del M2 afectan a la fluidez, diversidad léxica y densidad léxica y si esas diferencias son significativas estadísticamente.

**Fase 1a:** Fluidez y diversidad léxica de las muestras «F» y «1D» en el M1.

En la tabla siguiente presentamos los totales de *tokens* y *types* de los que partimos para realizar el análisis.

Muestras «F» y «1D»		
	«F» (n=71)	«1D» (n=34)
<i>Tokens</i>	9.088	3.097
<i>Types</i>	6.157	2.248

Tabla 121. Totales de *Tokens* y *Types* del M1. Muestras «F» y «1D»

Como el tamaño de las muestras es muy diferente, los totales tanto de *tokens* como de *types* no aportan una idea clara sobre la diferencia que puede haber entre ambas. Así, a continuación presentamos las medias según los descriptores obtenidos en la estadística descriptiva (tablas 53 y 54, para la muestra «F»; tablas 64 y 65, para la muestra «1D») para tener una idea más aproximada.

Muestras «F» y «1D»		
	«F» (n=71)	«1D» (n=34)
<i>Tokens</i>	128,0000	91,0882
<i>Types</i>	86,7183	66,1176

Tabla 122. Medias de *Tokens* y *Types* del M1. Muestras «F» y «1D»

Comprobamos en la tabla de medias que la muestra «F» registra valores superiores a los de la muestra «1D». Estos datos nos hacen entender que las producciones de la muestra «F» tienen mayor fluidez, es decir, que tienen una mayor extensión que las de la muestra «1D». A partir de estos valores, vamos a ver si la diferencia que observamos es significativa estadísticamente o no.

El primer paso es comprobar la bondad de ajuste a la normalidad para poder decidir la prueba estadística más apropiada para realizar el análisis.

Como vemos en los resultados de la tabla 123, no podemos rechazar la  $H_0$  de igualdad, es decir, que hay ajuste a la normalidad, ya que el valor del p-valor en ambos casos es superior al valor de significación 0.05. Así, el siguiente paso es comprobar la igualdad de varianzas. Para ello aplicamos el test F-Snedecor.

Bondad de ajuste a la normalidad	
F_M1	1D_M1
Para W = 0.9902 y n = 71	Para W = 0.9453 y n = 34
p-valor = <b>0.8570 &gt; 0.05</b>	p-valor = <b>0.0886 &gt; 0.05</b>

Tabla 123. Prueba de *Shapiro-Wilk* para las muestras «F» y «1D» en el M1

Los resultados de la prueba F-Snedecor hacen que tengamos que rechazar la  $H_0$  de igualdad de varianzas, ya que el valor del p-valor es inferior al valor de significación de 0.05, como podemos comprobar en la siguiente figura.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:       no igual
Estadístico de contraste F:  0.1409
p-valor:                     0.0007E-8

```

Figura 32. Prueba de F-Snedecor

Así, es la prueba estadística t-Student de varianzas diferentes la que vamos a aplicar para la comparación del grado de fluidez entre las muestras «F» y «1D».

Tras realizar el tratamiento estadístico, debemos rechazar la  $H_0$  de igualdad (figura 33), ya que el resultado del p-valor es inferior al valor de significación de 0.05 (p-valor=0.0003E-3 < 0.05). Este resultado indica que la diferencia observada entre la muestra «F» y «1D» es significativa estadísticamente. Así, podemos concluir que la muestra «F» presenta mayor fluidez que la muestra «1D».

```

t-Student
-----
Hipótesis Nula:                diferencia de medias = 0.0000
Hipótesis Alternativa:       no igual
t-Student (Satterthwaite):    -6.1795
p-valor:                     0.0003E-3
    
```

Figura 33. Prueba T-Student para varianzas diferentes

Una vez comprobada el grado de fluidez, pasamos a comparar la diversidad léxica de ambas muestras. Esta comparación la vamos a llevar a cabo a partir del índice Uber U cuya fórmula es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Así, a partir de los totales que hemos presentado en la tabla 124, obtenemos el índice Uber U de 0.17 para la muestra «F» y de 0.14 para la muestra «1D».

Para averiguar qué prueba estadística es la más apropiada para realizar la comparación, lo primero es calcular la bondad de ajuste a la normalidad con el test de Shapiro-Wilk. Los resultados que hemos obtenido los presentamos en la siguiente tabla.

<b>Bondad de ajuste a la normalidad</b>	
<b>F-1_M1</b>	<b>1D_M1</b>
W Shapiro-Wilk: 0.9701	W Shapiro-Wilk: 0.9673
p-valor Shapiro-Wilk: <b>0.0883 &gt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.3916 &gt; 0.05</b>

Tabla 124. Prueba de *Shapiro-Wilk* de las muestras «F» y «1D» para el M1

Los resultados que muestra la tabla hacen que no podamos rechazar la  $H_0$  de ajuste a la normalidad, pues el p-valor es superior al valor de significación de 0.05. Al tener muestras que siguen el ajuste a la normalidad, el siguiente paso es aplicar la prueba F-Snedecor para comprobar si las varianzas son iguales o diferentes.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:        no igual
Estadístico de contraste F:    0.6788
p-valor:                       0.1760

```

Figura 34. Prueba de F-Snedecor

El p-valor del test F-Snedecor indica que no podemos rechazar la  $H_0$  de igualdad de varianzas, ya que su valor es superior al valor de significación de 0.05 (p-valor = 0.1760 > 0.05). Por tanto, el test estadístico que vamos a aplicar para comprobar si la diferencia que arroja el índice Uber U es significativa o no es la t-Student de varianzas iguales.

```

t-Student
-----
Hipótesis Nula:                diferencia de medias = 0.0000
Hipótesis Alternativa:        no igual
t-Student:                    4.5242
p-valor:                       0.0002E-1

```

Figura 35. Prueba de t-Student para varianzas iguales

El tratamiento estadístico indica que debemos rechazar la  $H_0$  de igualdad, pues como vemos en la figura 35 el valor del p-valor es inferior al valor de significancia de 0.05. Así, podemos afirmar que la diferencia es significativa estadísticamente. Eso quiere decir que el tipo de texto de la muestra «F» presenta mayor diversidad léxica que el tipo de texto de la muestra «1D».

### **Fase 1b:** Densidad léxica de las muestras «F» y «1D» en el M1.

En esta fase de análisis vamos a comprobar qué grado de densidad léxica presenta cada una de las muestras, «F» y «1D» en el M1, para comparar a continuación si entre una muestra y otra hay diferencia significativa estadísticamente.

Para analizar la densidad léxica, partimos de los siguientes totales.

Muestras «F» y «1D» M1		
	«F» (n=71)	«1D» (n=34)
PL	7.457	2.458
PF	1.631	639

Tabla 125. Totales de palabras léxicas y palabras funcionales para el M1

En el gráfico 165 podemos visualizar las diferencias que observamos en la tabla. Pero teniendo en cuenta que la muestra «F» casi dobla en producciones escritas a la muestra «1D», vamos a realizar el tratamiento estadístico a partir de la media de palabras léxicas y funcionales de ambas muestras. En la tabla 126 y en el gráfico 166 presentamos la media, tanto de palabras léxicas como funcionales de cada una de las muestras.

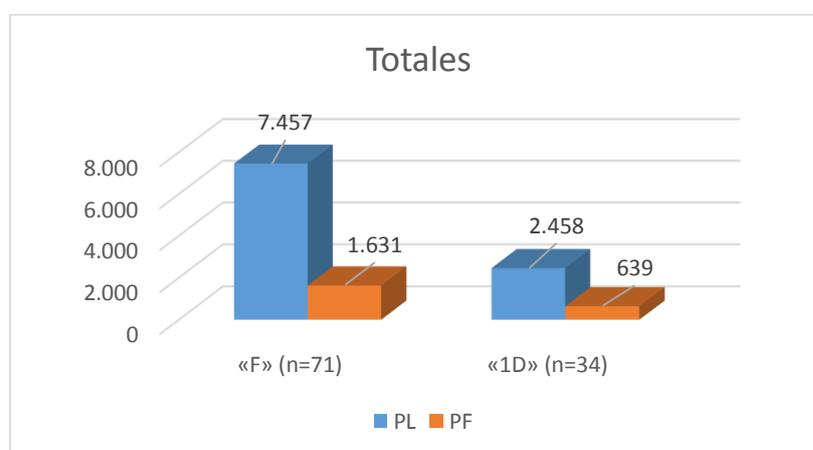


Gráfico 165. Totales de palabras léxicas y palabras funcionales para el M1

Muestras «F» y «1D» M1		
	«F» (n=71)	«1D» (n=34)
PL	0,8205	0,7902
PF	0,1794	0,2098

Tabla 126. Media de palabras léxicas y palabras funcionales para el M1

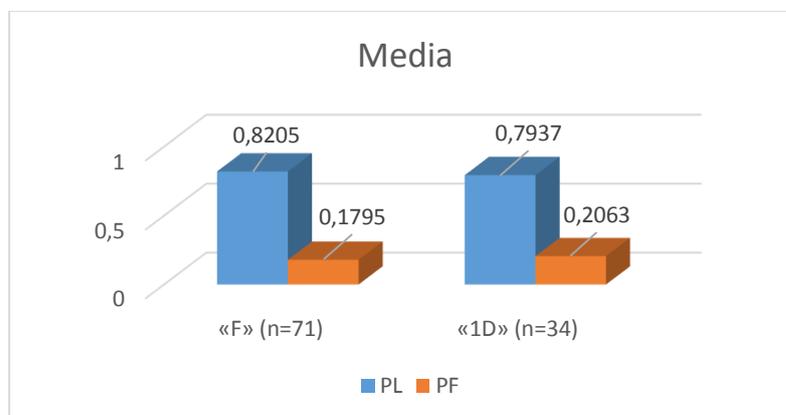


Gráfico 166. Media de palabras léxicas y palabras funcionales para el M1

Comprobamos, por un lado, que las diferencias tan marcadas observadas en los totales entre palabras léxicas y funcionales en cada una de las muestras se mantienen en el valor de la media; por otro lado, vemos que los valores de la media entre una muestra y otra están muy próximos, siendo el valor más alto para PL el de la muestra «F», mientras que el valor más alto de PF corresponde a la muestra «1D».

Así, vamos a realizar las pruebas estadísticas necesarias para verificar si esas diferencias son significativas estadísticamente o no.

En primer lugar, vamos a realizar la prueba de Shapiro—Wilk para comprobar la bondad de ajuste a la normalidad para decidir el test más apropiado para llevar a cabo el tratamiento estadístico.

Bondad de ajuste a la normalidad	
PL	PF
W Shapiro-Wilk: 0.8917	W Shapiro-Wilk: 0.9065
p-valor Shapiro-Wilk: <b>0.0002E-9 &lt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.0003E-8 &lt; 0.05</b>

Tabla 127. Prueba de *Shapiro-Wilk* para la muestra «F»

Bondad de ajuste a la normalidad	
PL	PF
W Shapiro-Wilk: 0.8723	W Shapiro-Wilk: 0.9004
p-valor Shapiro-Wilk: <b>0.0002E-5 &lt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.0005E-4 &lt; 0.05</b>

Tabla 128. Prueba de *Shapiro-Wilk* para la muestra «1D»

Los resultados del test de Shapiro-Wilk hacen que tengamos que rechazar la  $H_0$  de ajuste a la normalidad pues el p-valor es inferior al valor de significación de 0.05. Eso nos indica que la prueba estadística que debemos aplicar es el test U Mann-Whitney (Wilcoxon), una prueba no paramétrica, para comprobar si las diferencias en la media de PL («F»: media=0.8205; «1D»: media=0.7937) y PF («F»: media=0.1795; «1D»: media=0.2063) en la muestra «F» y «1D» son significativas estadísticamente.

En la figura 36 verificamos que la diferencia encontrada entre las PL y las PF de la muestra «F» es significativa estadísticamente al tener que rechazar la  $H_0$  de igualdad, pues el p-valor es inferior (p-valor = 0.0000 < 0.05) al valor de significación de 0.05. Así, este resultado indica que las producciones de la muestra «F» son textos densos léxicamente al constatar que la diferencia entre PL y PF es significativa.

**Estadístico de Mann-Whitney**

-----  
**Hipótesis Nula:** igualdad de distribuciones  
**Hipótesis Alternativa:** distribución PL\_F\_M1 no igual distribución PF\_F\_M1

**U<sub>xy</sub> = 7956.0000; E[U<sub>xy</sub>] = 40328.0000; V[U<sub>xy</sub>] = 3815385.7284**  
**U<sub>yx</sub> = 72700.0000; E[U<sub>yx</sub>] = 40328.0000; V[U<sub>yx</sub>] = 3815385.7284**

**Estadístico de contraste de U: 16.5730**  
**p-valor de U: 0.0000**

Figura 36. Prueba U Mann-Whitney para la muestra «F»

A continuación pasamos a comprobar si la diferencia en la muestra «1D» entre PL (media=0,7937) y PF (media=0,2063) es significativa estadísticamente.

Comprobamos en la figura 37 que debemos rechazar la  $H_0$  de igualdad al obtener un p-valor inferior al valor de significación 0.05 (p-valor = 0.0000 < 0.05). Por tanto, las producciones de la muestra «1D», al igual que las de la muestra «F», son textos densos léxicamente, pues la diferencia entre PL y PF es significativa estadísticamente.

**Estadístico de Mann-Whitney**

-----  
**Hipótesis Nula:** igualdad de distribuciones  
**Hipótesis Alternativa:** distribución PL\_1D\_M1 no igual distribución PF\_1D\_M1

**U<sub>xy</sub> = 2381.0000; E[U<sub>xy</sub>] = 9248.0000; V[U<sub>xy</sub>] = 419154.5092**  
**U<sub>yx</sub> = 16115.0000; E[U<sub>yx</sub>] = 9248.0000; V[U<sub>yx</sub>] = 419154.5092**

**Estadístico de contraste de U: 10.6067**  
**p-valor de U: 0.0000**

Figura 37. Prueba U Mann-Whitney para la muestra «1D»

Una vez constatado que podemos considerar ambas muestras como muestras densas léxicamente, vamos a compararlas entre sí para ver si la muestra «F», que presenta una media más alta de PL (media = 0.8205) que la muestra «1D» (media = 0.7937), registra una densidad léxica mayor estadísticamente significativa. Para ello, aplicamos nuevamente el test U Mann-Whitney a la media de PL de ambas muestras.

### Estadístico de Mann-Whitney

Hipótesis Nula: igualdad de distribuciones

Hipótesis Alternativa: distribución PL\_F\_M1 no igual distribución PL\_1D\_M1

$U_{xy} = 28028.5000$ ;  $E[U_{xy}] = 19312.0000$ ;  $V[U_{xy}] = 1354418.6645$

$U_{yx} = 10595.5000$ ;  $E[U_{yx}] = 19312.0000$ ;  $V[U_{yx}] = 1354418.6645$

Estadístico de contraste de U: -7.4897

p-valor de U: 0.0007E-10

Figura 38. Prueba U Mann-Whitney para las muestras «F» y «1D» del M1

El resultado que arroja el test estadístico indica que debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0007E-10). Así, podemos afirmar que las producciones de la muestra «F» son más densas léxicamente a pesar de que el valor de las medias de PL entre esta muestra y la muestra «1D» es pequeño.

### Fase 2a: Fluidez y diversidad léxica de las muestras «F» y «1D» en el M2.

En la tabla siguiente presentamos los totales de *tokens* y *types* de los que partimos para realizar el análisis.

Muestras «F» y «1D» M2		
	«F» (n=82)	«1D» (n=31)
<i>Tokens</i>	13.144	3.571
<i>Types</i>	8.592	2.572

Tabla 129. Totales de *Tokens* y *Types* de las muestras «F» y «1D» en el M2

A continuación presentamos el gráfico correspondiente a estos totales para visualizar más claramente las diferencias.

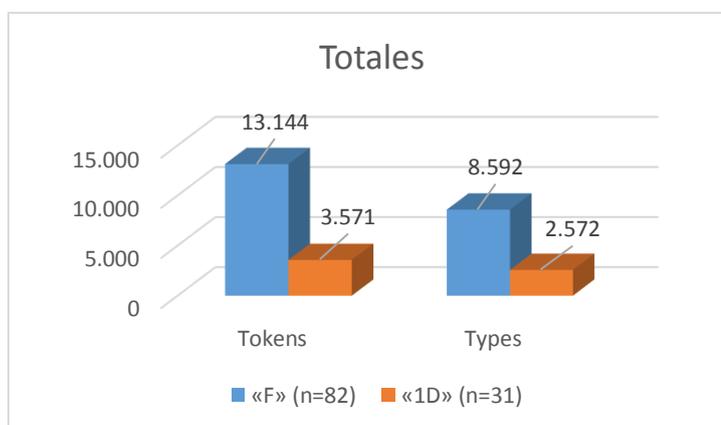


Gráfico 167. Totales de *Tokens* y *Types* de las muestras «F» y «1D» en el M2

Pero como el tamaño de las muestras es muy diferente, los totales tanto de *tokens* como de *types* no aportan una idea clara sobre la diferencia que puede haber entre ambas. Así, a continuación presentamos las medias a partir de los descriptores de la estadística descriptiva (tablas 53 y 54, para la muestra «F»; tablas 64 y 65, para la muestra «1D».

Podemos comprobar a través de la tabla 130 y del gráfico 168 que los valores de las medias de la muestra «F» son superiores a los de la muestra «1D», igual que sucede con los valores del M1. Estos datos confirman que las producciones de la muestra «F» tienen una mayor extensión que las de la muestra «1D». A partir de estos valores, vamos a ver si la diferencia que observamos es significativa estadísticamente o no.

Muestras «F» y «1D» M2		
	«F» (n=82)	«1D» (n=31)
<i>Tokens</i>	160,2927	115,1935
<i>Types</i>	104,7805	82,9677

Tabla 130. Medias de *Tokens* y *Types* del M2. Muestras F y 1D

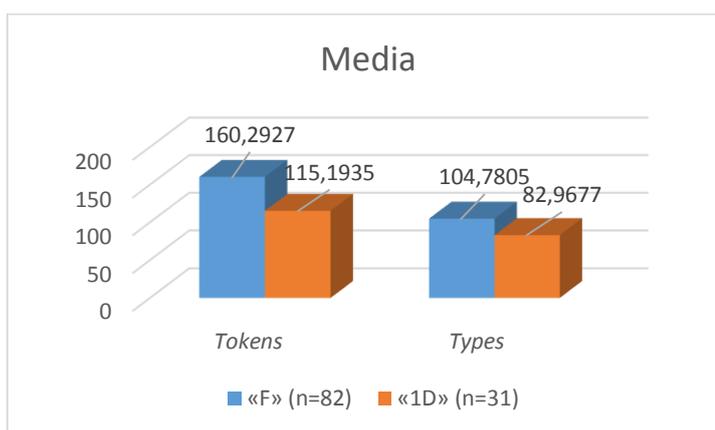


Gráfico 168. Medias de *Tokens* y *Types* de las muestras «F» y «1D» en el M2

El primer paso es comprobar la bondad de ajuste a la normalidad de la variable *tokens* para poder decidir la prueba estadística más apropiada para realizar el análisis de la fluidez entre ambas muestras.

Bondad de ajuste a la normalidad	
F_M2	1D_M2
Para W = 0.9858 y n = 82	Para W = 0.9661 y n = 31
p-valor = <b>0.5069</b> > 0.05	p-valor = <b>0.4197</b> > 0.05

Tabla 131. Prueba de *Shapiro-Wilk* para la variable *Tokens* de las muestras «F» y «1D» en el M2

Como vemos en los resultados de la tabla 131, no podemos rechazar la  $H_0$  de ajuste a la normalidad, ya que el p-valor en ambos casos es superior al valor de significación 0.05.

Así, el siguiente paso es comprobar la igualdad de varianzas. Para ello aplicamos el test F-Snedecor.

Los resultados de la prueba F-Snedecor hacen que tengamos que rechazar la  $H_0$  de igualdad de varianzas, ya que el p-valor es inferior al valor de significación de 0.05, como podemos comprobar en la siguiente figura.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:       no igual
Estadístico de contraste F:  2.6785
p-valor:                      0.0034

```

Figura 39. Prueba de F-Snedecor para el contraste de varianzas

Así, es la prueba estadística t-Student de varianzas diferentes la que vamos a aplicar para la comparación del grado de fluidez entre las muestras «F» y «1D» del M2.

```

t-Student
-----
Hipótesis Nula:                diferencia de medias = 0.0000
Hipótesis Alternativa:       no igual
t-Student (Satterthwaite):    7.3103
p-valor:                      0.0001E-6

```

Figura 40. Prueba T-Student para varianzas diferentes

Tras realizar el tratamiento estadístico, debemos rechazar la  $H_0$  de igualdad (figura 40), ya que el p-valor es inferior al valor de significación de 0.05 (p-valor=0.0001E-6 < 0.05). Este resultado indica que la diferencia observada entre la muestra «F» y «1D» es significativa estadísticamente. Así, podemos concluir que la muestra «F» presenta una mayor fluidez que la muestra «1D».

Una vez comprobada el grado de fluidez, pasamos a comparar la diversidad léxica de ambas muestras. Esta comparación la vamos a llevar a cabo a partir del índice Uber U cuya fórmula es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Así, a partir de los totales que hemos presentado en la tabla 132, obtenemos el índice Uber U de 0.18 para la muestra «F» y de 0.14 para la muestra «1D».

Para averiguar qué prueba estadística es la más apropiada para comprobar la diversidad léxica de las muestras «F» y «1D» en el M2, lo primero es calcular la bondad de ajuste a la normalidad con el test de Shapiro-Wilk. Los resultados que hemos obtenido los presentamos en la siguiente tabla.

<b>Bondad de ajuste a la normalidad</b>	
<b>F_M2</b>	<b>1D_M2</b>
W Shapiro-Wilk: 0.9824	W Shapiro-Wilk: 0.9454
p-valor Shapiro-Wilk: <b>0.3212 &gt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.1168 &gt; 0.05</b>

Tabla 132. Prueba de *Shapiro-Wilk* para el M2. Muestras F y 1D

Los resultados que muestra la tabla hacen que no podamos rechazar la  $H_0$  de ajuste a la normalidad, pues el p-valor es inferior al valor de significación de 0.05. Al tener muestras que siguen el ajuste a la normalidad, el siguiente paso es aplicar la prueba F-Snedecor para comprobar si las varianzas son iguales o diferentes.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:        no igual
Estadístico de contraste F:    1.1715
p-valor:                       0.6397

```

Figura 41. Prueba de F-Snedecor para el contraste de varianzas

El p-valor del test F-Snedecor indica que no podemos rechazar la  $H_0$  de igualdad, ya que su valor es superior al valor de significación de 0.05 ( $p\text{-valor} = 0.6397 > 0.05$ ). Por tanto, vamos a aplicar la prueba t-Student de varianzas iguales para el tratamiento estadístico y comprobar si la diferencia que arroja el índice Uber U es significativa estadísticamente.

```

t-Student
-----
Hipótesis Nula:                diferencia de medias = 0.0000
Hipótesis Alternativa:        no igual
t-Student:                    4.9848
p-valor:                       0.0002E-2

```

Figura 42. Prueba de t-Student para varianzas iguales

El tratamiento estadístico indica que debemos rechazar la  $H_0$  de igualdad, pues como vemos en la figura anterior el p-valor es inferior al valor de significación de 0.05 ( $p$ -

valor=0.0002E-2<0.05). Así, podemos afirmar que la diferencia es significativa estadísticamente. Eso quiere decir que el tipo de texto de la muestra «F» presenta mayor diversidad léxica que el tipo de texto de la muestra «1D».

**Fase 2b:** Densidad léxica de las muestras «F» y «1D» en el M2.

En esta fase de análisis vamos a comprobar qué grado de densidad léxica presenta cada una de las muestras, «F» y «1D» en el M2, para comparar a continuación si entre una muestra y otra hay diferencia significativa estadísticamente.

Para analizar la densidad léxica, partimos de los siguientes totales.

Muestras «F» y «1D» M2		
	«F» (n=82)	«1D» (n=31)
PL	10.615	2.839
PF	2.529	732

Tabla 133. Totales de palabras léxicas y palabras funcionales para el M2

En el gráfico 169 podemos visualizar las diferencias que observamos en la tabla.

Teniendo en cuenta que la muestra «F» es muy superior a la muestra «1D», vamos a realizar el tratamiento estadístico a partir de la media de palabras léxicas y funcionales de ambas muestras. En la tabla 134 y en el gráfico 170 presentamos la media tanto de palabras léxicas como funcionales de cada una de las muestras.

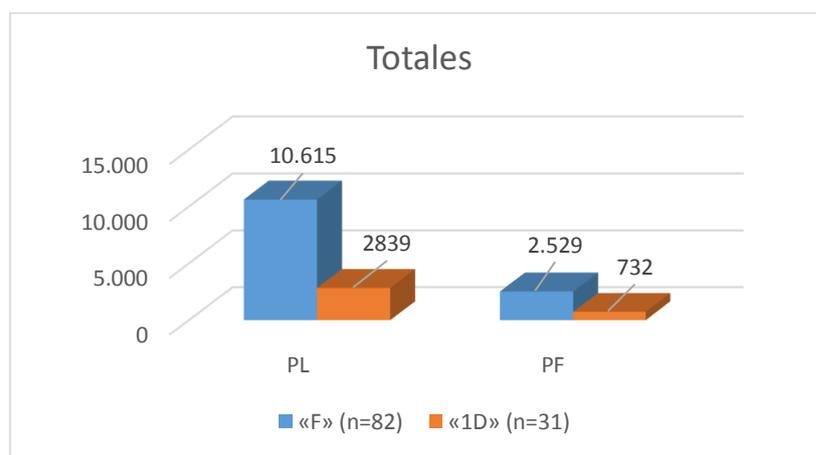


Gráfico 169. Totales de palabras léxicas y palabras funcionales para el M2

Muestras «F» y «1D» M2		
	«F» (n=82)	«1D» (n=31)
<b>PL</b>	0,8076	0,7950
<b>PF</b>	0,1924	0,2050

Tabla 134. Media de palabras léxicas y palabras funcionales para el M2

Comprobamos, por un lado, que las diferencias tan marcadas observadas en los totales entre PL y PF en cada una de las muestras se mantienen en la media; por otro lado, vemos que los valores de la media entre muestras están muy próximos.

Así, vamos a realizar las pruebas estadísticas necesarias para verificar si esas diferencias son significativas estadísticamente o no.

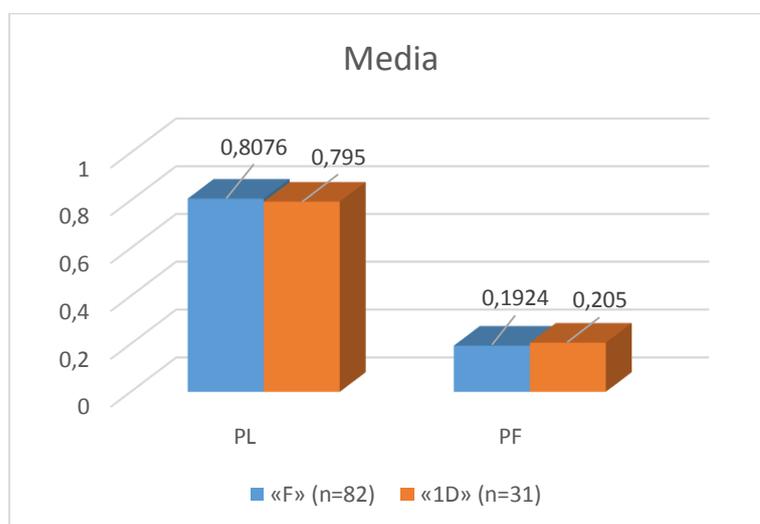


Gráfico 170. Media de palabras léxicas y palabras funcionales para el M2

El primer paso es comprobar la bondad de ajuste a la normalidad para decidir qué prueba estadística es la más apropiada para llevar a cabo el análisis de la densidad léxica en las muestras «F» y «1D» del M2.

En las siguientes tablas mostramos los resultados de aplicar el test de Shapiro-Wilk a las muestras «F» y «1D».

Bondad de ajuste a la normalidad	
PL	PF
W Shapiro-Wilk: 0.8426	W Shapiro-Wilk: 0.4881
p-valor Shapiro-Wilk: <b>0.0001E-13 &lt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.0001E-25 &lt; 0.05</b>

Tabla 135. Prueba de *Shapiro-Wilk* para la muestra «F» M2

Bondad de ajuste a la normalidad	
PL	PF
W Shapiro-Wilk: 0.9696	W Shapiro-Wilk: 0.8864
p-valor Shapiro-Wilk: <b>0.0067 &lt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.0003E-4 &lt; 0.05</b>

Tabla 136. Prueba de *Shapiro-Wilk* para la muestra «1D» M2

Podemos comprobar en las tablas anteriores que los valores del p-valor son inferiores al valor de significación 0.05, por tanto, debemos rechazar la  $H_0$  de ajuste a la normalidad. Así, el test que vamos a utilizar en el tratamiento estadístico es la prueba no paramétrica U de Mann-Whitney (Wilcoxon).

En la siguiente figura verificamos que la diferencia encontrada entre las PL (media=0.8076) y las PF (media=0.1924) de la muestra «F» es significativa estadísticamente al tener que rechazar la  $H_0$  de igualdad, pues el p-valor es inferior (p-valor = 0.0000 < 0.05) al valor de significación de 0.05. Así, este resultado indica que las producciones de la muestra «F» son textos densos léxicamente al constatar que la diferencia entre PL y PF es significativa.

```

Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa: distribución PL-F no igual distribución PF_F

Uxy = 14222.0000; E[Uxy] = 53792.0000; V[Uxy] = 5872165.5969
Uyx = 93362.0000; E[Uyx] = 53792.0000; V[Uyx] = 5872165.5969

Estadístico de contraste de U: 16.3293
p-valor de U: 0.0000

```

Figura 43. Prueba U Mann-Whitney para la muestra «F»

A continuación mostramos los resultados obtenidos en la muestra «1D».

```

Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa: distribución PL_1D no igual distribución PF_1D

Uxy = 2920.5000; E[Uxy] = 7688.0000; V[Uxy] = 318090.4980
Uyx = 12455.5000; E[Uyx] = 7688.0000; V[Uyx] = 318090.4980

Estadístico de contraste de U: 8.4531
p-valor de U: 0.0000

```

Figura 44. Prueba U Mann-Whitney para la muestra «1D»

Comprobamos en la figura anterior que debemos rechazar la  $H_0$  de igualdad al obtener un p-valor inferior al valor de significación 0.05 (p-valor = 0.0000 < 0.05). Por tanto, las producciones de la muestra «1D», al igual que las de la muestra «F», son textos densos léxicamente, pues la diferencia entre PL y PF es significativa estadísticamente.

Una vez constatado que podemos considerar ambas muestras como muestras densas léxicamente, vamos a compararlas entre sí para ver si la muestra «F», que presenta una media más alta de PL (media=0.8076) que la muestra «1D» (media = 0,7950), registra una densidad léxica mayor estadísticamente significativa. Para ello, aplicamos nuevamente el test U Mann-Whitney a la media de PL de ambas muestras.

```

Estadístico de Mann-Whitney
-----
Hipótesis Nula:          igualdad de distribuciones
Hipótesis Alternativa: distribución PL-F no igual distribución PL_1D

Uxy = 28694.0000; E[Uxy] = 20336.0000; V[Uxy] = 1534456.5052
Uyx = 11978.0000; E[Uyx] = 20336.0000; V[Uyx] = 1534456.5052

Estadístico de contraste de U:   -6.7472
p-valor de U:                   0.0002E-7

```

Figura 45. Prueba U Mann-Whitney para las muestras «F» y «1D» del M2

El resultado que arroja el test estadístico indica que debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0002E-7). Así, podemos afirmar que las producciones de la muestra «F» son más densas léxicamente a pesar de que el valor de las medias de PL entre esta muestra y la muestra «1D» es mínimo.

Una vez analizado cómo se comportan las muestras «F» y «1D» en cada uno de los estadios (M1; M2) en los que se ha centrado la investigación, pasamos a comprobar si el aumento en números totales que hemos observado que se produce en el M2 respecto al M1 es estadísticamente significativo. Para ello, en la siguiente fase 3 vamos a centrar el análisis en comparar la muestra «F» en el M1 y el M2 y la muestra «1D» en el M1 y el M2. Igual que en las fases anteriores, vamos a comparar la fluidez (fase 3a), la diversidad léxica (fase 3b) y la densidad léxica (fase 3c) entre el M1 y el M2 de las muestras «F» y «1D».

### Fase 3a: Fluidez en las muestras «F» y «1D» entre el M1 y el M2

Como ya hemos visto en las fases anteriores, para analizar la fluidez de las muestras partimos del número total de palabras (*tokens*). A continuación presentamos la tabla con los valores absolutos de cada una de las muestras y el gráfico correspondiente para visualizar con más detalle las diferencias entre las muestras en el M1 y el M2.

M1		M2	
«F» (n=71)	9.088	«F» (n=82)	13.144
«1D» (n=34)	3.097	«1D» (n=31)	3.571

Tabla 137. Total de *tokens* de las muestras «F» y «1D» en el M1 y el M2

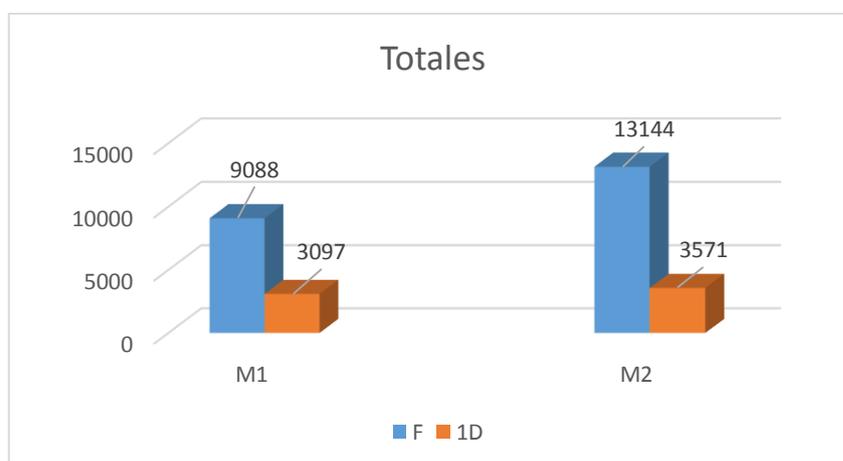


Gráfico 171. Total de *tokens* de las muestras «F» y «1D» en el M1 y el M2

Como podemos apreciar, las diferencias más llamativas las encontramos en la muestra «F», la cual aumenta en número total de palabras en el M2 respecto al M1 de forma considerable. Mientras que el aumento que se produce en la muestra «1D» no es tan vistoso.

No obstante, vamos a ver qué dicen los valores de la media para constatar si esas diferencias en números absolutos se confirman.

M1		M2	
«F» (n=71)	128.0000	«F» (n=82)	160.2927
«1D» (n=34)	86.7183	«1D» (n=31)	115.1935

Tabla 138. Media de *tokens* de las muestras «F» y «1D» en el M1 y el M2

Observamos en la tabla 138 y en el gráfico 172 que los valores de la media también son superiores en el M2 respecto al M1. Así, confirmamos que las producciones escritas de las muestras del M2 (final de curso) son más extensas que las del inicio de curso, M1.

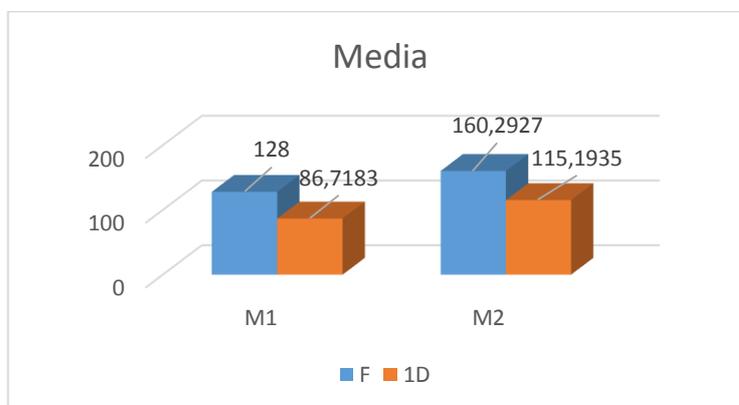


Gráfico 172. Media de *tokens* de las muestras «F» y «1D» en el M1 y el M2

Pero para verificar si esas diferencias son significativas, vamos a aplicar las pruebas estadísticas necesarias.

El primer paso es comprobar la bondad de ajuste a la normalidad para poder decidir qué test es el más apropiado para llevar a cabo el análisis.

Bondad de ajuste a la normalidad M1	
«F»	«1D»
W Shapiro-Wilk: 0.9902	W Shapiro-Wilk: 0.9453
p-valor Shapiro-Wilk: <b>0.8570 &gt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.0886 &gt; 0.05</b>

Bondad de ajuste a la normalidad M2	
«F»	«1D»
W Shapiro-Wilk: 0.9858	W Shapiro-Wilk: 0.9661
p-valor Shapiro-Wilk: <b>0.5069 &gt; 0.05</b>	p-valor Shapiro-Wilk: <b>0.4197 &gt; 0.05</b>

Tabla 139. Prueba de Shapiro-Wilk para las muestras «F» y «1D» en el M1 y el M2

Como vemos en los resultados de la tabla 139, no podemos rechazar la  $H_0$  de ajuste a la normalidad, pues todos los valores del p-valor son superiores al valor de significación de 0.05. Por tanto, al ser muestras que siguen la bondad de ajuste a la normalidad, para el tratamiento estadístico vamos a utilizar la prueba Anova, un test paramétrico para la comparación de más de dos muestras. En este caso, estamos comparando cuatro.

Como podemos observar en el resultado que proporciona la prueba Anova (figura 46), debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0002E-19).

**Anova Un Factor**

---

Variable Respuesta: M1-M2  
 Variable Explicativa: V\_Expl\_F-1D\_M1-M2  
 Número de Casos: 218

---

	Suma de Cuadrados	G.L.	Cuadrado Medio	F-valor	p-valor
Entre Grupos	132368.9183	3	44122.9728	46.9663	0.0002E-19
Dentro Grupos	201044.5496	214	939.4605		
Total (corr.)	333413.4679	217			

---

Figura 46. Prueba Anova para fluidez de las muestras «F» y «1D» en el M1 y el M2

Como el resultado que nos indica la prueba Anova es que hay diferencia, pero no entre qué grupos, para determinar, así, si podemos rechazar la  $H_0$  de igualdad, pasamos a realizar «comparaciones múltiples a posteriori» con el objetivo de averiguar de dónde proviene la diferencia. La misma prueba Anova proporciona la opción de realizar estas «comparaciones múltiples a posteriori». También se puede recurrir a pruebas estadísticas especializadas en la comparación de dos muestras, como las que hemos utilizado en las fases anteriores (t-Student o Mann-Whitney).

En la figura 47, vemos que salvo en una de las combinaciones, hay diferencia estadísticamente significativa en todas las demás. Para nuestra investigación, vamos a destacar la comparación que se establece entre las muestras:

- (i) «F» del M1 y del M2
- (ii) «1D» del M1 y del M2

Observamos que en dicho cotejo, el resultado de las «comparaciones múltiples a posteriori» señala que en estas muestras se produce diferencia significativa estadísticamente.

Así, podemos concluir que el aumento que hemos observado en las medias de estas muestras en el M2 respecto al M1 es significativo estadísticamente. Por tanto, podemos afirmar que tras una instrucción de 120 horas siguiendo el enfoque comunicativo y una

enseñanza mediante tareas, los sujetos son capaces de escribir textos más fluidos en el mismo tiempo, 30 minutos, que los sujetos del M1.

Las otras diferencias marcadas con asterisco en el test de comparaciones múltiples son las que ya hemos analizado con más detalle en las fases anteriores.

**Anova Un Factor, Comparaciones Múltiples**

Variable Respuesta: **Fluidez**  
 Variable Explicativa: **V\_Expl\_F-1D\_M1-M2**  
 Número de Casos: **218**

Método: **LSD al 95.00%**

V_Expl_F-1D_M1-M2	N	Media	Grupos Homogéneos
M1-1D	34	91.0882	X
M2-1D	31	115.1935	X
M1-F	71	128.0000	X
M2-F	82	160.2927	X

Contraste	Diferencia	+/- Límite
M1-F VS M2-F	*-32.2927	*9.7940
M1-F VS M1-1D	*36.9118	*12.6002
M1-F VS M2-1D	12.8065	13.0059
M2-F VS M1-1D	*69.2044	*12.3235
M2-F VS M2-1D	*45.0991	*12.7380
M1-1D VS M2-1D	*-24.1053	*15.0033

\* Diferencia estadísticamente significativa.

Figura 47. Prueba de Comparaciones múltiples para la fluidez de las muestras en el M1 y el M2

**Fase 3b:** Diversidad léxica en las muestras «F» y «1D» entre el M1 y el M2

En las fases anteriores hemos analizado la diversidad léxica de las muestras «F» y «1D» en el M1 y el M2 por separado. En esta fase, vamos a comprobar si las diferencias que hemos ido observando entre el M1 y el M2 son significativas estadísticamente.

Para realizar este análisis vamos a aplicar el índice Uber U cuya fórmula es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Para ello, como ya hemos visto en las fases anteriores, partimos del total de palabras (*tokens*) y del total de palabras distintas (*types*) de cada una de las muestras. A continuación mostramos la tabla con dichos totales.

<i>Tokens</i>			
M1		M2	
«F» (n=71)	9.088	«F» (n=82)	13.144
«1D» (n=34)	3.097	«1D» (n=31)	3.571
<i>Types</i>			
M1		M2	
«F» (n=71)	6.157	«F» (n=82)	8.592
«1D» (n=34)	2.248	«1D» (n=31)	2.572

Tabla 140. Total de *tokens* y *types* de las muestras «F» y «1D» en el M1 y el M2

Una vez aplicada la fórmula Uber U, los resultados obtenidos los mostramos en la siguiente tabla.

Índice Uber U			
M1		M2	
«F»	<b>0.17</b>	«F»	<b>0.18</b>
«1D»	<b>0.14</b>	«1D»	<b>0.14</b>

Tabla 141. Índice Uber U para las muestras «F» y «1D» en el M1 y el M2

Como podemos ver en la tabla, el valor del índice Uber U aumenta en la muestra «F» en el M2 respecto al M1, sin embargo en la muestra «1D» es el mismo para ambos momentos (U= 0.14). Para comprobar si hay diferencia significativa entre las muestras del M1 y del M2, vamos a realizar las pruebas estadística que a continuación detallamos.

En primer lugar es averiguar si las muestras siguen la bondad de ajuste a la normalidad. Para ello, aplicamos el test de Shapiro-Wilk.

Bondad de ajuste a la normalidad			
M1		M2	
«F» (n=71)	W Shapiro-Wilk: 0.9701 p-valor Shapiro-Wilk: <b>0.0883 &gt; 0.05</b>	«F» (n=82)	W Shapiro-Wilk: 0.9824 p-valor Shapiro-Wilk: <b>0.3212 &gt; 0.05</b>
«1D» (n=34)	W Shapiro-Wilk: 0.9673 p-valor Shapiro-Wilk: <b>0.3916 &gt; 0.05</b>	«1D» (n=31)	W Shapiro-Wilk: 0.9454 p-valor Shapiro-Wilk: <b>0.1168 &gt; 0.05</b>

Tabla 142. Prueba de *Shapiro-Wilk* de las muestras «F» y «1D» para el M1 y el M2

Los resultados del test de Shapiro-Wilk hacen que no podamos rechazar la  $H_0$  de ajuste a la normalidad, ya que todos los valores del p-valor son superiores al valor de significación de 0.05. Como las muestras siguen la bondad de ajuste a la normalidad, para el tratamiento estadístico vamos a utilizar la prueba Anova, un test paramétrico para la comparación de más de dos muestras.

**Anova Un Factor**

---

Variable Respuesta: UberU\_F-1D\_M1-M2  
 Variable Explicativa: V\_Expl\_F-1D\_M1-M2  
 Número de Casos: 218

---

	Suma de Cuadrados	G.L.	Cuadrado Medio	F-valor	p-valor
Entre Grupos	0.0726	3	0.0242	17.1295	0.0005E-6
Dentro Grupos	0.3025	214	0.0014		
Total (corr.)	0.3751	217			

---

Figura 48. Prueba Anova para la diversidad léxica de las muestras «F» y «1D» en el M1 y el M2

Como podemos observar en la figura anterior, el resultado que proporciona la prueba Anova hace que debamos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0005E-6).

Como la prueba Anova solo indica si hay diferencia significativa, pero no entre qué grupos, pasamos a realizar las «comparaciones múltiples a posteriori» que ofrece la misma prueba Anova con el objetivo de averiguar entre qué muestras se da la diferencia.

En la figura 49, lo primero que llama la atención es que las muestras «F» en el M1 y el M2, y las muestras «1D» en el M1 y el M2 aparecen como homogéneas. Es por eso que vemos que aparecen marcadas con asterisco solo cuatro de las seis posibles combinaciones para comparar las muestras de dos en dos. Es decir, que dichas muestras presentan diferencia estadísticamente significativa. Mientras que las muestras M1-F y M2-F, por un lado, y M1-1D y M2-1D, por otro, al ser consideradas como homogéneas, no están marcadas con asterisco. Eso quiere decir que no presentan diferencia estadísticamente significativa. En el caso de la muestra «1D» se entiende que no haya diferencia, pues el valor del índice Uber U es el mismo para ambas muestras ( $U=0.14$ ). Sin embargo, en la otra muestra, «F», sí hay diferencia en el valor del índice Uber U, aunque es pequeña ( $U=0.17$ , para el M1;  $U=0.18$ ,

para el M2). Puesto que la respuesta esperada era encontrar diferencia significativa, decidimos aplicar pruebas estadísticas específicas para la comparación de dos muestras (t-Student o Mann-Whitney), ya que hemos constatado en las fases anteriores la consistencia de los resultados de dichas pruebas. De esta manera, podemos comprobar exactamente el p-valor resultante de comparar la diferencia en el índice Uber U de ambas muestras.

**Anova Un Factor, Comparaciones Múltiples**

**Variable Respuesta:** UberU\_F-1D\_M1-M2  
**Variable Explicativa:** V\_Expl\_F-1D\_M1-M2  
**Número de Casos:** 218

**Método:** LSD al 95.00%

V_Expl_F-1D_M1-M2	N	Media	Grupos Homogéneos
M1-1D	34	0.1329	X
M2-1D	31	0.1397	X
M1-F	71	0.1679	X
M2-F	82	0.1799	X

Contraste	Diferencia	+/- Límite
M1-F VS M2-F	-0.0120	0.0120
M1-F VS M1-1D	*0.0349	*0.0155
M1-F VS M2-1D	*0.0282	*0.0160
M2-F VS M1-1D	*0.0469	*0.0151
M2-F VS M2-1D	*0.0402	*0.0156
M1-1D VS M2-1D	-0.0067	0.0184

\* Diferencia estadísticamente significativa.

Figura 49. Prueba: Comparaciones múltiples para la diversidad léxica de las muestras en el M1 y el M2.

Así, pasamos a aplicar este tratamiento estadístico a estas muestras, M1-F y M2-F, con el objetivo de confirmar si realmente no presentan diferencia estadísticamente significativa.

Para llevar a cabo el tratamiento estadístico más adecuado en la comparación de las muestras M1-F y M2-F, una vez comprobada la bondad de ajuste a la normalidad (tabla 42), el siguiente paso es aplicar el test de F-Snedecor para el contraste de varianzas.

```

F-Snedecor
-----
Hipótesis Nula:                cociente de varianzas= 1.0000
Hipótesis Alternativa:       no igual
Estadístico de contraste F:  0.7864
p-valor:                      0.3039

```

Figura 50. Prueba F-Snedecor: contraste de varianzas de la muestra «F» en el M1 y el M2

El resultado que proporciona el test hace que no podamos rechazar la  $H_0$  de igualdad, pues el p-valor es superior al valor de significación de 0.05 ( $p\text{-valor}=0.3039 > 0.05$ ). Así, será la prueba t-Student de varianzas iguales la que vamos a utilizar para comprobar si la diferencia en el índice Uber U entre la muestra «F» del M1 ( $U=0.17$ ) y del M2 ( $U=0.18$ ) es significativa estadísticamente o no, como ha señalado el resultado de la Anova.

```

t-Student
-----
Hipótesis Nula:                diferencia de medias = 0.0000
Hipótesis Alternativa:       no igual
t-Student:                   -2.0022
p-valor:                      0.0471

```

Figura 51. Prueba t-Student de varianzas iguales para la muestra «F» entre el M1 y el M2

El resultado de la prueba t-Student de varianzas iguales hace que debamos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 ( $p\text{-valor}=0.0471 < 0.05$ ). Vemos con este resultado que una prueba estadística específica para la comparación de dos muestras, afina más en su resultado y es capaz de detectar las pequeñas diferencias que presentan las muestras.

Así, podemos decir que en la muestra «F» entre el M1 y el M2 se produce un aumento significativo estadísticamente de la diversidad léxica. Es decir, los sujetos del M2, tras una instrucción de 120 horas, son capaces de escribir con mayor diversidad léxica relatos en pasado a partir de una frase dada. Recordemos que este tipo de texto es el que se trabajó en el aula siguiendo una enseñanza mediante tareas.

En cuanto al índice Uber U de la muestra «1D», hemos visto en la tabla 141 que es el mismo para el M1 y el M2 ( $U=0.14$ ). No obstante, vamos a verificar con el tratamiento estadístico específico para la comparación de dos muestras que no es posible rechazar la  $H_0$  de igualdad, como hemos visto con el resultado de la Anova.

Como ya hemos comprobado que las muestras siguen la bondad de ajuste a la normalidad, vamos a llevar a cabo la prueba F-Snedecor para el contraste de varianzas.

```
F-Snedecor  
-----  
Hipótesis Nula:                cociente de varianzas= 1.0000  
Hipótesis Alternativa:       no igual  
Estadístico de contraste F:  1.3674  
p-valor:                      0.3894
```

Figura 52. Prueba F-Snedecor: contraste de varianzas de la muestra «1D» en el M1 y el M2

El resultado que proporciona el test hace que no podamos rechazar la  $H_0$  de igualdad, pues el p-valor es superior al valor de significación de 0.05 ( $p\text{-valor}=0.3894 > 0.05$ ). Así, será la prueba t-Student de varianzas iguales la que vamos a utilizar para verificar la igualdad de la muestra «1D» del M1 y del M2.

```
t-Student  
-----  
Hipótesis Nula:                diferencia de medias = 0.0000  
Hipótesis Alternativa:       no igual  
t-Student:                   -0.6934  
p-valor:                      0.4906
```

Figura 53. Prueba t-Student de varianzas iguales para la muestra «1D» entre el M1 y el M2

El resultado de la prueba t-Student de varianzas iguales hace que no podamos rechazar la  $H_0$  de igualdad, pues el p-valor es superior al valor de significación de 0.05 ( $p\text{-valor}=0.4906 > 0.05$ ). De esta manera confirmamos el resultado que ya hemos visto en la prueba Anova, es decir, que no se produce aumento de la diversidad léxica entre el M1 y el M2 de la muestra «1D».

### **Fase 3c:** Densidad léxica en las muestras «F» y «1D» entre el M1 y el M2.

En esta fase de análisis vamos a comprobar si las diferencias en cuanto a la densidad léxica observadas en cada una de las muestras, «F» y «1D» entre el M1 y el M2, son significativas estadísticamente o no.

Para analizar si la diferencia en la densidad léxica de dichas muestras es significativa, y teniendo en cuenta que en las fases anteriores ya hemos comprobado que las muestras por separado son densas léxicamente, en esta fase vamos a realizar la comparación entre el M1 y el M2 a partir de las palabras léxicas.

Así, a continuación presentamos las tablas y los gráficos correspondientes a los totales y las medias de las muestras «F» y «1D» entre el M1 y el M2.

Total de Palabras Léxicas			
M1		M2	
PL: «F» (n=71)	<b>7.457</b>	PL: «F» (n=82)	<b>10.615</b>
PL: «1D» (n=34)	<b>2.458</b>	PL: «1D» (n=31)	<b>2.839</b>

Tabla 143. Total de PL de las muestras «F» y «1D» para el M1 y el M2

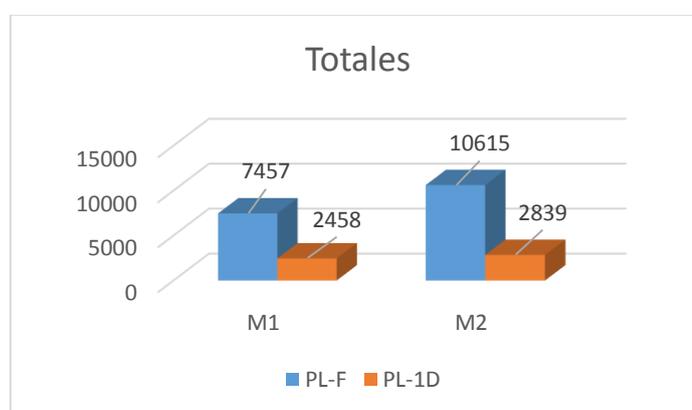


Gráfico 173. Total de PL de las muestras «F» y «1D» para el M1 y el M2

Media de Palabras Léxicas			
M1		M2	
PL: «F» (n=71)	<b>0,8205</b>	PL: «F» (n=82)	<b>0,8076</b>
PL: «1D» (n=34)	<b>0,7902</b>	PL: «1D» (n=31)	<b>0,7950</b>

Tabla 144. Media de PL de las muestras «F» y «1D» para el M1 y el M2

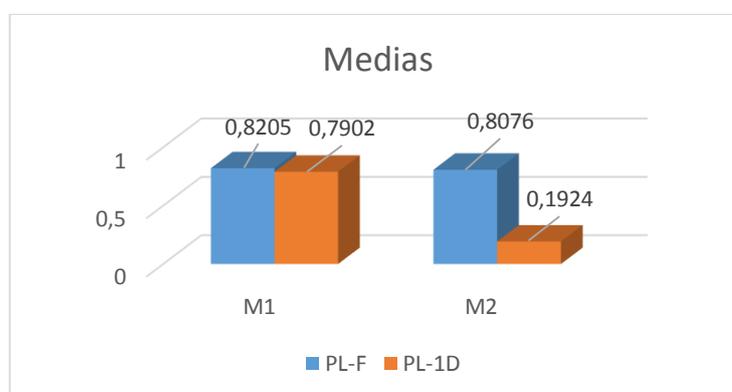


Gráfico 174. Media de PL de las muestras «F» y «1D» para el M1 y el M2

Observamos en la tabla y en el gráfico de totales de PL que las diferencias entre la muestra «F» del M1 y del M2 es llamativa, más de tres mil palabras de diferencia en el M2. Sin embargo, al comprobar los valores de la media, vemos que el valor más alto corresponde al M1 y no al M2, como sería de esperar por el número total de PL que presenta la muestra del M2. En cuanto a la muestra «1D», es el M2 el que presenta el valor más alto de media de PL, aunque con poca diferencia respecto al M1.

Para comprobar si esas diferencias que apreciamos son significativas o no, a continuación explicamos los pasos a seguir para decidir la prueba estadística más apropiada para realizar la comparación.

En primer lugar, vamos a comprobar la bondad de ajuste a la normalidad de las muestras. Para ello, aplicamos el test Shapiro-Wilk.

Bondad de ajuste a la normalidad		
	M1	M2
PL «F»	W Shapiro-Wilk: 0.8917 p-valor Shapiro-Wilk: <b>0.0002E-9 &lt; 0.05</b>	W Shapiro-Wilk: 0.8426 p-valor Shapiro-Wilk: <b>0.0001E-13 &lt; 0.05</b>
PL «1D»	W Shapiro-Wilk: 0.8723 p-valor Shapiro-Wilk: <b>0.0002E-5 &lt; 0.05</b>	W Shapiro-Wilk: 0.9696 p-valor Shapiro-Wilk: <b>0.0067 &lt; 0.05</b>

Tabla 145. Prueba de *Shapiro-Wilk* para la muestra «F» y «1D» en el M1 y el M2

El resultado que arroja el test de Shapiro-Wilk hace que debemos rechazar la  $H_0$  de ajuste a la normalidad, ya que todos los valores del p-valor son inferiores al valor de significación de 0.05. Por tanto, como son muestras que no siguen la bondad de ajuste a la normalidad, aplicaremos la prueba estadística no paramétrica de Kruskal-Wallis para la comparación de más de dos muestras. Esta prueba es la equivalente a la Anova.

Como vemos en la figura 54, el p-valor es inferior al valor de significación de 0.05 (p-valor=0.0000), esto significa que debemos rechazar la  $H_0$  de igualdad. Pero al igual que la prueba Anova, el test de Kruskal-Wallis solo señala que hay diferencia significativa, pero no entre qué muestras. Así, el siguiente paso es realizar las «comparaciones múltiples a posteriori» para comprobar dónde se encuentran las diferencias.

Podemos ver en la figura 54 que el test de comparaciones múltiples presenta como homogéneas las muestras «F» del M1 y del M2, y las muestras «1D» del M1 y del M2. De ahí que dichas comparaciones no las marque con asterisco. Es decir, que según el resultado del test de Kruskal-Wallis, no presentan diferencia estadísticamente significativa.

### Kruskal-Wallis

Variable Respuesta: Med\_PL-F-ID\_M1-M2  
Variable Explicativa: V\_Exp1\_PL-M1-M2  
Número de Casos: 872

Grupos	n	Suma de Rangos Rm	Rango Medio
PL_F_M1	284	115651.5000	407.2236
PL_ID_M1	136	81303.0000	597.8162
PL-F_M2	328	117318.5000	357.6784
PL-ID_M2	124	66355.0000	535.1210

Estadístico de Kruskal-Wallis (sin corrección por empates): 110.7599  
Estadístico de Kruskal-Wallis (con corrección por empates): 110.8011  
Grados de Libertad: 3  
p-valor: 0.0000

Figura 54. Prueba Kruskal-Wallis para la densidad léxica de las muestras «F» y «ID» en el M1 y el M2

Pero al igual que en la fase anterior, ante este resultado no esperado, decidimos constatarlo con pruebas estadísticas específicas para la comparación de dos muestras. Así, pasamos a aplicar el test de Mann-Whitney (Wilcoxon) para la comparación de dos muestras, ya que no siguen la bondad de ajuste a la normalidad (tabla 145).

El resultado del test de Mann-Whitney hace que debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0087 < 0.05). Así, la diferencia entre la muestra «F» del M1 (media=0,8205) y la muestra del M2 (media=0,8076) es significativa estadísticamente. Por tanto, debemos considerar que la muestra del M1 es más densa léxicamente.

En cuanto a la muestra «ID», también aplicamos el test de Mann-Whitney (Wilcoxon) para comprobar si la diferencia observada (M1; media = 0,7902; M2: media = 0,7950) entre el M1 y el M2 era significativa estadísticamente.

Kruskal-Wallis, Comparaciones Múltiples

Variable Respuesta: Med\_PL-F-1D\_M1-M2  
 Variable Explicativa: V\_Expl\_PL-M1-M2  
 Número de Casos: 872

Método: Dunn al 95.0%

Med_PL-F-1D_M1-M2	N	Rango Medio	Grupos Homogéneos
PL-F_M2	328	357.6784	X
PL_F_M1	284	407.2236	X
PL-1D_M2	124	535.1210	X
PL_1D_M1	136	597.8162	X

Contraste	Diferencia	+/- Limite
PL-F_M2 VS PL-1D_M2	*177.4426	*70.0378
PL_1D_M1 VS PL-1D_M2	-62.6952	82.4931
PL_F_M1 VS PL-1D_M2	*127.8974	*71.5107
PL_1D_M1 VS PL-F_M2	*-240.1378	*67.7585
PL_F_M1 VS PL-F_M2	-49.5452	53.8506
PL_F_M1 VS PL_1D_M1	*190.5926	*69.2799

\* Diferencia estadísticamente significativa.

Figura 55. Prueba: Comparaciones múltiples para la densidad léxica de las muestras «F» y «1D» en el M1 y el M2

Estadístico de Mann-Whitney

Hipótesis Nula: igualdad de distribuciones  
 Hipótesis Alternativa: distribución PL\_F\_M1 no igual distribución PL-F\_M2

U<sub>xy</sub> = 40858.0000; E[U<sub>xy</sub>] = 46576.0000; V[U<sub>xy</sub>] = 4755764.8128  
 U<sub>yx</sub> = 52294.0000; E[U<sub>yx</sub>] = 46576.0000; V[U<sub>yx</sub>] = 4755764.8128

Estadístico de contraste de U: 2.6220  
 p-valor de U: 0.0087

Figura 56. Prueba Mann-Whitney para la densidad léxica de la muestra «F» entre el M1 y el M2

#### Estadístico de Mann-Whitney

-----  
Hipótesis Nula: igualdad de distribuciones  
Hipótesis Alternativa: distribución PL\_ID\_M1 no igual distribución PL-ID\_M2  
  
U<sub>xy</sub> = 6987.0000; E[U<sub>xy</sub>] = 8432.0000; V[U<sub>xy</sub>] = 366525.9174  
U<sub>yx</sub> = 9877.0000; E[U<sub>yx</sub>] = 8432.0000; V[U<sub>yx</sub>] = 366525.9174  
  
Estadístico de contraste de U: 2.3868  
p-valor de U: 0.0170

Figura 57. Prueba Mann-Whitney para la densidad léxica de la muestra «1D» entre el M1 y el M2

El p-valor del test de Mann-Whitney hace que debamos rechazar la  $H_0$  de igualdad, pues es inferior al valor de significación de 0.05 ( $p\text{-valor}=0.0170 < 0.05$ ). Así, podemos afirmar que la diferencia es significativa estadísticamente. Eso quiere decir que la muestra «1D» del M2 presenta mayor densidad léxica que la del M1, aunque sea poca.

#### Resumen de la pregunta 4:

Para dar respuesta a esta cuarta pregunta, primero hemos mirado de clarificar qué muestra, «F» o «1D», presentaba más fluidez, diversidad y densidad léxica en el M1 y en el M2 para, a continuación, comprobar si había habido mejoría en la calidad de los textos entre ambos momentos.

En el M1, hemos podido constatar que es la muestra «F» la que exhibe mejores resultados en fluidez, diversidad léxica y densidad léxica. Los resultados de la prueba estadística t-Student de varianzas diferentes hizo que tuviéramos que rechazar la  $H_0$  de igualdad al registrar un  $p\text{-valor}=0.0003E-3$  ante una media de *tokens* de  $m=128.0000$  para la muestra «F» y una  $m=91.0882$  para la muestra «1D». Así, podemos decir que la muestra «F» presenta mayor **fluidez** que la muestra «1D». Igualmente, tras el tratamiento estadístico para el análisis de la diversidad léxica con la prueba t-Student de varianzas iguales, el resultado del  $p\text{-valor}=0.0002E-1$  hizo que tuviéramos que rechazar también la  $H_0$  de igualdad ante los índices Uber U de  $U=0.17$  para la muestra «F» y  $U=0.14$  para la muestra «1D». Es decir, también la muestra «F» revela más **diversidad léxica** que la muestra «1D». En cuanto a la **densidad léxica**, hemos visto, en un primer momento, que las producciones escritas de cada una de las dos muestras las podemos considerar densas léxicamente al mostrar una proporción mayor de PL que de PF; así nos lo demostró el análisis del test estadístico de

Mann-Whitney (Wilcoxon) al registrar un  $p$ -valor=0.0000. En un segundo momento de análisis, comparamos qué muestra presentaba mayor densidad léxica y, una vez más, fue la muestra «F» la que destacó. El resultado del test de Mann-Whitney (Wilcoxon) hizo de tuviéramos que rechazar la  $H_0$  de igualdad al registrar un  $p$ -valor=0.0007E-10.

El análisis de la fluidez, diversidad léxica y densidad léxica en las muestras «F» y «1D» en el M2 revelan, igual que en el M1, que es la muestra «F» la que presenta mejores resultados. En **fluidez**, por ejemplo, la comparativa de las medias de *tokens* ( $m=160.2927$  para la muestra «F» y  $m=115.1935$  para la muestra «1D») con la prueba t-Student de varianzas diferentes nos indica que la diferencia es estadísticamente significativa al tener que rechazar la  $H_0$  de igualdad ante un  $p$ -valor=0.0001E-6<0.05. También las diferencias del índice Uber U ( $U=0.18$  para la muestra «F» y  $U=0.14$  para la muestra «1D») revelaron ser significativas estadísticamente al obtener un  $p$ -valor=0.0002E-2<0.05 con la t-Student de varianzas iguales. Así, podemos decir que la muestra «F» presenta mayor **diversidad léxica** que la muestra «1D». Respecto a la **densidad léxica**, hemos comprobado que tanto la muestra «F» como la «1D» están conformadas por textos escritos densos léxicamente al contener una media de PL superior a la de PF y así lo ha constatado la prueba de Mann-Whitney (Wilcoxon) al registrar un  $p$ -valor=0.0000<0.05. En referencia a qué muestra presentaba mayor densidad léxica, la muestra «F» o la «1D», el  $p$ -valor=0.0002E-7<0.05, tras el análisis con el test de Mann-Whitney (Wilcoxon), hizo que tuviéramos que rechazar la  $H_0$  de igualdad. Así, una vez más, tenemos que decir que la muestra «F» es más densa léxicamente que la muestra «1D».

Para comparar las muestras entre el M1 y el M2, hemos recurrido a la prueba Anova y Kruskal-Wallis para la comparación de más de dos muestras, pero también a las pruebas t-Student de varianzas iguales y de Mann-Whitney (Wilcoxon) como pruebas específicas para la comparación de dos muestras para verificar los resultados del primer análisis.

En **fluidez**, hemos comprobado que la diferencia en la media de *tokens*, superior en las muestras del M2, era significativa estadísticamente. El resultado de la Anova y de las «comparaciones múltiples a posteriori» hicieron que tuviéramos que rechazar la  $H_0$  de igualdad, indicando, así, que las muestras «F» y «1D» presentaban mayor fluidez en el M2 respecto al M1. En **diversidad léxica**, la Anova mostró que había diferencia significativa estadísticamente ( $p$ -valor=0.0005E-6<0.05) entre las muestras del M1 y del M2. Pero al aplicar el test de «comparaciones múltiples a posteriori», éste indicó que no había diferencia

significativa estadísticamente entre las muestras «F» y «1D» entre el M1 y el M2, aunque sí entre otras combinaciones. Así, decidimos verificar esos datos con pruebas específicas para la comparación de dos muestras. Los resultados de la t-Student de varianzas iguales demostró que había diferencia significativa entre la muestra «F» del M1 ( $U=0.17$ ) y del M2 ( $U=0.18$ ) al registrar un  $p\text{-valor}=0.0471 < 0.05$  y, así, tener que rechazar la  $H_0$  de igualdad. En cuanto a la muestra «1D», al compartir el mismo índice Uber  $U$  en el M1 y el M2 ( $U=0.14$ ) el resultado del t-Student de varianzas iguales simplemente verificó el de la Anova de no diferencia significativa estadísticamente al no poder rechazar la  $H_0$  de igualdad al registrar un  $p\text{-valor}=0.4906 > 0.05$ . Podemos decir, por tanto, que la muestra «F» presenta más diversidad léxica en el M2 si aceptamos como válidos los resultados del test de Mann-Whitney (Wilcoxon). Sin embargo, si nos atenemos a los resultados de la Anova, hemos de decir que no se produce mejoría en cuanto a diversidad léxica tras una instrucción de 120 horas. El análisis de la **densidad léxica** mediante el test de Kruskal-Wallis y las «comparaciones múltiples a posteriori» reveló que no había diferencia significativa entre las medias de PL de las muestras «F» y «1D» entre el M1 y el M2. Sin embargo, al aplicar el test Mann-Whitney (Wilcoxon) para confirmar los resultados, comprobamos que las diferencias sí eran significativas estadísticamente al obtener un  $p\text{-valor}$  inferior al valor de significación de 0.05:  $p\text{-valor}=0.0087$  para la muestra «F» y un  $p\text{-valor}=0.0170$  para la muestra «1D». De esta manera, en el caso de la muestra «F», al presentar una media de PL mayor en el M1 que en el M2, tenemos que decir que no hay mejoría. Mientras que en la muestra «1D» sí hay mejoría en el M2 respecto al M1.

### 6.2.3 Resultados y análisis de comparar el grupo NN y el grupo N

En este apartado vamos a centrar nuestro análisis en la comparativa del grupo de NN y N con la finalidad de dar respuesta a la quinta pregunta de investigación.

**Pregunta 5: En las producciones escritas de los sujetos no nativos y los sujetos nativos, ¿hay diferencia en cuanto a fluidez, diversidad léxica y densidad léxica?**

Para responder a esta última pregunta, vamos a comparar las dos muestras, «F» y «1D», entre el grupo de NN (M1 y M2) y el grupo de N. Así, aquí estamos comparando las muestras por estadios, es decir, por niveles de lengua: (i) nivel inicial B1, (ii) nivel final B1, y (iii) nivel nativo. El objetivo es comprobar hasta qué punto las muestras del grupo de NN están próximas a las muestras del grupo de N.

La comparación la vamos a llevar a cabo en dos fases:

(i) En la fase uno, vamos a analizar cómo se comporta la muestra «F» en los tres niveles de lengua mencionados en cuanto a fluidez del texto (fase 1a), diversidad léxica (fase 1b) y densidad léxica (fase 1c).

(ii) En la segunda fase de análisis, vamos a centrarnos en la muestra «1D». También vamos a analizar la fluidez (fase 2a), la diversidad léxica (fase 2b) y la densidad léxica (fase 2c).

**Fase 1a:** Fluidez de la muestra «F» por nivel de lengua: NN-M1, NN-M2, N.

Para analizar la fluidez de la muestra «F», partimos del total de palabras escritas en cada uno de los niveles de lengua. En la siguiente tabla mostramos los valores absolutos.

Muestras «F» por estadios			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
<i>Tokens</i>	9.088	13.144	9.661

Tabla 146. Número total de tokens por nivel de lengua

Como vemos, si nos basamos en los números absolutos, es el estadio NN-M2 el que presenta un valor más alto. Pero, como ya hemos visto en los análisis anteriores, la comparativa la vamos a realizar a partir de la media de *tokens* que presentamos en la siguiente tabla (Tabla 147).

Muestras «F» por estadios			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
Media Tokens	128.0000	160.2927	175.6545

Tabla 147. Media de *tokens* por nivel de lengua

En estos valores vemos que se produce un aumento sucesivo en la media de *tokens* por nivel de lengua. Siendo, así, la muestra del grupo de N la que presenta una media mayor a las muestras del grupo de NN. En los siguientes gráficos podemos visualizar las diferencias que presentan los totales y las medias por nivel de lengua.

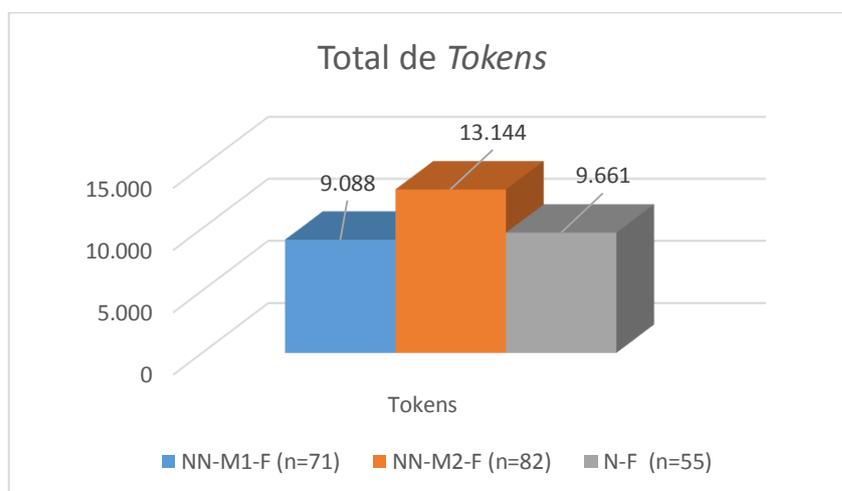


Gráfico 175. Total de *tokens* por nivel de lengua

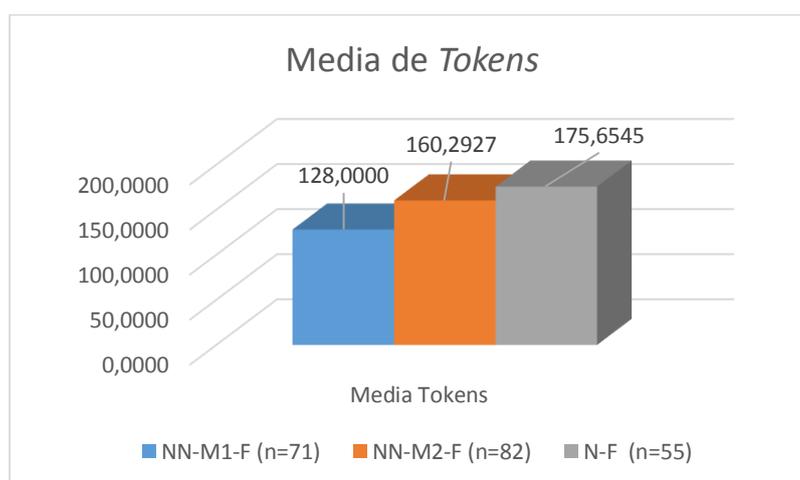


Gráfico 176. Media de *tokens* por nivel de lengua

Pero para comprobar si esas diferencias que se aprecian son significativas, vamos a realizar el tratamiento estadístico adecuado. Para ello, lo primero es comprobar la bondad de ajuste a la normalidad de las tres muestras.

Para comprobar la bondad de ajuste a la normalidad, aplicamos el test de Shapiro-Wilk.

Bondad d ajuste a la normalidad	
NN-M1 n = 71	Para W = 0.9902 p-valor = <b>0.8570 &gt; 0.05</b>
NN-M2 n = 82	Para W = 0.9858 p-valor = <b>0.5069 &gt; 0.05</b>
N n=55	Para W = 0.9666 p-valor = <b>0.5069 &gt; 0.05</b>

Tabla 148. Prueba de Shapiro-Wilk para *tokens* por nivel de lengua

Los resultados del test hacen que no podamos rechazar la  $H_0$  de ajuste a la normalidad, ya que el p-valor de las tres muestras es superior al valor de significación de 0.05. Por tanto, como son muestras que siguen la bondad de ajuste a la normalidad, el test que vamos a aplicar es la Anova, prueba paramétrica para la comparación de más de dos grupos.

#### Anova Un Factor

Variable Respuesta: M1-M2-N\_F  
 Variable Explicativa: Grupos  
 Número de Casos: 208

	Suma de Cuadrados	G.L.	Cuadrado Medio	F-valor	p-valor
Entre Grupos	76940.6986	2	38470.3493	38.1530	<b>0.0008E-11</b>
Dentro Grupos	206705.4120	205	1008.3191		
Total (corr.)	283646.1106	207			

Figura 58. Test Anova para la fluidez de las muestras NN-M1-F, NN-M2-F, N-F

Como vemos en la figura superior, el p-valor=0.0008E-11<0.05 obliga a rechazar la  $H_0$  de igualdad al presentar un valor inferior al valor de significación de 0.05. Este resultado lo que nos indica es que hay diferencia, pero no entre qué grupos; por tanto, a continuación

aplicamos el test de «comparaciones múltiples» para identificar dónde se producen las diferencias.

**Anova Un Factor, Comparaciones Múltiples**

**Variable Respuesta:** M1-M2-N\_F  
**Variable Explicativa:** Grupos  
**Número de Casos:** 208

**Método:** LSD al 95.00%

Grupos	N	Media	Grupos Homogéneos
NN-M1_F	71	128.0000	X
NN-M2_F	82	160.2927	X
N_F	55	175.6545	X

Contraste	Diferencia	+/- Limite
NN-M1_F VS NN-M2_F	*-32.2927	*10.1491
NN-M1_F VS N_F	*-47.6545	*11.2459
NN-M2_F VS N_F	*-15.3619	*10.9117

\* Diferencia estadísticamente significativa.

Figura 59. Test Anova-«comparaciones múltiples» para la fluidez de las muestras NN-M1-F, NN-M2-F, N-F

Vemos en la figura 59 que los grupos no presentan homogeneidad entre ellos, por tanto es lógico que haya diferencia significativa estadísticamente entre las distintas muestras. Comprobamos, en primer lugar, que se da diferencia significativa entre las muestras NN-M1\_F y NN-M2\_F, como ya hemos confirmado en respuesta a la pregunta cuatro de esta investigación. También hay diferencia significativa entre las dos muestras del grupo de NN y el de N. Es decir, ante estos resultados, y partiendo de la media de *tokens* que hemos presentado en la tabla 147, podemos afirmar que la muestra del grupo de N presenta mayor fluidez que las otras dos muestras.

**Fase 1b:** Diversidad léxica de la muestra «F» por nivel de lengua: NN-M1, NN-M2, N.

Para analizar qué muestra presenta mayor diversidad léxica, partimos de los datos de *tokens* y *types* que presentamos en la tabla siguiente.

Muestras «F»			
	NN-M1-F (n=71)	NN-M2-F (n=82)	N-F (n=55)
<i>Tokens</i>	9.088	13.144	9.661
<i>Types</i>	6.157	8.592	6.341

Tabla 149. Número total de *tokens* y *types* de la muestra «F» por nivel de lengua

En el siguiente gráfico podemos visualizar estos valores.

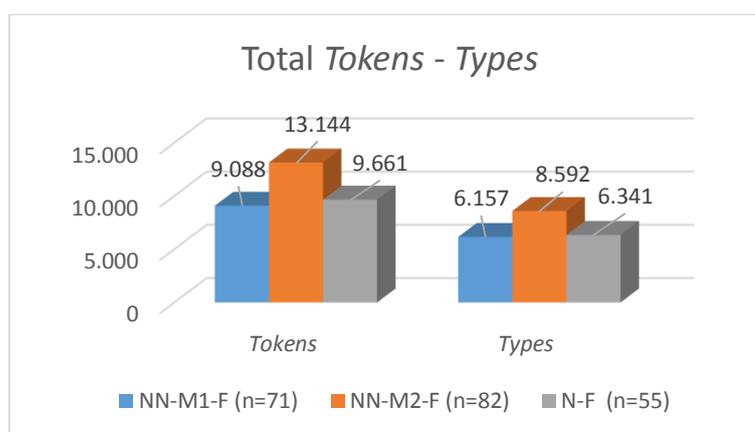


Gráfico 177. Total de *tokens* y *types* de la muestra «F» por estadios por nivel de lengua

A partir de estos datos, vamos a calcular el índice Uber U de cada una de las muestras. La fórmula del índice Uber U es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Los resultados obtenidos al aplicar la fórmula del índice Uber U son los que presentamos en la siguiente tabla:

Índice Uber U	
NN-M1_F n = 71	0.17
NN-M2_F n = 82	0.18
N_F n=55	0.18

Tabla 150. Cálculo del índice Uber U para la muestra «F» por nivel de lengua

Vemos en la tabla 150 que el índice Uber U es el mismo para la muestra NN-M2\_F (nivel lengua final B1) y la de N\_F (nivel lengua nativa); mientras que la muestra que

corresponde al nivel de lengua B1 inicial registra un índice Uber U algo inferior. Para averiguar qué prueba estadística es la más apropiada para verificar si estos valores presentan diferencias significativas, vamos a comprobar, en primer lugar, la bondad de ajuste a la normalidad con el test de Shapiro-Wilk.

Bondad d ajuste a la normalidad	
NN-M1 n = 71	Para W = 0.9299 p-valor = <b>0.0007 &lt; 0.05</b>
NN-M2 n = 82	Para W = 0.9824 p-valor = <b>0.3212 &gt; 0.05</b>
N n=55	Para W = 0.9847 p-valor = <b>0.7073 &gt; 0.05</b>

Tabla 151. Prueba de Shapiro-Wilk para índice Uber U de la muestra «F» por nivel de lengua

El resultado de la prueba de Shapiro-Wilk indica que debemos recurrir a una prueba no paramétrica para realizar el tratamiento estadística, ya que el p-valor de una de las muestras es inferior al valor de significación de 0.05 y por tanto debemos rechazar la  $H_0$  de ajuste a la normalidad. Así, vamos a aplicar el test de Kruskal-Wallis para la comparación de más de dos muestras.

#### Kruskal-Wallis

Variable Respuesta: M1\_F\_Uber  
Variable Explicativa: Grupos  
Número de Casos: 208

Grupos	n	Suma de Rangos Rm	Rango Medio
NN-M1_F	71	6203.5000	87.3732
NN-M2_F	82	9169.0000	111.8171
N_F	55	6363.5000	115.7000

Estadístico de Kruskal-Wallis (sin corrección por empates): 8.8652  
Estadístico de Kruskal-Wallis (con corrección por empates): 8.8654  
Grados de Libertad: 2  
p-valor: **0.0119**

Figura 60. Test Kruskal-Wallis para la diversidad léxica de las muestras NN-M1-F, NN-M2-F, N-F

Vemos en la figura 60 que debemos rechazar la  $H_0$  de igualdad al obtener un resultado de p-valor inferior al valor de significación de 0.05 (p-valor = 0.0119 < 0.05). Eso quiere decir que hay diferencia significativa entre las muestras. Pero para averiguar entre qué muestras se da la diferencia, debemos aplicar el test de «comparaciones múltiples».

En la figura 61, observamos que hay homogeneidad entre los grupos NN-M2\_F y el grupo N\_F. Recordemos que estos grupos presentan el mismo índice Uber U (U=0.18), por tanto, era de esperar que no presentaran diferencia, como se constata en el resultado que presenta el test de «comparaciones múltiples» al no marcar estos grupos con asterisco.

Sin embargo, sí marca asterisco, por tanto diferencia estadísticamente significativa, entre las dos muestras del grupo de NN (NN-M1\_F y NN-M2\_F) y también entre NN-M1\_F y N\_F. Es decir, la muestra del nivel de lengua de B1 final presenta la misma diversidad léxica que la muestra de lengua de nivel nativo. Mientras que la muestra del nivel de lengua B1 inicial presenta una diversidad léxica inferior a las otras dos muestras. Así, podemos decir que la instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas da como resultado una diversidad léxica igual a la del grupo de nivel de lengua nativa.

**Kruskal-Wallis, Comparaciones Múltiples**

**Variable Respuesta:** M1\_F\_Uber  
**Variable Explicativa:** Grupos  
**Número de Casos:** 208

**Método:** Dunn al 95.0%

M1_F_Uber	N	Rango Medio	Grupos Homogéneos
NN-M1_F	71	87.3732	X
NN-M2_F	82	111.8171	X
N_F	55	115.7000	X

Contraste	Diferencia	+/- Limite
NN-M1_F VS N_F	*28.3268	*25.8823
NN-M2_F VS N_F	3.8829	25.1131
NN-M2_F VS NN-M1_F	*-24.4438	*23.3581

\* Diferencia estadísticamente significativa.

Figura 61. Test Kruskal-Wallis-«comparaciones múltiples» para la diversidad léxica de las muestras NN-M1-F, NN-M2-F, N-F

**Fase 1c:** Densidad léxica de la muestra «F» por nivel de lengua: NN-M1, NN-M2, N.

En esta fase de análisis, vamos a comprobar si hay diferencia de densidad léxica entre las muestras del grupo de NN y el grupo de N. Para ello, vamos a tener en cuenta las palabras léxicas que presentan cada una de las muestras. Veamos lo totales y las medias correspondientes.

Palabras léxicas		
	Totales	Media
NN_M1_F	7.457	0.8205
NN_M2_F	10.615	0.8076
N_F	7.102	0.7351

Tabla 152. Totales y medias de PL de la muestra «F» por nivel de lengua

Presentamos a continuación los gráficos para ver con más claridad los valores que muestra la tabla.

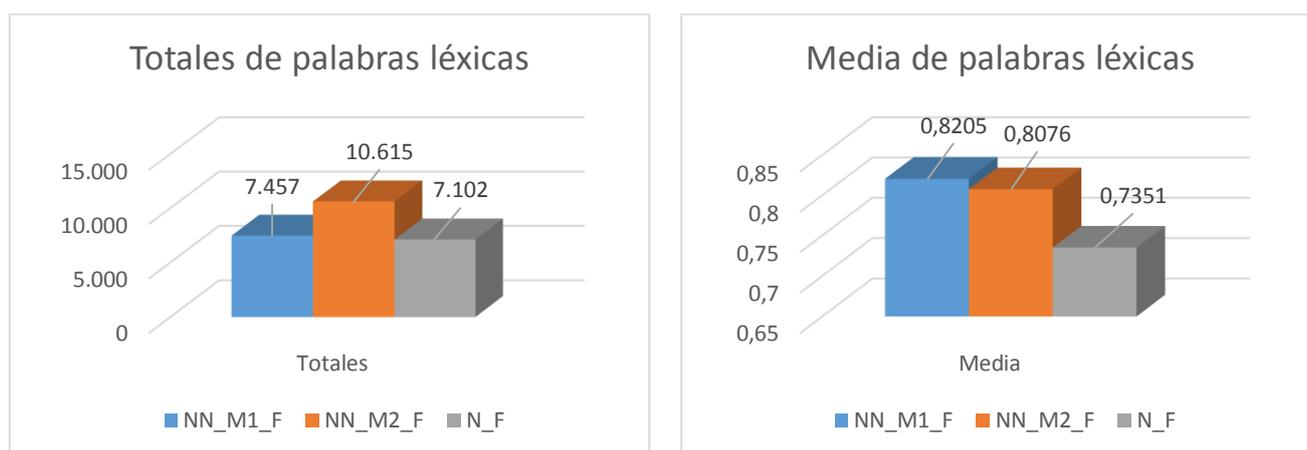


Gráfico 178. Totales y medias de PL de la muestra «F» por nivel de lengua

Observamos en los dos gráficos superiores que mientras en el de totales es la muestra del M2 del grupo de no nativos la que registra el valor más alto de palabras léxicas; en el gráfico de medias, el valor de la media más alto corresponde al grupo NN\_M1\_F (nivel de lengua B1 inicial) y este valor va disminuyendo progresivamente según aumenta el nivel de lengua, registrando el valor más bajo el grupo nativo.

Para comprobar si esas diferencias son significativas estadísticamente, primero debemos aplicar el test Shapiro-Wilk para verificar la bondad de ajuste a la normalidad y decidir qué tratamiento estadístico es el más apropiado.

Bondad d ajuste a la normalidad	
NN-M1 n = 71	Para W = 0.8917 p-valor = <b>0.0002E-9 &lt; 0.05</b>
NN-M2 n = 82	Para W = 0.8426 p-valor = <b>0.0001E-13 &lt; 0.05</b>
N n=55	Para W = 0.8732 p-valor = <b>0.0001E-8 &lt; 0.05</b>

Tabla 153. Prueba de Shapiro-Wilk para la densidad léxica de la muestra «F» por nivel de lengua

El resultado que arroja el test de Shapiro-Wilk hace que debamos rechazar la  $H_0$  de ajuste a la normalidad, ya que todos los p-valor son inferiores al valor de significación 0.05. Por tanto, como son muestras que no siguen la bondad de ajuste a la normalidad, vamos a utilizar el test Kruskal-Wallis, una prueba no paramétrica para la comparación de más de dos muestras.

#### Kruskal-Wallis

Variable Respuesta: Med\_PL\_F  
 Variable Explicativa: Grupos  
 Número de Casos: 832

Grupos	n	Suma de Rangos Rm	Rango Medio
NN_M1_F	284	121849.5000	429.0475
NN_M2_F	328	124871.0000	380.7043
N_F	220	99807.5000	453.6705

Estadístico de Kruskal-Wallis (sin corrección por empates): 13.3141  
 Estadístico de Kruskal-Wallis (con corrección por empates): 13.3208  
 Grados de Libertad: 2  
 p-valor: **0.0013**

Figura 62. Test Kruskal-Wallis para la densidad léxica de las muestras NN-M1-F, NN-M2-F, N-F

Comprobamos en la figura superior que debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 (p-valor = 0.0013 < 0.05). Eso nos indica que hay diferencia significativa entre alguna de las muestras, pero para saber dónde están esas diferencias debemos aplicar el test de «comparaciones múltiples».

Observamos en el test de «comparaciones múltiples» (Figura 63) que aparecen como homogéneas las muestras NN\_M1-F y N-F. Es decir, la del nivel de lengua de B1 inicial y la de nivel de lengua nativa. Por tanto, las diferencias que hemos apreciado en los valores de

la media ( $m=0.8205$  para NN\_M1;  $m=0.7351$  para N) no registran diferencia significativa estadísticamente. Eso quiere decir que estos dos niveles de lengua presentan una densidad léxica similar. Sin embargo, la muestra del nivel de lengua B1 final sí registra diferencia significativa estadísticamente con el grupo N y con el grupo NN\_M1. Así, podemos decir que el grupo NN\_M2 presenta mayor densidad léxica que el grupo de N, pero menos que el grupo de NN\_M1.

#### Kruskal-Wallis, Comparaciones Múltiples

Variable Respuesta: Med\_PL\_F  
 Variable Explicativa: Grupos  
 Número de Casos: 832

Método: Dunn al 95.0%

Med_PL_F	N	Rango Medio	Grupos Homogéneos
NN_M2_F	328	380.7043	X
NN_M1_F	284	429.0475	X
N_F	220	453.6705	X

Contraste	Diferencia	+/- Límite
NN_M1_F VS N_F	24.6229	51.6595
NN_M2_F VS N_F	*72.9662	*50.1242
NN_M2_F VS NN_M1_F	*48.3433	*46.6213

\* Diferencia estadísticamente significativa.

Figura 63. Test Kruskal-Wallis-«comparaciones múltiples» para la densidad léxica de las muestras NN-M1-F, NN-M2-F, N-F

**Fase 2a:** Fluidez de la muestra «1D» por nivel de lengua: NN-M1, NN-M2, N.

En esta parte de la investigación, nos vamos a centrar en la muestra «1D». Igual que en la fase anterior, vamos a ver cuál es el total de palabras que presenta cada una de las muestras.

Muestras «1D»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Tokens</i>	3.097	3.571	9.746

Tabla 154. Número total de *tokens* por nivel de lengua

Como vemos en la tabla, el número de sujetos participantes es muy diferente en las muestras del grupo de NN y el grupo de N. Así, los resultados en valores absolutos también varían. Es por ello que a continuación mostramos la media de palabras que registra cada grupo y poder de esta manera ver más claramente qué muestra puede presentar mayor fluidez.

Muestras «1D»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Tokens</i>	91.0882	115.1935	157.1935

Tabla 155. Media de *tokens* por nivel de lengua

Para apreciar mejor las diferencias que observamos en las tablas anteriores, presentamos los gráficos correspondientes a total y media de *tokens* de la muestra «1D».

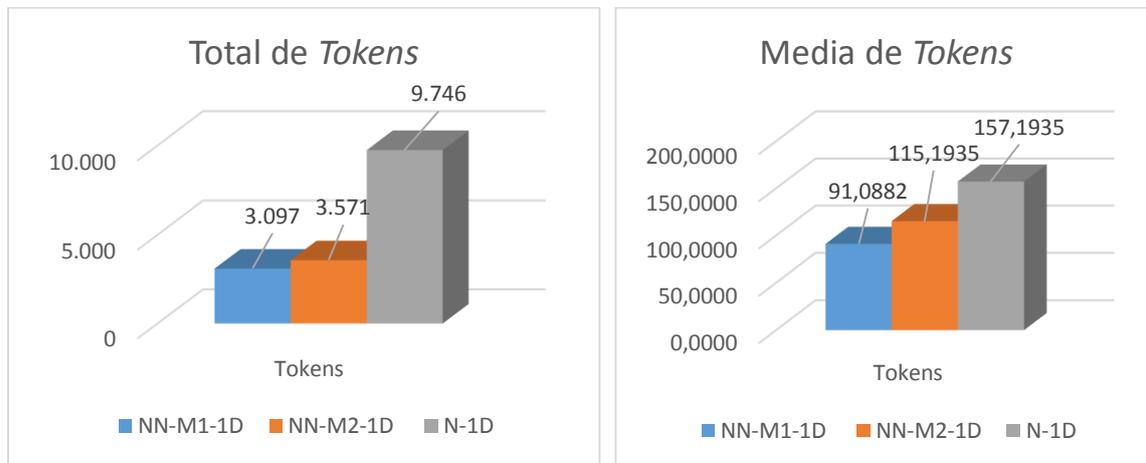


Gráfico 179. Totales y medias de *Tokens* de la muestra «1D» por nivel de lengua

Observamos en los gráficos que tanto en números totales como en la media de *tokens*, es la muestra del grupo de nativos la que registra los valores más altos. Para verificar si esas diferencias entre muestras son significativas estadísticamente, vamos, en primer lugar, a comprobar la bondad de ajuste a la normalidad para decidir qué prueba estadística es la más apropiada para realizar la comparación.

Bondad d ajuste a la normalidad	
NN-M1 n = 34	Para W = 0.9453 p-valor = <b>0.0886 &gt; 0.05</b>
NN-M2 n = 31	Para W = 0.9661 p-valor = <b>0.4197 &gt; 0.05</b>
N n=62	Para W = 0.9762 p-valor = <b>0.2713 &gt; 0.05</b>

Tabla 156. Prueba de Shapiro-Wilk para la fluidez de la muestra «1D» por nivel de lengua

Una vez comprobado con el test de Shapiro-Wilk que no podemos rechazar la  $H_0$  de ajuste a la normalidad, pues los p-valor son superiores al valor de significación de 0.05, recurrimos a la prueba paramétrica Anova para realizar la comparación de las tres muestras.

#### Anova Un Factor

Variable Respuesta:		tokens_M1-M2-N			
Variable Explicativa:		Grupos			
Número de Casos:		127			
	Suma de Cuadrados	G.L.	Cuadrado Medio	F-valor	p-valor
Entre Grupos	104052.1816	2	52026.0908	66.3781	<b>0.0003E-16</b>
Dentro Grupos	97189.2514	124	783.7843		
Total (corr.)	201241.4331	126			

Figura 64. Test Anova para la fluidez de las muestras NN-M1-1D, NN-M2-1D, N-1D

Como vemos en la figura superior, el p-valor=0.0003E-16<0.05 obliga a rechazar la  $H_0$  de igualdad al presentar un valor inferior al valor de significación de 0.05. Este resultado lo que nos indica es que hay diferencia, pero no entre qué grupos; por tanto, a continuación aplicamos el test de «comparaciones múltiples» para identificar dónde se producen las diferencias.

Vemos en la figura 65 que los grupos no presentan homogeneidad entre ellos, por tanto es normal que vayamos a encontrar diferencia significativa estadísticamente entre las distintas muestras. Comprobamos, en primer lugar, que hay diferencia significativa entre las muestras NN-M1\_F y NN-M2\_F, como ya hemos confirmado en respuesta a la pregunta cuatro de esta investigación. También hay diferencia significativa entre las dos muestras del grupo de NN y el de N. Vemos que se repite el mismo esquema que en la muestra «F», es decir, al igual que en la muestra «1D», aquí también podemos afirmar que la muestra del

grupo de N presenta mayor fluidez que las otras dos muestras. Por tanto, el grupo de NN está lejos de producir textos tan extensos como el de N.

**Anova Un Factor, Comparaciones Múltiples**

**Variable Respuesta:** tokens\_M1-M2-N  
**Variable Explicativa:** Grupos  
**Número de Casos:** 127

**Método:** LSD al 95.00%

Grupos	N	Media	Grupos Homogéneos
NN_M1_1D	34	91.0882	X
NN_M2_1D	31	115.1935	X
N_1D	62	157.1935	X

Contraste	Diferencia	+/- Límite
NN_M1_1D VS NN_M2_1D	*-24.1053	*13.7607
NN_M1_1D VS N_1D	*-66.1053	*11.8251
NN_M2_1D VS N_1D	*-42.0000	*12.1891

\* Diferencia estadísticamente significativa.

Figura 65. Test Anova-«comparaciones múltiples» para la fluidez de las muestras NN-M1-1D, NN-M2-1D, N-1D

**Fase 2b:** Diversidad léxica de la muestra «1D» por nivel de lengua: NN-M1, NN-M2, N.

Para analizar la diversidad léxica de la muestra «1D» en cada uno de los grupos, partimos de los datos de *tokens* y *types* que presentamos en la tabla 157.

Muestras «F»			
	NN-M1-1D (n=34)	NN-M2-1D (n=31)	N-1D (n=62)
<i>Tokens</i>	3.097	3.571	9.746
<i>Types</i>	2.248	2.572	6.607

Tabla 157. Número total de *tokens* y *types* de la muestra «1D» por nivel de lengua

Observamos en la tabla superior que es nuevamente la muestra del grupo de N la que presenta los valores más altos tanto en número en *tokens* como en *types*. En el gráfico siguiente podemos visualizar estas diferencias.

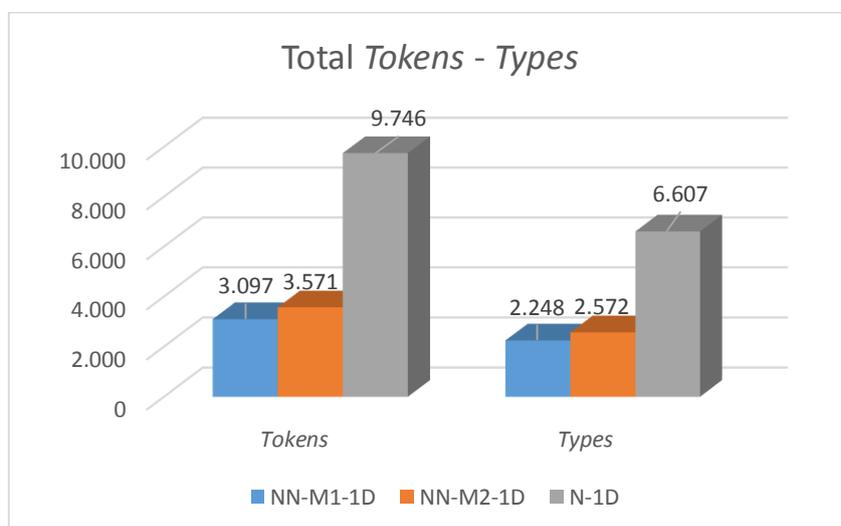


Gráfico 180. Totales de *Tokens* y *Types* de la muestra «1D» por nivel de lengua

A partir de estos valores, vamos a calcular el índice Uber U de cada una de las muestras para poder apreciar con más detalle la proporción de palabras diferentes que presenta cada grupo.

La fórmula del índice Uber U es:

$$U = \frac{(\log \text{tokens})^2}{\log \text{tokens} - \log \text{types}}$$

Una vez aplicada la fórmula, los resultados obtenidos son los siguientes.

Índice Uber U	
NN-M1_1D n = 34	0.14
NN-M2_1D n = 31	0.14
N_1D n=62	0.17

Tabla 158. Cálculo del índice Uber U para la muestra «1D» por nivel de lengua

Como podemos ver en la tabla 158, las dos muestras del grupo de NN registran el mismo índice Uber U; mientras que la muestra del grupo de N presenta un índice algo mayor. El siguiente paso para el análisis de estos datos es comprobar la bondad de ajuste a la normalidad de las muestras para decidir qué prueba estadística es la más adecuada. Así, pasamos a aplicar el test de Shapiro-Wilk.

Bondad d ajuste a la normalidad	
NN-M1 n = 34	Para W = 0.9788 p-valor = 0.7355 > <b>0.05</b>
NN-M2 n = 31	Para W = 0.9454 p-valor = 0.1168 > <b>0.05</b>
N n=62	Para W = 0.9491 p-valor = <b>0.0120 &lt; 0.05</b>

Tabla 159. Prueba de Shapiro-Wilk para la diversidad léxica de la muestra «1D» por nivel de lengua

El resultado de la prueba de Shapiro-Wilk indica que debemos utilizar una prueba estadística no paramétrica para el análisis de la densidad léxica, ya que el p-valor de la muestra del grupo de N es inferior al valor de significación de 0.05 (p-valor = 0.0120 < 0.05) y, por tanto, debemos rechazar la  $H_0$  de ajuste a la normalidad. Aunque las dos muestras del grupo de NN sí cumplen con la bondad de ajuste a la normalidad, siempre que haya una que no lo siga, se ha de recurrir a una prueba no paramétrica. Así, aplicaremos el test de Kruskal-Wallis para realizar la comparación.

#### Kruskal-Wallis

Variable Respuesta: Uber-U  
Variable Explicativa: Grupos  
Número de Casos: 127

Grupos	n	Suma de Rangos Rn	Rango Medio
M1_ID	34	1560.0000	45.8824
M2_ID	31	1617.0000	52.1613
N_ID	62	4951.0000	79.8548

Estadístico de Kruskal-Wallis (sin corrección por empates): 22.9507  
Estadístico de Kruskal-Wallis (con corrección por empates): 22.9512  
Grados de Libertad: 2  
p-valor: **0.0001E-1**

Figura 66. Test Kruskal-Wallis para la diversidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D

Comprobamos en la figura 66 que con el resultado que arroja el test de Kruskal-Wallis (p-valor = 0.0001E-1 < 0.05) debemos rechazar la  $H_0$  de igualdad. Es decir, que entre las muestras hay diferencia significativa estadísticamente. Pero para verificar entre qué muestras se produce la diferencia, tratamos los datos con el test de «comparaciones múltiples».

Los resultados que ofrece el test de «comparaciones múltiples» confirma la igualdad entre las dos muestras del grupo de NN. Ambas aparecen en la figura 67 marcadas como grupos homogéneos y no están señaladas con el asterisco que indica la diferencia significativa estadísticamente. Sin embargo, sí aparecen con asterisco, diferencia estadísticamente significativa, las muestras del grupo de NN en relación con la muestra del grupo de N.

Ante estos resultados, vemos, por un lado, que en las muestras del grupo de NN no se produce mejoría, como ya comprobamos en respuesta a la pregunta cuatro de la investigación. Por otro lado, comprobamos que los textos del grupo de NN están lejos de los del grupo de N en cuanto a diversidad léxica. Es decir, los nativos son capaces de escribir textos con mayor variedad de palabras que los no nativos.

**Kruskal-Wallis, Comparaciones Múltiples**

**Variable Respuesta:** Uber-U  
**Variable Explicativa:** Grupos  
**Número de Casos:** 127

**Método:** Dunn al 95.0%

Uber-U	N	Rango Medio	Grupos Homogéneos
M1_ID	34	45.8824	X
M2_ID	31	52.1613	X
N_ID	62	79.8548	X

Contraste	Diferencia	+/- Límite
M2_ID VS M1_ID	-6.2789	21.8811
N_ID VS M1_ID	*-33.9725	*18.8032
N_ID VS M2_ID	*-27.6935	*19.3819

\* Diferencia estadísticamente significativa.

Figura 67. Test Kruskal-Wallis-«comparaciones múltiples» para la diversidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D.

**Fase 2c:** Densidad léxica de la muestra «1D» por nivel de lengua: NN-M1, NN-M2, N.

En esta fase de análisis, vamos a comprobar si hay diferencia de densidad léxica entre las muestras del grupo de NN y el grupo de N. Para ello, vamos a tener en cuenta las palabras léxicas que presentan cada una de las muestras. Veamos lo totales y las medias correspondientes en la tabla siguiente.

Palabras léxicas		
	Totales	Media
<b>NN_M1_1D</b>	3.097	0.7902
<b>NN_M2_1D</b>	3.571	0.7950
<b>N_1D</b>	9.746	0.7321

Tabla 160. Totales y medias de PL de la muestra «1D» por nivel de lengua

Como podemos observar en la tabla, los valores de la media de PL están muy próximos entre sí, pero destaca el de la muestra de NN\_M2 (nivel de lengua final B1) sobre las otras dos muestras. Llama la atención, no obstante, el valor de la media de la muestra del grupo de N, pues es el más bajo de las tres muestras. A continuación mostramos los gráficos para visualizar estas diferencias.

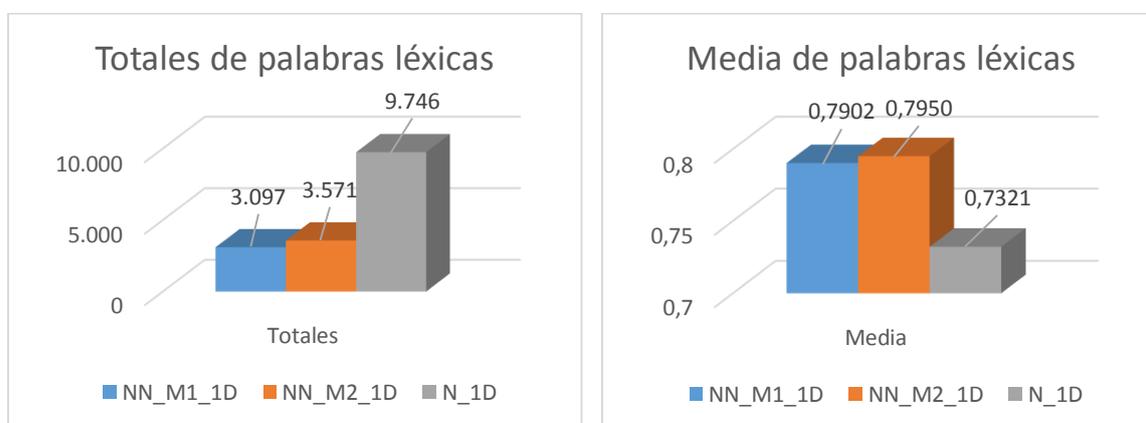


Gráfico 181. Totales y medias de PL de la muestra «1D» por nivel de lengua

Para verificar si las diferencias que apreciamos en los gráficos son estadísticamente significativas, primero vamos a aplicar el test de Shapiro-Wilk para comprobar la bondad de ajuste a la normalidad y decidir qué prueba estadística vamos a utilizar para la comparación.

Bondad d ajuste a la normalidad	
NN-M1 n = 34	Para W = 0.8723 p-valor = <b>0.0002E-5 &lt; 0.05</b>
NN-M2 n = 31	Para W = 0.8754 p-valor = <b>0.0009E-5 &lt; 0.05</b>
N n=62	Para W = 0.8625 p-valor = <b>0.0004E-10 &lt; 0.05</b>

Tabla 161. Prueba de Shapiro-Wilk para la densidad léxica de la muestra «1D» por nivel de lengua

El resultado que arroja el test de Shapiro-Wilk hace que debamos rechazar la  $H_0$  de ajuste a la normalidad, ya que todos los p-valor son inferiores al valor de significación 0.05. Por tanto, como son muestras que no siguen la bondad de ajuste a la normalidad, vamos a utilizar el test Kruskal-Wallis, una prueba no paramétrica para la comparación de más de dos muestras.

#### Kruskal-Wallis

Variable Respuesta: Med\_PL  
Variable Explicativa: Grupos  
Número de Casos: 508

Grupos	n	Suma de Rangos Rm	Rango Medio
PL_M1-ID	136	41904.5000	308.1213
PL_M2-ID	124	38764.5000	312.6169
PL_N_ID	248	48617.0000	196.0363

Estadístico de Kruskal-Wallis (sin corrección por empates): 76.9233  
Estadístico de Kruskal-Wallis (con corrección por empates): 76.9488  
Grados de Libertad: 2  
p-valor: **0.0000**

Figura 68. Test Kruskal-Wallis para la densidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D

Comprobamos en la figura superior que debemos rechazar la  $H_0$  de igualdad, pues el p-valor es inferior al valor de significación de 0.05 ( $p\text{-valor} = 0.0000 < 0.05$ ). Eso nos indica que hay diferencia significativa entre alguna de las muestras, pero para saber dónde están esas diferencias debemos aplicar el test de «comparaciones múltiples».

Observamos en el test de «comparaciones múltiples» (Figura 69) que aparecen como homogéneas las dos muestras del grupo de NN. Por tanto, las diferencias que hemos apreciado en los valores de la media ( $m=0.7902$  para NN\_M1;  $m=0.7950$  para NN\_M2) no

registran diferencia significativa estadísticamente. Eso quiere decir que estos dos niveles de lengua presentan una densidad léxica similar.

Como ya hemos señalado más arriba, la muestra del grupo de N sí registra diferencia significativa estadísticamente con el grupo NN\_M1 y con el grupo NN\_M2. Pero esa diferencia lo que nos indica es que las muestras del grupo de NN presentan mayor densidad léxica que la muestra del grupo de N, ya que la media de PL en esos grupos es superior.

#### Kruskal-Wallis, Comparaciones Múltiples

Variable Respuesta: Med\_PL  
 Variable Explicativa: Grupos  
 Número de Casos: 508

Método: Dunn al 95.0%

Med_PL	N	Rango Medio	Grupos Homogéneos
PL_N_1D	248	196.0363	X
PL_M1-1D	136	308.1213	X
PL_M2-1D	124	312.6169	X

Contraste	Diferencia	+/- Límite
PL_M2-1D VS PL_M1-1D	-4.4956	43.6270
PL_N_1D VS PL_M1-1D	*112.0850	*37.4903
PL_N_1D VS PL_M2-1D	*116.5806	*38.6441

\* Diferencia estadísticamente significativa.

Figura 69. Test Kruskal-Wallis-«comparaciones múltiples» para la densidad léxica de las muestras NN-M1-1D, NN-M2-1D, N-1D

#### Resumen de la pregunta cinco:

Para responder a la quinta pregunta de nuestro trabajo, hemos comparado entre sí las muestras que conforman nuestro corpus y que pertenecen a tres niveles de lengua, a saber: (i) nivel no nativo B1 inicial, (ii) nivel no nativo B1 final y (iii) nivel nativo. La finalidad de esta comparativa era ver si producciones escritas por sujetos no nativos estaban próximas a producciones de sujetos nativos en fluidez, diversidad léxica y densidad léxica. Como el

corpus está compuesto, además, por dos tipos de texto, «F» y «1D», hemos comparado estas muestras por separado para observar si los resultados se diferenciaban entre un tipo de texto y otro o si, por el contrario, se comportaban de forma similar.

En cuanto a **fluidez**, hemos de decir que los resultados han sido similares en ambas muestras, «F» y «1D». La media de *tokens* siempre ha ido de menos a más en relación a los tres niveles de lengua mencionados. Así, tras el tratamiento estadístico, hemos tenido que rechazar la  $H_0$  de igualdad entre las tres muestras (NN\_M1; NN\_M2 y N), ya que la prueba Anova y el test de «comparaciones múltiples» han confirmado que había diferencia estadísticamente significativa entre las tres. Es decir, de un estadio a otro, la fluidez en el texto escrito ha aumentado de forma significativa, siendo la muestra del grupo de nativos la que mayor fluidez presenta con una media de  $m=175.6545$  para la muestra «F» y  $m=157.1935$  para la muestra «1D», lejos de las medias de las muestras del grupo de NN (para la muestra «F»:  $m=128.0000$  y  $m=160.2927$ , NN\_M1 y NN\_M2, respectivamente; para la muestra «1D»:  $m=91.0882$  y  $m=115.1935$ , NN\_M1 y NN\_M2, respectivamente). Así, podemos afirmar que en este aspecto los nativos son capaces de escribir textos más extensos que los sujetos no nativos cuando se encuentran bajo las mismas condiciones, realizar la tarea en 30 minutos y dentro del aula.

El análisis de la **diversidad léxica** ha presentado resultados diferentes en la muestra «F» y «1D». Mientras que en la primera el test de Kruskal-Wallis y el test de «comparaciones múltiples» indicaban igualdad en las muestras de NN\_M2 y N, ya que ambos grupos presentaban un índice Uber  $U = 0.18$ , superior al  $U = 0.17$  del grupo NN\_M1; en la segunda muestra, «1D», la igualdad se daba entre las muestras del grupo de NN con un índice Uber  $U = 0.14$  para ambas y un índice  $U = 0.17$  para el grupo N. Así, al contrario que en la fluidez donde se producía un aumento progresivo por estadio, aquí hemos visto que no hay aumento progresivo de un estadio a otro, sino que en el caso de la muestra «F» hay igualdad de diversidad léxica entre el grupo NN\_M2 y el grupo N; y en el caso de la muestra «1D» también hay igualdad, pero entre las muestras del grupo NN.

Respecto a la **densidad léxica**, llama la atención que en los dos tipos de texto, «F» y «1D», el valor más bajo para la media de PL lo registra la muestra del grupo N. Así, hemos de decir que éste presenta menor densidad léxica, ya que el tratamiento estadístico con la prueba de Kruskal-Wallis y el test de «comparaciones múltiples» han señalado diferencia estadísticamente significativa entre el grupo N y NN\_M2, en la muestra «F»; y N y NN\_M1 y N y NN\_M2 en la muestra «1D».

## CAPÍTULO 7

### DISCUSIÓN

En este capítulo discutimos los resultados que hemos presentado y analizado en el capítulo anterior a partir de los datos obtenidos de las distintas muestras de nuestro corpus. Para llevar a cabo la discusión, vamos a seguir, igual que en el capítulo seis, el orden marcado por las preguntas de investigación agrupadas en tres secciones: (i) comparación del texto «F» y «1D»: preguntas uno y dos; (ii) comparación del grupo NN-M1 y M2: preguntas tres y cuatro; (iii) comparación del grupo N y del grupo NN: pregunta cinco.

#### 7.1 Comparación de las muestras del texto «F» y del texto «1D»

Aquí vamos a discutir los resultados correspondientes a la primera y segunda pregunta de investigación. En estas preguntas, nos centramos en los elementos léxicos que pueden caracterizar los dos géneros textuales que forman el corpus de nuestra investigación y vamos a discutir, asimismo, los resultados observados en el capítulo anterior sobre el comportamiento de los grupos N y NN ante las dos tareas que realizaron.

##### Pregunta 1:

¿Hay diferencias léxicas o sintácticas que distingan los relatos en pasado en los que se exige más creatividad (inventar una historia) frente a los que exigen cierta introspección (relatar una experiencia personal)?

---

Sí hay diferencia entre las producciones de los sujetos NN y los sujetos N. Pero estas diferencias varían dependiendo de si nos referimos a la fluidez, la diversidad léxica o la densidad léxica y dependiendo, asimismo, de si se trata de la muestra «F» o «1D».

---

Como ya hemos indicado en el capítulo anterior, las herramientas a las que hemos recurrido para caracterizar los dos tipos de textos con los que hemos trabajado son la lista de frecuencias, los *clusters* y las concordancias, que forman parte de la aplicación informática AntConc.

Tras analizar los datos con estas herramientas y contrastar las dos muestras de nuestro corpus, «F» y «1D», con el corpus CREA, constatamos:

- (i) que existe coincidencia con el corpus CREA y con los trabajos de Muñoz Lobo<sup>30</sup> (2012) y el de Cantos y Sánchez<sup>31</sup> (2011) en que las primeras palabras de la lista de frecuencias (*WordList*) corresponden a palabras funcionales.
- (ii) que el listado de frecuencias con la lista de *stopwords* ha proporcionado una primera caracterización del tipo de texto de la muestra «1D». Las palabras léxicas que han aparecido en el nuevo listado de frecuencias remitían claramente al campo semántico de la educación. Cruz (2012) ya señalaba este elemento como una ventaja de trabajar con este tipo de herramienta. Sin embargo, como vimos en el apartado de ‘resultados’, con la muestra «F» no fue posible caracterizar ante qué tipo de texto estábamos, pues las palabras léxicas del listado de frecuencias con *stopwords* no proporcionaban un hilo conductor para extraer el tema.
- (iii) que, por un lado, las herramientas ‘*clusters*’ y ‘concordancias’ han ayudado a caracterizar y mostrar las diferencias de los dos tipos de muestras. Mediante el análisis de las palabras seleccionadas hemos comprobado que la muestra «1D» se caracteriza por una presencia mayor de agrupamientos y concordancias en las que domina la función expresiva del lenguaje frente a la referencial de la muestra «F». Por otro lado, en el ya citado trabajo de Muñoz Lobo (*op.cit.*), que también analiza agrupamientos, hay que mencionar uno de los agrupamientos que analiza, el de la palabra ‘tiempo’. En su trabajo indica que seleccionó esta palabra léxica, que aun no siendo la primera de la lista de frecuencias, sí es la que más combinaciones ofrece. En nuestra investigación también analizamos los agrupamientos y las concordancias de esta misma palabra, ya que fue una de las que coincidían en las dos muestras, «F» y «1D», dentro de las cincuenta primeras palabras de la lista de frecuencias con *stopwords*. Constatamos, como Muñoz Lobo (*op.cit.*), que los agrupamientos y las concordancias con esta palabra léxica son abundantes y ricos.  
  
Otro trabajo a destacar que también estudia agrupamientos es el de las autoras Verspoor y Smiskova (2012). Ellas trabajan con lo que llaman

---

<sup>30</sup> Su trabajo se basa en dos corpus del español, el CREA y el corpus de la ‘Universidad de Brigham Young’.

<sup>31</sup> Analizan el corpus ‘Cumbre’ para el español y el corpus ‘Lacell’ para el inglés.

‘*chunks*’ (secuencias cortas de palabras, de fórmulas). Pero mientras ellas recurren a una selección de ‘*chunks*’ para indagar si hay progreso en la producción escrita, en nuestra investigación la finalidad era identificar los agrupamientos más frecuentes para caracterizar los dos tipos de muestras con las que hemos trabajado.

En esta línea, también encontramos coincidencias con el estudio de van Esch *et al.* (2006) en el que analizan la calidad discursiva de textos argumentativos, por un lado para averiguar si se producía progreso entre el primer año de estudio y el segundo y, por otro lado, para extraer las características del género argumentativo y comprobar si los textos de los sujetos no nativos se aproximan a los de los nativos. Vemos, así, que es relevante identificar las características principales de los géneros textuales para saber qué hacen los nativos y no nativos.

**Pregunta 2:**

**Entre los sujetos no nativos y los sujetos nativos que participaron en la investigación, ¿hay diferencias léxicas o sintácticas que los distingan?**

---

Sí hay diferencia léxica en el campo semántico de la expresión de sentimientos y en las expresiones para cuantificar. Los sujetos nativos presentan una mayor diversidad léxica en esos dos aspectos. Sin embargo, en líneas generales, hemos comprobado una similitud tanto en agrupamientos como en concordancias en los dos grupos.

---

Para indagar si entre el grupo N y NN había diferencias en el léxico o estructuras utilizadas en la producción de las dos tareas, hemos recurrido, igual que en la pregunta uno, a las herramientas de la lista de frecuencias, los *clusters* y las concordancias de la aplicación informática AntConc.

Tras analizar los datos con estas herramientas y contrastar las dos muestras de nuestro corpus, «F» y «1D», con el corpus CREA, destacamos:

- (i) la igualdad con la pregunta uno de la presencia de palabras funcionales en las primeras posiciones de la lista de frecuencias y las coincidencias con los trabajos allí citados.

- (ii) la semejanza en los agrupamientos y concordancias de las palabras seleccionadas en las dos muestras analizadas (grupo N y NN). En esta línea, encontramos trabajos que también han comparado las producciones escritas de nativos y no nativos, aunque hayan analizado otros aspectos de la competencia escrita. Entre esos trabajos, resaltamos los de Teijeira *et al.* (2005), de Haan y van Esch (2005), Díaz y Lucha (2010; 2012). En todas estas investigaciones observaron que los resultados de los sujetos no nativos no mostraban diferencias significativas respecto de los nativos. Así, por ejemplo, el trabajo de Teijeira *et al.* (*op. cit.*), investigación de metodología cualitativa, distinguen la similitud en la manera de organizar la argumentación los sujetos no nativos en comparación con los nativos. Asimismo, resaltan el hecho de que los sujetos no nativos, aun estando en un nivel no muy avanzado de sus estudios, son capaces de escribir textos coherentes. En este punto, podemos relacionar los resultados de nuestra investigación con los de Teijeira *et al.* (*op. cit.*), ya que los sujetos no nativos que participaron en nuestra investigación, aun teniendo un nivel B1 según el MCER, muestran resultados próximos a los sujetos nativos. También, como hemos indicado, Díaz y Lucha (*op. cit.*) señalan similitudes entre los resultados de nativos y no nativos; pero en este caso, las autoras hacen referencia a que esa igualdad se produce en estudiantes que no han recibido una instrucción específica, punto coincidente con los estudiantes no nativos de nuestra investigación que asistían a un curso general de español.
- (iii) las diferencias encontradas en el léxico para expresar sentimientos o cuantificación. En este aspecto el grupo N presenta una mayor diversidad léxica que el grupo NN. También es destacable un cierto predominio de la función expresiva del lenguaje en la muestra del grupo N frente a la del grupo NN. Los trabajos de Díaz *et al.* (2001; 2003), estudiando la calidad del texto y la proximidad entre las producciones escritas de nativos y no nativos, también encontraron diferencias significativas que alejaban el texto no nativo del modelo del nativo.

## 7.2 Comparación del grupo NN- M1 y M2

Aquí vamos a discutir los resultados correspondientes a la tercera y cuarta pregunta de investigación. En estas preguntas, como en la quinta, el análisis que hemos realizado es de carácter cuantitativo y pseudolongitudinal.

### Pregunta 3:

Tras un período de instrucción de 120 horas siguiendo el enfoque comunicativo y una enseñanza mediante tareas, ¿se observa progreso en cuanto a fluidez, diversidad léxica y densidad léxica en la producción escrita de los sujetos no nativos participantes de la investigación entre el M1 (inicio del curso) y el M2 (final del curso)?

---

Sí hay diferencia en las variables de fluidez y diversidad léxica, pero no en la variable densidad léxica. Hemos constatado que tras un período de instrucción de 120 horas, los estudiantes son capaces de escribir textos más extensos y con más diversidad léxica.

---

Las medidas de fluidez, diversidad léxica y densidad léxica que hemos seleccionado para evaluar la calidad general del texto (Wolfe-Quintero et al., 1998, Celaya *et al.*, 2001, Polio, 2001, Ortega, 2004, Torras *et al.*, 2006) son medidas cuantitativas que otras investigaciones de carácter también cuantitativo y longitudinal o pseudolongitudinal han utilizado en los estudios realizados en el ámbito de la Adquisición de Segundas Lenguas. El análisis cuantitativo, como ya hemos visto, es más habitual en el campo del inglés como L2/LE que en el del español L2/LE, en el que los estudios cualitativos son más recurrentes. De ahí nuestro interés en acercarnos al análisis de la producción escrita por aprendices de español como L2 desde una perspectiva cuantitativa.

Como hemos visto en el capítulo uno, el estado de la cuestión, y en el capítulo tres, marco metodológico, hay una mayor tradición de investigaciones cualitativas que cuantitativas. Igualmente, existe un predominio de estudios transversales frente a longitudinales o pseudolongitudinales. De esa ausencia de este tipo de trabajos, como ya hemos indicado más arriba, nace nuestro interés en presentar aquí una investigación cuantitativa y pseudolongitudinal.

Los datos que hemos analizado en el capítulo seis han reportado los siguientes resultados:

- (i) Los valores tanto en **fluidez** - considerada como el número total de palabras escritas en un período de tiempo determinado (en nuestra investigación fueron 30 minutos)- como en **diversidad léxica** en el M2 son superiores a los del M1. Recordemos que la medida utilizada para valorar la **diversidad léxica** -que pone en relación el número total de palabras (*tokens*) con el número total de palabras diferentes (*types*) presentes en un texto- fue *Uber U* (Dugast, 1978; Jarvis, 2002). Con esta fórmula queríamos evitar las posibles anomalías en los resultados que en casos de textos de extensión diferente (como en nuestra investigación) la tradicional medida de la ratio *types/tokens* (RTT) puede originar.

Las pruebas estadísticas utilizadas para comprobar si la diferencia observada era significativa demostraron que sí era estadísticamente significativa. Por tanto, estos datos corroboran los resultados de otros trabajos longitudinales anteriores que constataron también mejoría en las producciones escritas. Así, por ejemplo, los trabajos de Martínez Arbeláiz (2004), De Haan y van Esch (2005), van Esch *et al.* (2006), Celaya y Navés (2009) y Baerlocher (2013), siguiendo una metodología cuantitativa, analizaron también la progresión en fluidez y los resultados en todos ellos verificaron un aumento significativo estadísticamente. Es decir, que los alumnos fueron capaces de escribir un número mayor de palabras después de un período de instrucción; constatando de esta manera que a mayor número de horas de instrucción, mejores resultados en fluidez escrita.

Otras investigaciones longitudinales, pero de cariz más cualitativo y que focalizaron en otros aspectos y no tanto en la calidad del texto (Mancilla *et al.*, 2015; Sasaki, 2009) también observaron mejoría en las producciones escritas. Sasaki (*op. cit.*), que centra su trabajo en los estudiantes con estancia en el extranjero (*Stay Abroad SA*), remarca el hecho de que tanto los estudiantes con estancia en el extranjero como los que permanecieron en su universidad (en Japón) mejoraron el nivel general de la L2, pero “the only the SA group

improved their L2 writing ability” (Sasaki, 2009: 52).

También nos ayuda a entender este progreso el trabajo, también longitudinal, de Schoonen *et al.* (2009) que remarcan la interrelación entre los recursos cognitivos y los lingüísticos, permitiendo retomar las experiencias previas para mejorar su actuación en situaciones posteriores. En el caso de nuestra investigación, los sujetos pudieron recurrir a dichos recursos para mejorar su actuación en la realización de la segunda tarea.

- (ii) No hay diferencia estadísticamente significativa en la variable **densidad léxica**. Los resultados revelaron que las producciones escritas tanto en el M1 como en el M2 eran densas léxicamente, pues en ambos momentos la proporción de las PL era superior a la de PF. No obstante, la diferencia observada entre la muestra de la primera tarea y la segunda no hay diferencia. Esto lo podemos atribuir al hecho de que en un período corto como es un curso general de 120 horas es insuficiente para mejorar en este aspecto. También vemos que estos resultados se relacionan con los discutidos en las preguntas uno y dos, más centradas en el análisis del comportamiento del léxico. Allí hemos visto que no se dan grandes diferencias entre las distintas muestras.

**Pregunta 4:**

¿Hay diferencia en los resultados de fluidez, diversidad léxica y densidad léxica en aquellas producciones que se trabajaron siguiendo una enseñanza mediante tareas (inventar una historia) frente a las que no (explicar una experiencia personal)?

---

Sí hay diferencia en las variables de fluidez y diversidad léxica. La muestra «F» presenta valores más altos que la muestra «1D». Sin embargo, en densidad léxica las pruebas estadísticas reportaron resultados contradictorios.

---

El interés en saber si hay diferencia en la fluidez, diversidad léxica y densidad léxica entre la muestra «F», trabajada en el aula siguiendo una enseñanza mediante tareas, y la muestra «1D», a la que no se le dedicó un tratamiento específico, radica en discernir hasta qué punto una enseñanza mediante tareas puede influir en una mejor calidad del texto escrito.

Así, para responder a esta inquietud, (i) en primer lugar valoramos la calidad de las muestras «F» y «1D» por separado en el M1 y el M2 y (ii) después comparamos los resultados del M1 y del M2 para ver si se había producido progreso.

Al analizar las muestras «F» y «1D», comprobamos que tanto en el M1 como en el M2, la muestra «F» reportó mejores resultados en fluidez, diversidad léxica y densidad léxica.

Pero al comparar si se había producido progreso en las producciones escritas de las dos muestras, observamos resultados interesantes. El tratamiento estadístico utilizado mostró:

- (i) que se había producido progreso en cuanto a la **fluidez** tanto en la muestra «F» como en la «1D», y esa diferencia registrada se comprobó que era estadísticamente significativa. Ese aumento, tras una instrucción de 120 horas, lo podemos relacionar con resultados similares de otros trabajos longitudinales como los de Martínez Arbelaiz (2004), De Haan y van Esch (2005), van Esch *et al.* (2006) y Baerlocher (2013), en los que se constató que a mayor número de horas de instrucción, mejores resultados en fluidez escrita. Es decir, los alumnos fueron capaces de escribir un mayor número de palabras en el segundo momento de recogida de datos cumpliendo con las mismas instrucciones y en el mismo contexto que en el M1, al inicio del curso.
  
- (ii) que en cuanto a la **diversidad léxica** hubo progreso en la muestra «F», pero no así en la muestra «1D». También se verificó que la diferencia entre el M1 y el M2 en la muestra «F» era estadísticamente significativa. Igual que en el caso de la variable fluidez, estos resultados se pueden atribuir al número de horas de instrucción, como en los trabajos ya mencionados. Pero, además, hay otro factor en este caso que hay que resaltar. Como vemos, es la muestra «F» frente a la «1D» la que presenta progreso significativo. Es decir, el texto trabajado en el aula (inventar una historia) siguiendo una enseñanza mediante tareas es el que muestra mejores resultados. Trabajos como los de Díaz y Lucha (2010, 2012) y la investigación de Verspoor y Smiskova (2012), también realizaron tratamientos diferenciados en dos grupos. En Díaz y Lucha (*op. cit.*) se llevó a cabo una enseñanza mediante técnicas de FoF (Focus on Form) en uno de los

grupos, mientras que Verspoor y Smiskova (*op. cit.*) trabajaron de forma intensiva una mayor aportación de *input* al aula en uno de los grupos. En nuestro caso el tratamiento diferenciado no ha sido con dos grupos, sino con el tipo de texto que se trabajó. No obstante, creemos que existen vínculos entre estas investigaciones.

Los resultados que obtuvieron Díaz y Lucha (*op. cit.*) en ambos estudios avalan que el grupo que recibió el tratamiento específico mejoró su rendimiento y obtuvo mejores resultados. Nuestra investigación coincide con esos estudios verificando, así, que un tratamiento instruccional específico ayuda a progresar en la producción escrita. En cambio, los resultados de Verspoor y Smiskova (*op. cit.*) no mostraron progreso significativo entre el grupo que recibió el tratamiento instruccional con mayor aportación de *input* que el que siguió la enseñanza habitual.

- (iii) que la muestra «F» registra mayor **densidad léxica** que la muestra «1D» tanto en el M1 como en el M2. Sin embargo, al aplicar la prueba no paramétrica Kruskal-Wallis para la comparación de más de dos muestras (comparábamos las muestras «F» «1D» en el M1 y el M2, cuatro muestras en total), el resultado que arrojó fue que no había diferencias significativas. Ante este resultado, encontramos coincidencias, nuevamente, con el trabajo de Verspoor y Smiskova (*op. cit.*) que tampoco encontraron diferencias significativas entre el grupo que recibió mayor *input* respecto del que recibió una cantidad menor. Pero como este resultado nos llamó la atención, por no ser el esperado, decidimos aplicar la prueba no paramétrica de Mann-Whitney (específica para la comparación de dos muestras) y ésta indicó que sí había diferencia significativa estadísticamente, con lo que estos resultados estarían más próximos a los trabajos de Díaz y Lucha (*op. cit.*). No obstante, ante estos resultados opuestos, deberíamos realizar más pruebas para confirmar si hay o no diferencias significativas.

Como acabamos de ver, la muestra «F» es la que presenta valores más altos en las tres variables tanto en el M1 como en el M2. Como hemos indicado, el progreso que se produce en fluidez y diversidad léxica en el M2 sí podríamos considerar que es la enseñanza mediante tareas el catalizador del mejor resultado en las producciones escritas de la muestra «F», pues

fue este tipo de texto el que se trabajó en el aula siguiendo la enseñanza mediante tareas. Sin embargo, la mejor calidad del texto de la muestra «F» en el M1 no se puede decir que sea la consecuencia del número de horas de estudio de la lengua ni del tratamiento mediante tareas. Así, surge la duda de si esta diferencia entre la muestra «F» y la «1D» en el M2 se debe realmente a la instrucción o es más por motivos de tipología textual. Es decir, que los mejores resultados en la muestra de textos en la que debían inventar una historia resulta más fácil, pues no hay implicaciones personales, que escribir sobre una experiencia personal en la que es necesario realizar una introspección. También es interesante traer a colación las opiniones de los sujetos nativos que colaboraron en la investigación. Recordemos que estos realizaron las dos tareas de forma consecutiva en un mismo día y que en primer lugar se les administró la prueba «1D» por considerar que les resultaría más fácil escribir sobre algo que ellos mismos habían vivido. La sorpresa fue que la gran mayoría comentó que les era difícil recordar y explicar cómo había sido su primer día de clase de un segundo idioma.

### 7.3 Comparación del grupo NN y el grupo N: fluidez, diversidad léxica y densidad léxica en las muestras «F» y «1D»

Aquí vamos a discutir los resultados correspondientes a la quinta pregunta de investigación.

**Pregunta 5:**

En las producciones escritas de los sujetos no nativos y los sujetos nativos, ¿hay diferencia en cuanto a fluidez, diversidad léxica y densidad léxica en las muestras «F» y «1D»?

---

Sí hay diferencia entre las producciones de los sujetos NN y los sujetos N. Pero estas diferencias varían dependiendo de si nos referimos a la fluidez, la diversidad léxica o la densidad léxica y dependiendo, asimismo, de si se trata de la muestra «F» o «1D».

---

El objetivo de medir y comparar la calidad general del texto entre sujetos no nativos y nativos radica en el interés en llegar a conocer hasta qué punto hay o no diferencias en el desarrollo de la escritura en los distintos estadios de nivel de lengua que hemos tomado en consideración, a saber: nivel B1 inicial, nivel B1 final y nivel nativo; e indagar, también, si hay proximidad en fluidez, diversidad léxica y densidad léxica entre producciones escritas de no nativos y nativos. Al mismo tiempo, también pensamos que era relevante investigar si había diferencias en la calidad general del texto según su tipología (más creativa, muestra, «F», o más personal, más introspectiva, muestra «1D»). Consideramos que discernir entre estas cuestiones puede ayudar al docente como al aprendiz de ELE a tomar conciencia del nivel de lengua que posee realmente y focalizar en aquellos aspectos que sean necesarios reforzar.

Los resultados obtenidos en la variable **fluidez** textual reportaron diferencias entre las tres muestras analizadas. El tratamiento estadístico realizado a posteriori mediante las pruebas Anova y «comparaciones múltiples» nos ayudó a dilucidar si esas diferencias eran significativas estadísticamente. Los resultados permitieron constatar:

- (i) que había una diferencia estadísticamente significativa en el progreso que se produce entre el M1 y el M2 del grupo NN tanto en la muestra «F» como «1D».

Este aumento, tras una instrucción de 120 horas siguiendo un enfoque comunicativo y una enseñanza mediante tareas, viene a corroborar los resultados de otros trabajos longitudinales como los de Martínez Arbelaz (2004), De Haan y van Esch (2005), van Esch *et al.* (2006) y Baerlocher (2013). En todos ellos se produjo también una progresión clara en la fluidez de los textos. Es decir, que los alumnos fueron capaces de escribir un número mayor de palabras en el segundo momento de recogida de datos; constatando de esta manera que a mayor número de horas de instrucción, mejores resultados en fluidez escrita.

- (ii) que la superioridad en la media de *tokens* del grupo N respecto del grupo NN (M1 y M2) en los dos tipos de texto analizados marca también una diferencia estadísticamente significativa. Estos datos confirman los encontrados por Díaz *et al.* (2001; 2003), Díaz y Lucha (2010; 2012) en sus estudios. En ellos, las producciones escritas de los nativos reflejaron también valores más altos de fluidez. Llama la atención, sin embargo, el trabajo de De Haan y van Esch (*op. cit.*) que señalan que las producciones de los no nativos del segundo año registraron valores próximos o incluso, en algunos casos, mejores a los de los nativos. Esto enlaza y se justifica con la teoría de sistemas dinámicos (Verspoor y Smiskova, 2012) que defiende que cada individuo es diferente y que, por tanto, el desarrollo de la L2/LE, en este caso en el campo de la escritura, va a seguir un ritmo distinto, pudiendo llegar a superar los niveles esperados para cada estadio. En nuestra investigación decidimos tomar los datos en conjunto y no focalizar en individuos, pero observamos, como De Haan y van Esch (*op. cit.*) casos de sujetos no nativos que presentaban resultados en fluidez superiores a algunos de los sujetos nativos. Estas diferencias se pueden observar en el análisis descriptivo pormenorizado del comportamiento de las distintas categorías morfológicas analizadas (ver apartado 6.1).

Ante estos resultados podemos afirmar que el grupo de sujetos N es capaz de escribir textos más extensos que el grupo de sujetos NN. Esto se entiende en la medida que la fluidez mejora en función de las horas de instrucción, como hemos visto claramente con el grupo de sujetos NN. De ahí que en los distintos estadios de lengua (nivel no nativo B1 inicial; nivel no nativo B1 final, y nivel nativo) hayamos comprobado el aumento progresivo de la fluidez

textual. Asimismo, hemos verificado que la distancia en fluidez entre sujetos no nativos y nativos está lejos de poder considerarse próxima como en el estudio de De Haan y van Esch (*op. cit.*).

Los resultados que arrojó la medida Uber U mostraron comportamientos diferentes según la tipología textual y el grupo de sujetos. Con la prueba estadística Kruskal-Wallis comprobamos:

- (i) que en la muestra «F» había diferencia significativa estadísticamente entre los grupos NN-M1 y NN-M2 (como ya hemos comentado más arriba), por un lado y entre NN-M1 y N, por otro lado; pero no entre el grupo NN-M2 y el grupo de sujetos nativos. Este resultado de igualdad en la diversidad léxica entre el grupo más avanzado de no nativos y el grupo de nativos remite a los resultados que Díaz y Lucha (2010) encontraron en su investigación entre nativos y no nativos. Las autoras observaron que el grupo de no nativos que recibió instrucción con técnicas de FoF obtuvieron resultados próximos al grupo nativo, concluyendo así que no es la prerrogativa de ser ‘nativo’, sino la instrucción, en ocasiones, la que otorga la capacidad de alcanzar mejores resultados. En este caso podemos decir que nuestra investigación avala dichos resultados, pues recordemos que el grupo de no nativos siguió una enseñanza mediante tareas en la que se focalizó en la tipología textual del modelo «F». También hay que hacer referencia, nuevamente, al estudio de De Haan y van Esch (*op. cit.*); como ya hemos señalado en referencia a la fluidez, estos autores encontraron datos iguales o superiores en los resultados de sujetos no nativos respecto a los nativos. Aunque no se trata de la misma variable, no deja de ser llamativo que se dé este fenómeno en dos tipos de investigación que, en este caso, no analizan los mismos elementos. También resulta interesante resaltar aquí el trabajo de Mancilla *et al.* (2015). Recordemos que los autores llevaron a cabo una investigación longitudinal en la que hicieron el seguimiento de estudiantes universitarios nativos y no nativos de inglés. Su estudio se centró principalmente en el análisis de la complejidad sintáctica y una de sus conclusiones, y que podemos relacionar con nuestros resultados, es que no se produjeron diferencias significativas entre los estudiantes no nativos con mayor competencia y los nativos.

Como vemos, son varios los estudios que coinciden en no encontrar diferencias

significativas, aunque en diferentes aspectos, entre la producción escrita de sujetos no nativos de mayor competencia en comparación con los de menor competencia y los sujetos nativos.

- (ii) que en la muestra «1D» hay diferencia significativa estadísticamente entre el grupo N y el grupo NN (M1 y M2). Los resultados mostraron mayor diversidad léxica en el grupo de sujetos nativos. Resultado que esperábamos, pues partíamos de la premisa de que los sujetos nativos deberían presentar mayor diversidad léxica que los sujetos no nativos. Pero una vez analizados los datos de la muestra «F» y contrastarlos con los del trabajo de Díaz y Lucha (*op. cit.*), constatamos que la instrucción puede influir de manera decisiva en un desarrollo de la competencia comunicativa del estudiante no nativo.

En cuanto a la **densidad léxica**, retomando los resultados del capítulo anterior, llama la atención, por un lado, que la muestra «F» registra valores algo más altos en la media de PL que la muestra «1D», tanto para el grupo N como para el grupo NN. Es decir, que la producción escrita que requiere más creatividad por parte de los sujetos (inventar una historia) presenta un valor de media mayor de palabras que aportan información al texto. Por otro lado, hemos comprobado que las muestras del grupo N son las que presentan el valor más bajo de PL en los dos tipos de muestras. Este dato lo confirmamos al revisar el apartado dedicado a la estadística descriptiva (6.1.4; 6.1.5) en la que comprobamos que la media de palabras funcionales (pronombres, preposiciones, artículos y conjunciones) por parte del grupo N es más alta, tanto en la muestra «F» como «1D», que en las muestras del grupo NN\_M1 y NN\_M2. Con estos datos y tras realizar el tratamiento estadístico con la prueba Kruskal-Wallis observamos:

- (i) que en la muestra «F» había diferencia significativa entre el grupo N y el grupo NN-M2, pero no entre el grupo N y el NN-M1. Es decir, que el índice de densidad léxica (palabras que aportan contenido semántico al texto) es similar en el grupo de no nativos con menor competencia y el grupo de nativos.
- (ii) que en la muestra «1D» hay diferencia estadísticamente significativa entre el grupo N y el grupo NN (M1 y M2).

La mayor presencia de PF (palabras funcionales) en el grupo N nos hace pensar que las producciones escritas por los sujetos nativos son más complejas sintácticamente, ya que

la finalidad de las PF es ayudar a estructurar el texto mediante los elementos de relación como las preposiciones o como las conjunciones, de coordinación o subordinación. No obstante, el hecho de que no haya diferencia estadísticamente significativa en la muestra «F» entre el grupo NN-M1 y el grupo N no quiere decir que los textos de la muestra de los no nativos M1 sean más complejos sintácticamente. Si vamos al capítulo de resultados comprobamos que el grupo NN-M1 es el que presenta una media mayor de PL. Por tanto, no podemos concluir que estamos ante textos complejos sintácticamente.



## CAPÍTULO 8

### CONCLUSIONES

Como decíamos al principio, el campo de investigación sobre la expresión escrita en una lengua extranjera, si bien tiene una aparente solera y tradición, no es realmente así. Es cierto que los productos escritos en una L2 (fueran de traducción, composición, etc.) han sido objeto de estudio y análisis en el marco de la lingüística contrastiva y el análisis de errores, por ejemplo. Pero ya desde los 70 se llamó la atención sobre la ausencia de estudios centrados en el carácter comunicativo del uso de la lengua, la ausencia de estudios centrados en las características textuales propias o la inexistencia de estudios sobre procesos de escritura y procesos seguidos por los escritores en L1 y L2 en sí. Hemos revisado en el capítulo 1, en el estado de la cuestión algunos de los hitos que han centrado y puesto al día la investigación tanto desde la perspectiva de la enseñanza como desde la adquisición de una lengua extranjera.

En los estudios sobre procesos de escritura, ha resultado muy fértil la aplicación de metodologías de enseñanza como el enfoque por tareas, que han permitido superar la artificialidad comunicativa de la mayoría de actividades de producción escrita en metodologías anteriores. El uso de técnicas como el '*focus on form*', por ejemplo, han permitido llevar a cabo interesantes estudios de investigación-acción que han arrojado mucha luz sobre los procesos de escritura y las posibilidades de mejora de la competencia con técnicas procedentes de la psicolingüística aplicada a la enseñanza de lenguas. De ello nos ocupamos en una investigación previa (nuestro DEA) y en sucesivas publicaciones (de 2006 a 2015)<sup>32</sup>. En ellas nos hemos ocupado de destacar cómo el reto ha estado, pues, en plantear tareas concretas y evaluar los componentes lingüísticos (exponentes lingüísticos) no desde la perspectiva de los errores, sino dentro de la del desarrollo de la L2 (la perspectiva de la adquisición) y atendiendo a lo que sostenemos que es la competencia: una suma de subcompetencias (léxica, sintáctica, discursiva), que son las que se han podido medir en el trabajo efectuado mediante los indicadores seleccionados y con las herramientas elegidas para cada caso. Nos referimos, efectivamente, a la competencia léxica (y a la diversidad léxica y la densidad léxica esperables en un determinado subgénero y nivel de conocimiento de la L2), a la sintáctica (y a la previsión de estructuras y componentes por subgénero y

---

<sup>32</sup> Lucha, 2006; Díaz y Lucha, 2007, 2010, 2012, 2015; Baerlocher y Lucha, 2008; Lucha y Baerlocher, 2008; Lucha y Baerlocher, 2012

nivel), a la discursiva y a la retórica (o a los aspectos de coherencia, cohesión, rastreables en el uso de párrafos, marcadores, conectividad y anáforas y deícticos, por ejemplo). En cuanto a las herramientas, hemos adoptado la perspectiva práctica del profesor-investigador y hemos elegido las que se encuentran disponibles en la red, de libre acceso, cambiando y combinando recursos y ámbito de aplicación en un ejercicio de poner puertas al campo y obtener una imagen de la destreza objeto de estudio en sus dos momentos de cata.

En este sentido, creemos que ha sido novedoso aproximarse a la destreza escrita en español L2 y a la competencia escritora en un contexto no académico (ni universitario ni de escolarización) desde esta perspectiva y, además, hacer de ello un estudio longitudinal con el que ver, precisamente, el desarrollo de esta competencia como un componente más en un curso de lengua extranjera. Es, insistimos, particularmente novedoso abordar el estudio de este desarrollo **no en un curso de escritura**, ni un taller de escritura (que es donde se sitúan principalmente los estudios más conocidos y citados sobre el inglés L2: Polio 1991, Manchón 1998, 2007, 2008, 2009, 20014). Creemos que haber elegido como contexto un curso de carácter general resulta particularmente revelador para ver cómo se desarrolla la competencia escritora ‘no experta’ de alumnos de L2 y compararla con la producción de nativos que tampoco son expertos.

Es importante a estas alturas retomar cuál fue el objetivo inicial, el motor de este estudio, que no es específicamente un trabajo de adquisición de español L2. O, mejor dicho, cuál fue el motor de la investigación sobre escritura en L2 que emprendimos desde el ya lejano trabajo de investigación que plasmamos en el DEA. Este objetivo no ha sido otro que aplicar a la escritura del español L2 la lupa de análisis y caracterización basada en el estudio de los géneros escritos producidos por aprendices B1 y B2, el descubrimiento de sus rasgos idiosincrásicos, de sus coincidencias con las estructuras y estrategias nativas y de sus divergencias con el fin de que, una vez conocidas, se pueda abordar su trabajo sistemático y enseñanza, desde el conocimiento empírico de en qué consiste su variación.

Si en el DEA nos ocupamos de la escritura de textos de carácter específico en aprendices de B2, universitarios multilingües, identificando su especificidad y trabajándola a través de las técnicas de ‘*focus on form*’, en la presente tesis nos hemos atrevido con alumnos de carácter generalista, en un entorno como el de las Escuelas Oficiales de Idiomas y en un curso de lengua también general. Hemos seleccionado, también, dos (sub)géneros textuales distintos entre sí: el relato (que sí ha sido largamente objeto de estudio en contexto de L2 inglés, si bien no tanto para ELE) y la introspección personal valorativa. La apuesta

por identificar las diferencias entre géneros en el marco de un curso general de B1 era arriesgada. No obstante, ambas situaciones de escritura aparecen recogidas para ese nivel en el MCER, si bien no como productos escritos en un ámbito público, puesto que el segundo está más claramente dentro del ámbito personal y entrañaba una dificultad mayor. A lo largo del estudio hemos apuntado cómo la implicación emocional y el manejo de la memoria por parte de todas las poblaciones de sujetos (incluidos los nativos de control) han puesto de manifiesto la dificultad de la tarea. Cabe decir, no obstante, que después del estudio podemos afirmar que constituye un instrumento idóneo para calibrar y detectar al buen escritor tanto en L1 como en L2, puesto que es la prueba donde más dificultad (y peores resultados) se observa.

Centrándonos más concretamente en el estudio llevado a cabo en esta tesis, hemos destacado la importancia que tuvo para nosotros desde el principio el disponer de un volumen de datos cuantitativamente importante, para superar los problemas de la poca representatividad estadística que enfrentamos en la investigación anterior (DEA). Por esta razón, contamos con dos grupos en dos períodos de recogida distintos, cuya producción se obtuvo en dos momentos: al inicio y al fin del curso. En algunos casos, como se detalla en el capítulo de metodología, fue posible contar con dos recogidas de algunos individuos, si bien la técnica empleada ha consistido en una recogida de datos ‘falsos longitudinales’, compensando el no mantenimiento del mismo grupo en el tiempo con dos recogidas transversales de grupos similares (equivalentes en nivel, verificado mediante test de nivel y equivalentes estadísticamente mediante pruebas estadísticas para seleccionar solo lo verdaderamente comparable). El hecho de contar con dos pruebas de elicitación, de dos géneros textuales distintos, en dos momentos cronológicamente distintos y con una población de sujetos numerosa nos ha permitido conseguir unos datos robustos para contrastar con las hipótesis postuladas, que buscan establecer los parámetros de variación por (sub)género, por familiarización con la lengua y la práctica de las tareas (variable temporal).

Como hemos indicado en la síntesis de la discusión de datos, las hipótesis se han verificado en su mayoría, como sigue.

- (1) *Hay diferencias entre un relato en el que se exige cierto grado de creatividad, como es inventar una historia (muestra «F»), frente a un relato en el que hay que explicar una experiencia personal (muestra «ID»).*

La primera conclusión a la que hemos llegado es que sí hay diferencias apreciables entre un género textual y otro. Constatamos, por ejemplo, que el texto de la muestra «1D», en la lista de frecuencias con *stopwords*, presenta un conjunto de palabras léxicas que remiten claramente al campo semántico de la educación, mientras que la muestra «F» registra un léxico más variado, ya que al tratarse de un tema más abierto, con más libertad creativa, las opciones del léxico a utilizar son casi ilimitadas. Esto hace difícil determinar la temática del texto en cuestión.

La segunda conclusión relevante en este apartado es que también la muestra «1D» presenta agrupamientos y concordancias que caracterizan a este género textual dentro de la función expresiva del lenguaje frente a la muestra «F», en la que domina la función referencial. Es decir, que a pesar de que la instrucción de la tarea exigía que la historia estuviera contada en primera persona del singular, se trata de un género textual que no requiere de implicación personal. Por tanto, la expresión de sentimientos o emociones quedan en un segundo plano; mientras que en los textos de la muestra «1D», la implicación personal de lo que se está explicando es absoluta, pues atañe a uno mismo, y queda reflejada claramente en el léxico que se selecciona y en las estructuras utilizadas.

Por tanto, podemos concluir que nuestra primera hipótesis se confirma, ya que estamos ante géneros textuales que se distinguen fácilmente por el tipo de léxico y estructuras que presentan uno y otro.

No obstante, verificamos que sí hay una semejanza entre ambas muestras en la lista de frecuencias sin *stopword*, y es que las primeras posiciones están ocupadas por palabras funcionales. Es decir, palabras cuya finalidad es simplemente gramatical. Pero hemos de indicar que este hecho es algo habitual, pues es algo recurrente de otro tipo de lenguas.

- (2) *Pensamos que las producciones escritas de los sujetos no nativos y nativos van a presentar entre ellas diferencias en cuanto al léxico y a las estructuras sintácticas.*

Al comparar el comportamiento de los sujetos nativos y los no nativos en las producciones escritas, la primera conclusión alcanzada es la existencia de igualdad en el tipo de agrupamientos y concordancias que muestran los dos grupos. Es decir, que las producciones de los sujetos no nativos, aun teniendo solo un nivel B1 según el MCER, muestran resultados próximos a los sujetos nativos.

Tras el análisis de los resultados, la siguiente conclusión a la que hemos llegado es que sí hay diferencias léxicas llamativas entre ambos grupos. El grupo N registra mayor diversidad léxica que el grupo NN a la hora de expresar sentimientos o emociones, hecho que está íntimamente relacionado con la función del lenguaje predominante en el texto, la expresiva; mientras que en la muestra del grupo NN destaca la función referencial del lenguaje.

Por tanto, respecto a esta segunda hipótesis, no podemos concluir al cien por cien que haya diferencias en las producciones escritas de los sujetos no nativos y nativos. Sí las hemos encontrado respecto al léxico, pero no en cuanto a estructuras.

- (3) *Hay progreso en el resultado final de las producciones escritas del grupo NN-M2 respecto del grupo NN-M1 tras una instrucción de 120h.*

Para verificar esta hipótesis hemos recurrido al tratamiento estadístico de los datos y hemos constatado que sí se produce progreso estadísticamente significativo en las producciones escritas por los sujetos del grupo NN-M2 (final de curso) frente al grupo NN-M1 (inicio del curso); pero solamente en las variables de fluidez y diversidad léxica. En cuanto a la densidad léxica no se aprecia mejora entre el inicio y final de curso. Es decir, que la proporción de palabras léxicas que utilizan no varía durante el período del curso académico.

Así, no podemos concluir con rotundidad, al igual que en el caso anterior, que haya progreso al cien por cien.

- (4) *Hay mejores resultados en cuanto a calidad de texto en la muestra de las producciones escritas que se trabajaron mediante tareas que en la muestra que no recibió tratamiento específico de enseñanza mediante tareas.*

La primera conclusión clara a la que llegamos en este apartado es que la muestra «F» presenta mejores resultados que la muestra «1D» en las variables de fluidez, diversidad léxica y densidad léxica tanto en el M1 como en el M2.

La siguiente conclusión relevante es que sí hay progreso en la muestra «F» respecto a la fluidez y la diversidad léxica entre el M1 y el M2; y que ese progreso es estadísticamente significativo. Hay que hacer hincapié que esta fue la muestra que se trabajó en el aula con una enseñanza mediante tareas, lo que avala este mejoramiento. Sin embargo, la muestra «1D», que no recibió instrucción específica, no experimenta progreso entre el M1 y el M2.

En cuanto a la densidad léxica, realizamos dos tipos de pruebas estadísticas para confirmar los resultados obtenidos y debemos decir que no podemos aportar una conclusión concluyente, pues los resultados de las pruebas son contradictorios. Mientras que una indica que no hay diferencia significativa estadísticamente, la otra sí diferencia estadística.

A pesar del progreso claro que se produce en la muestra «F» en fluidez y diversidad léxica en el M2, no podemos mostrar un acuerdo absoluto con la hipótesis de partida, pues surge la duda de si esta diferencia entre la muestra «F» y la «1D» en el M2 se debe realmente a la instrucción específica o es más por motivos de tipología textual, ya que la muestra «F» registraba mejores resultados que la muestra «1D» incluso antes de iniciar la enseñanza mediante tareas.

- (5) *Los sujetos nativos van a presentar mejores resultados en calidad de texto que los sujetos no nativos en las producciones escritas.*

Una de las primeras conclusiones sobre la comparación de la competencia escritora entre el grupo de N y NN es que la media de *tokens* del grupo N

respecto del grupo NN (M1 y M2) en los dos tipos de texto analizados marca también una diferencia estadísticamente significativa. Es decir, que el grupo de sujetos N es capaz de escribir textos más extensos que el grupo de sujetos NN. Esto se entiende en la medida que la fluidez mejora en función de las horas de instrucción, como hemos visto claramente con el grupo de sujetos NN. De ahí que en los distintos estadios de lengua (nivel no nativo B1 inicial; nivel no nativo B1 final, y nivel nativo) hayamos comprobado el aumento progresivo de la fluidez textual.

Otra conclusión interesante es que el grupo NN-M2 en la muestra «F», el más avanzado del grupo de no nativos, y el grupo de sujetos nativos presentan igualdad respecto a la diversidad léxica.

En cuanto a la medida de ‘diversidad léxica’ en la muestra «1D», las conclusiones alcanzadas es que también registra diferencia significativa estadísticamente entre el grupo N y el grupo NN (M1 y M2).

Nuevamente no podemos mostrar acuerdo absoluto con la hipótesis, pues, como acabamos de ver, hay similitudes en el grupo N y el grupo NN en algunos aspectos.

A la luz, pues, de lo abordado en este estudio, que no deja de ser parcial y mejorable (especialmente si se replica en el marco de un proyecto y con un equipo de investigación que permita subsanar las limitaciones de un trabajo unipersonal y amplíe la colaboración a expertos en estadística, psicolingüística, etc.), podemos concluir que el desarrollo de la competencia escritora se produce gradualmente y para los subgéneros estudiados (si bien con diferencias entre sí) en un curso de lengua de carácter general, sin fines académicos y sin que la escritura sea objeto de foco específico. Y que el lapso de tiempo seleccionado, un curso, es un marco temporal suficiente para emprender una descripción y apuntar implicaciones didácticas.

Este desarrollo de la competencia se da para todos los grupos estudiados, sin que se detecten diferencias especialmente atribuibles a la L1 de los aprendices. Asimismo, las diferencias detectadas entre los dos momentos y los dos géneros, que se decanta hacia la

mejor prestación en uno de los subgéneros estudiados (relato en 1ª ps), se da también en el grupo nativo de control, lo que pone de manifiesto que la nativéz, *per se*, no es tampoco una garantía de invarianza. Más bien los datos ponen de manifiesto, curiosamente, la variación entre las dos recogidas entre los mismos nativos. Este último punto nos parece interesante por cuanto que pone sobre la mesa la necesidad de considerar de forma distinta y no monolítica la producción ‘modelo’ nativa, toda vez que tiene implicaciones en la comparación como ‘vara de medir’ para la producción no nativa (esta sí siempre caracterizada por la variación). Habrá que ver cómo se recoge esta observación en el marco de la evaluación y el ‘*bench marking*’ de la competencia no nativa por destrezas y por (sub)géneros en futuras propuestas didácticas y de evaluación.

## REFERENCIAS BIBLIOGRÁFICAS

- AGUSTÍN LLACH, M. P. (2007). "Lexical errors as Writing Quality Predictors", *Studia Linguistica*, 61 (1), pp. 1-19.
- AGUSTÍN LLACH, M. P. (2011a). *Lexical Errors and Accuracy in Foreign Language Writing*. Bristol: Multilingual Matters.
- ALBA QUIÑONES, V. (2009). "El análisis de errores en el campo del español como lengua extranjera: algunas cuestiones metodológicas", *Revista Nebrija de Lingüística Aplicada*, 5 (3), pp. 1-16.
- ALBA, J.M. y ZANÓN, J. (1999). "Unidades didácticas para la enseñanza del español/LE en los Institutos Cervantes". En J. Zanón (coord.), *La enseñanza mediante tareas*. Madrid: Edinumen.
- ALCÓN SOLER, E. (2002). *Bases lingüísticas y metodológicas para la enseñanza de la lengua inglesa*. Universidad Jaume I, Servicio de Comunicación y Publicaciones.
- ALVAR EZQUERRA, M., BLANCO RODRÍGUEZ, M.J. y PÉREZ LAGOS, F. (1994). "Diseño de un corpus español en el marco de un corpus europeo". En M. Alvar Ezquerro y J.A. Villena Ponsada (eds.), *Estudios para un corpus de español*. Anejo nº 7 de *Analecta Malacitana*. Málaga: Universidad de Málaga, pp. 9-29.
- ARNAUD, P. J. L. (1984). "The lexical richness of L2 written productions and the validity of vocabulary tests". En T. Culhane, C. Klein-Braley, y D.K. Stevenson (eds.), *Practice and Problems in Language Testing*. University of Essex: Colchester, pp. 14-28.
- ARNDT, V. (1987). "Six writers in search of texts: A protocol-based study of L1 and L2 writing", *ELT Journal*, 41 (4), pp. 257-267.
- ÁVILA, R. (1986). "Léxico infantil en México: Palabras, tipos, vocablos". En R. Ávila, *Actas del Congreso de II Congreso Internacional sobre el español de América*, México, D.F.: Universidad Nacional Autónoma de México, pp. 510-517.
- BACHMAN, L. (1990). "Habilidad lingüística comunicativa". En Llobera *et al.* (1995). *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Edelsa, pp. 105-129.

- BAERLOCHER, C. (2013). *Los errores léxicos en textos escritos en español por alumnos universitarios brasileños en formación como profesores de español lengua extranjera*. Tesis doctoral. Universidad de Barcelona.
- BAERLOCHER, C. y LUCHA, R.M. (2008). “La evaluación objetiva en pruebas subjetivas”. En C. Baerlocher y R.M. Lucha, *Comunicación del XXXVI congreso AESLA*, Universidad de Almería.
- BAERLOCHER, C. y LUCHA, R.M. (2013). “Aportaciones metodológicas y didácticas derivadas de un análisis de errores de ELE en un contexto de lenguas próximas”. En C. Baerlocher y R.M. Lucha, *Comunicación en el XXIV Congreso Internacional ASELE La enseñanza del español como LE/L2 en el siglo XXI*. Universidad de Jaén.
- BAKER, M. (1993). “Corpus Linguistics and Translation Studies: Implications and Applications”. En M. Baker, M. G. Francis y E. Tognini-Bonelli (eds.), *Text and Technology. In Honor of John Sinclair*. Filadelfia/Amsterdam: John Benjamins, pp. 233-250.
- BARALO, M. (1996). “Innatismo e interacción en la adquisición de una lengua extranjera”. En M. Baralo, *ASELE, Actas VII*.
- BARCIA, E. (2002). “El descubrimiento y la construcción de los implícitos culturales.” En A. M. de la Peña Ortega (coord.), *Lenguas para abrir camino*. Madrid: Ministerio de Educación y Ciencia, pp. 227-268.
- BELCHER, D. y HIRVELA, A. (2001). *Linking literacies. Perspectives on L2 Reading-Writing connections*. Ann Arbor: the University of Michigan Press.
- BELCHER, D. y HIRVELA, A. (2008). *The oral-literate connection. Perspectives on L2 Speaking, Writing and other Media interaction*. Ann Arbor: the University of Michigan Press.
- BIBER, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- BREEN, M. (1987). “Learner contributions to task design.” In C. Candlin & D. Murphy (eds.), *Language learning tasks*. London: Prentice Hall.
- BRITTON, J. et al. (1975). *The development of writing abilities*. London: Macmillan.

- BYRAM, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Johannesburg: Multilingual Matters Ltd.
- BYRAM, M. y FLEMING, M. (2001). *Perspectivas interculturales en el aprendizaje de idiomas*. Madrid: Cambridge University Press.
- BYRNES, H. (2011). "Beyond writing as language learning or content learning. Construing foreign language writing as meaning-making". En R. Manchón (ed.), 2011. *Learning-to-Write and Writing-to-Learn*. Amsterdam/Philadelphia: John Benjamins, pp. 133-153.
- BYRNES, H. (2012). *Task based language writing. Issues, Research and Practice*. Amsterdam/Philadelphia: John Benjamins.
- BYRNES, H., MAXIM, H. H. y NORRIS, J. (2010). "Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment", *Modern Language Journal*, 94 (Supplement 1).
- CAMPBELL, C. (1990). "Writing with others' words: using background reading text in academic composition". En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press. pp. 211-230.
- CANALE, M. (1983). "De la competencia comunicativa a la pedagogía comunicativa del lenguaje". En Llobera et al. (1995). *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Edelsa, pp. 63-83.
- CANALE, M. y SWAIN, M. (1980). "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics*, 1, pp. 1-47. Versión en español: "Fundamentos teóricos de los enfoques comunicativos", *Signos*, (17) pp. 56-61 y (18) pp. 78-91, 1996.
- CANDLIN, C. (1987). "Towards task-based language learning". En C. Candlin & D. Murphy (eds.), *Language learning tasks*. London: Prentice Hall.
- CANTOS GOMEZ, P. (2002). "Do we need statistics when we have Linguistics?" *Delta*, 18 (2), pp. 233 – 271.
- CANTOS, P. y SÁNCHEZ, A. (2011). "El inglés y el español desde una perspectiva cuantitativa y distributiva: equivalencias y contrastes". En P. Cantos y A. Sánchez, *Estudios Ingleses de la Universidad Complutense*, 19, pp. 15-44.

- CARROLL, J. B. (1964). "Words, meanings, y concepts", *Harvard Educational Review*, 334, pp. 178-202.
- CELAYA, M.L. y NAVÉS, T. (2009). "Age-related Differences and Associated Factors in Foreign Language Writing. Implications for L2 Writing Theory and School Curricula". En R. M. Manchón (ed.), *Writing in Foreign Language Contexts. Learning, Teaching, and Research*. Bristol, UK: Multilingual Matters.
- CELAYA, M.L., PÉREZ-VIDAL, C. y TORRAS, M.R. (2000-2001). "Matriz de criterios de mediación para la determinación del perfil de competencia lingüística escrita en inglés (LE)", *RESLA*, 14, pp. 87-98.
- CENOS, J. (1996). "La competencia comunicativa: su origen y componentes." En J. Cenoz y J. F. Valencia (eds.), *La Competencia Pragmática: elementos lingüísticos y psicosociales*. Bilbao: Servicio Editorial. Universidad del País Vasco, pp. 95- 114.
- CHARLES, M., PECORARI, D. y HUNSTON, S. (eds) (2009). *Academic Writing: At the interface of Corpus and Discourse*. London: Continuum.
- CIMASKO, T. y REICHELT, M. (eds.) (2011). *Foreign Language Writing Instruction: Principles and Practices*. Parlor Press.
- CONSEJO DE EUROPA (1996). *Marco común europeo de referencia para las lenguas: enseñanza, aprendizaje, evaluación*. Estrasburgo.
- CORROS, F. J. (2005). "Aspectos pragmáticos, sociolingüísticos e interferencias culturales en la enseñanza de ELE en Estados Unidos". En F. J. Corros, *XVI Congreso de ASELE, La Competencia Pragmática o la Enseñanza del Español como Lengua Extranjera*. pp. 213-221.
- CROOKES, G. (1989). "Planning and interlanguage variation". *Studies in Second Language Acquisition*, 11, pp. 367-83.
- CRUZ PIÑOL, M. (2012). *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: ArcoLibros.
- CUMMING, A. (1986). "Intentional learning as a principle for ESL writing instruction: A case study". En P. Lightbown & S. Firth (eds.), *TESL Canada Journal, Special Issue 1*, pp. 69-83.

- CUMMING, A. (2009). "The Contribution of Studies of Foreign Language Writing to Research, Theories and Policies ". En R. Manchón (ed.), *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Multilingual Matters.
- CUMMING, A. (2010). "Theories, frameworks and heuristics. Some reflections on inquiry and SL writing". En Silva, T. y P. Kei Matsuda (eds.), *Practicing Theory in SL Writing*. West Lafayette. IN: Parlor Press, pp. 48-71.
- CUMMING, A. (ed.) (2006). *Goals for Academic Writing. ESL students and their instructors*. John Benjamins Publishing Company.
- DE ANGELIS, G. y JESSNER, U. (2012). "Writing across languages in a bilingual context: A Dynamic Systems Theory approach". En R. Manchón, (ed.), *L2 Writing Development. Multiples Perspectives*. USA: de Gruyter Mouton.
- DE HAAN, P. y K. VAN ESCH (2005). "The development of writing in English and Spanish as foreign languages", *Assessing Writing*, 10, pp. 100-116.
- DÍAZ, L. y AYMERICH, M. (2003). *La destreza escrita*. Madrid: Edelsa.
- DÍAZ, L., BEL, A. y LÓPEZ, L. (2001). "Índices de calidad en un ejercicio de interpretación consecutiva. Investigación y práctica" En L. Díaz, A. Bel y L. López, Actas del I Congreso Internacional sobre la evaluación de la calidad en la interpretación de conferencias. Universidad de Granada, Centro Mediterráneo, Almuñecar. Granada.
- DÍAZ, L., BEL, A. y RUGGIA, A. (2003). "Escribir y evaluar textos de fines específicos con ayuda de recursos informáticos: nuevas tecnologías y EFE". En L. Díaz, A. Bel y A. Ruggia, Actas II Congreso Internacional de Español para Fines Específicos.
- DÍAZ, L. y LUCHA, R.M. (2007). "Análisis del desarrollo de competencias en las destrezas de mediación". En L. Díaz y R. M. Lucha, Comunicación XXV Congreso International AESLA, Universidad de Murcia, Murcia.
- DÍAZ, L. y LUCHA, R.M. (2010). "Teaching genres in an l2 writing course. The incidence of focus on form in written production or how to go a step further than natives", *Monografías MarcoEle*, (11), pp. 20-36.
- DÍAZ, L. y LUCHA, R.M. (2010). "I am writting expositive texts with an accent. Does it matter? Can I be automatically assessed?, En L. Díaz y R. M. Lucha, Comunicación Mapping Language Across Cultures: Language analysis in Cross-Cultural and

Intercultural (MLAC10), Universidad de Salamanca y Centro Cultural Hispano Japonés., Salamanca.

DÍAZ, L. y LUCHA, R.M. (2012). “Escribo con acento en español (l2) ¿eso importa? ¿puedo obtener feedback automáticamente?”, *Monografías MarcoEle*, (15), pp. 68-81.

DÍAZ, L. y LUCHA, R.M. (2015). “¿Progreso en mis textos escritos?”. En L. Díaz y R. M. Lucha, *Comunicación en el XXVIII Congreso Internacional de la Asociación Española de Lingüística Aplicada (AESLA), Análisis de la fluidez y la complejidad en producciones escritas mediante software informático*, Universidad Politécnica de Madrid. Madrid.

DÍAZ, L. y MARTÍNEZ, R. (1998). *E de escribir*. Barcelona: Difusión, colección *Tareas*.

DÖRNYEI, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

DÖRNYEI, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

DOUGHTY, C. y LONG, M.H. (eds.) (2003). *The Handbook of Second Language Acquisition*, Blackwell Publishers.

DOUGHTY, C. y WILLIAMS, J. (1998). “Problemas y terminología”. En C. Doughty y J. Williams (coords.), *Atención a la forma en la adquisición de segundas lenguas en el aula* (ed. Española, 2009). Madrid: Edinumen.

DUGAST, D. (1978). “Sur quoi se fonde la notion d’étendue théorique du vocabulaire?”, *Le français modern*, 1, pp. 25-32.

DUGAST, D. (1979). *Vocabulaire et stylistique. 1. Théâtre et dialogue*. Genève: Slatkine.

DUGAST, D. (1980). *La statistique lexicale*. Geneva, Switzerland: Slatkine.

DURAN, P., MALVERN, D., RICHARDS, B., y N. CHIPERE (2004). “Developmental trends in lexical diversity”, *Applied Linguistics*, 25, pp. 220-242.

DURANTI, A. (2000). *Antropología lingüística*. Madrid: Cambridge University Press

- EAGLES (1996). "Text Corpora Working Group Reading Guide" Documento Eagles *Expert Advisory Group on Language Engineering*, EAG-TCWG-FR-2.
- EGUILUZ PACHECO, J. y DE VEGA SANTOS, C. M. (1996). "Criterios para la evaluación de la producción escrita". *Expolingua*.  
[http://marcoele.com/descargas/expolingua1996\\_eguiluz-vega.pdf](http://marcoele.com/descargas/expolingua1996_eguiluz-vega.pdf) (consultado en línea 15 de abril de 2015)
- EGUILUZ PACHECO, J. y EGUILUZ PACHECO, A. (2004). "La evaluación de la expresión escrita". En J. Sánchez Lobato y I. Santos Gargallo (eds.), *Vademécum para la formación de profesores*. Madrid: SGEL.
- ELLIS, R. (1990). *Instructed Second Language Acquisition*. Oxford: Basil Blackwell 500.
- ELLIS, R. (1991). *Second Language Acquisition and Language Pedagogy*. Clevedon: Multilingual Matters.
- ELLIS, R. (1993). "The Structural Syllabus and Second Language Acquisition", *TESOL Quarterly*, 27 (1), pp. 91-115.
- ELLIS, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- ELLIS, R. (2001). *Form-Focused Instruction and Second Language Learning*. Blackwell Publishers.
- ELLIS, R. (2002). "Does Form-Focused Instruction affect the acquisition of implicit Knowledge", *SSLA*, 24, pp. 223-236.
- ELLIS, R. (2008). *The Study of Second Language Acquisition*. (2 ed.). Oxford: Oxford University Press.
- ELLIS, R. (2005). *La Adquisición de Segundas Lenguas en un contexto de Enseñanza. Análisis de las investigaciones existentes*. Departamento Técnico del Ministerio de Educación de Nueva Zelanda.
- ELLIS, R., y SHINTANI, N.. (2014). *Exploring Language Pedagogy through Second Language Acquisition Research*. Routledge.

- EMIG, J. (1971). *The composition processes of twelfth graders*. Urbana. III National Council of Teachers of English.
- ENGBER, C. (1995). "The Relationship of Lexical Proficiency to the Quality of ESL Compositions", *Journal of Second Language Writing*, 4(2), pp. 139-155.
- ESTÉVEZ, M. y FERNÁNDEZ, Y. (2006). *El componente cultural en la clase de EL/E*. Madrid: Edelsa.
- FAIGLEY, L y WITTE, S. (1981). "Analyzing revisions", *College Composition and Communication*, (32), pp. 400-414.
- FATHMAN, A.K. y WHALLEY, E. (1990). "Teacher responses to student writing: Focus on form versus content. En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Nueva York: Cambridge University Press, pp. 178-198.
- FIRTH, A. y J. WAGNER. (1997). "On discourse, communication, and (some) fundamental concepts in SLA research", *Modern Language Journal*, (81), pp. 285-300.
- FOSTER, P. y SKEHAN, P. (1996). "The influence of planning on performance in task-based learning", *Studies in Second Language Acquisition*, (18), pp. 299-324.
- FOSTER, P. y SKEHAN, P. (1999). "The influence of source of planning and focus of planning on task-based performance", *Language Teaching Research*, 3 (3), pp. 215-247.
- FRIEDLANDER, A. (1990). "Composing in English: Effects of a first language on writing in English as a second language. En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Nueva York: Cambridge University Press, pp. 109-125.
- GARCÍA GARCÍA, Pilar. (2004). "La cultura, ¿universo compartido? La didáctica intercultural en la enseñanza de idiomas", *RedELE* 0.
- GARCÍA PAREJO, I. (1993). "La expresión escrita en español l2: motivación y creatividad para el desarrollo de diferentes destrezas lingüísticas". En I. García Parejo, *ASELE*, Actas IV.  
[http://cvc.cervantes.es/Ensenanza/biblioteca\\_ele/asele/pdf/04/04\\_0289.pdf](http://cvc.cervantes.es/Ensenanza/biblioteca_ele/asele/pdf/04/04_0289.pdf)  
 (consultado en línea 15 de abril de 2015).

- GASS, S. y SELINKER, L. (2008). *Second Language Acquisition. An introductory Course* (3 ed.). London: Routledge.
- GEESLIN, K. L. (2014). *The handbook of Spanish Second Language Acquisition*. John Wiley & Sons, Inc
- GÓMEZ DEL ESTAL, M. y ZANÓN, J. (1999). “Tareas formales para la enseñanza de la gramática en la clase de español”. En J. Zanón (coord.), *La enseñanza del español mediante tareas*. Madrid: Edinumen.
- GUIRAUD, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Dordrecht: D. Reidel.
- HALLIDAY, M.A.K. y HASAN, R. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- HAN, Z. (2004). *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters Ltd.
- HERNÁNDEZ SAMPIERI, R., FERNÁNDEZ COLLADO, C. y BAPTISTA LUCIO, P. (1997). *Metodología de la investigación*. McGraw-Hill.
- HERSCHENSOHN, J. y YOUNG-SCHOLTEN, M. (eds.). *The Cambridge Handbook of Second Language Acquisition*. Cambridge University Press.
- HILLOCKS, G. (1986). *Research for written composition: New directions for teaching*. Urbana: Illinois.
- HYLAND, K. (2009). *Teaching and Researching Writing*. (2ª ed.). London: Longman.
- HYLAND, K. y HYLAND, F. (2006). “Contexts and issues in feedback on L2 writing: An introduction”. En K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues*. Nueva York: Cambridge University Press. pp. 1-22.
- HYMES, D. H. (1972). “Acerca de la competencia comunicativa”. En Llobera *et al.* (1995). *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Edelsa, pp. 27-47.

- JARVIS, S. (2002). "Short texts, best-fitting curves and new measures of lexical diversity", *Language Testing*, 19 (1), pp. 57–84.
- JARVIS, S. (2002). "Short texts best-fitting curves and new measures of lexical diversity", *Language Testing*, 19 (1), pp. 57-84.
- JARVIS, S., GRANT, L., BIKOWSKI, D. y FERRIS, D. (2003). "Exploring multiple profiles of highly rated learner compositions", *Journal of Second Language Writing*, 12, pp. 377–403.
- JOHNS, A. M. (1990). "L1 composition theories: Implications for developing theories of L2 composition". En B. Kroll (Ed.), *Second language writing: Research insights for the classroom*. Nueva York: Cambridge University Press, pp. 24-36.
- JOHNSON, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: The University of Chicago Press.
- KASTEN, S., DANTAS-WHITNEY, M., RILLING, S. y SAVOVA, L. (2010). *Effective second language writing*. Alexandria, Va: TESOL.
- KELLERMAN, E. (2001). "New uses for old language: cross-linguistic and cross-gestural influence in the narratives of non-speakers". En J. Cenoz, B. Hufeisen, y U. Jessner (eds.), *Cross-linguistic influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters, pp. 170–191.
- KRAMSCH, C. (2001). "El privilegio del hablante intercultural". En M. Byram y M. Fleming (eds), *Perspectivas interculturales en el aprendizaje de idiomas*. Madrid: Cambridge University Press.
- KRAMSCH, Cl. y WHITESIDE, A. (2007). "Three Fundamental Concepts in Second Language Acquisition and Their Relevance in Multilingual Contexts", *The Modern Language Journal*, (91).
- KRASHEN, S. D. (1977). "El modelo del monitor y la actuación de los adultos en L2". En J. Muñoz Licerias (1992) (ed.), *La adquisición de las lenguas extranjeras*. Madrid: Visor Dis.
- KRASHEN, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Nueva York: Longman.

- KROLL, B. (1990). *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press.
- KROLL, B. (ed.) (2003). *Exploring the Dynamics of Second Language Writing*. Cambridge: Cambridge University Press.
- LACÁMARA RUBERTE, P. (1989). “El estatuto del escrito dentro de un enfoque comunicativo. Hacia un modelo didáctico de la producción escrita”. En P. Lacámara Ruberte, *ASELE*, Actas I.  
[http://cvc.cervantes.es/Ensenanza/biblioteca\\_ele/asele/pdf/01/01\\_0323.pdf](http://cvc.cervantes.es/Ensenanza/biblioteca_ele/asele/pdf/01/01_0323.pdf)  
 (consultado en línea 15 de abril de 2015).
- LARSEN-FREEMAN, D. (1991). “Second Language Acquisition Research: Staking Out the Territory”, *TESOL Quarterly*, 25 (2), pp. 315-350.
- LARSEN-FREEMAN, D. (2006). “The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English”, *Applied Linguistics*, 27 (4) pp. 590-619.
- LARSEN-FREEMAN, D. y LONG, M. (1991). *An Introduction to Second Language Acquisition Research*. Londres: Longman.
- LAUER, J. M. y ASHER, J.W. (1988). *Composition Research: empirical designs*. Nueva York: Oxford University Press.
- 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*. Clevedon: Multilingual Matters, pp. 19-31.
- LAVIOSA, S. (1998). “The Corpus-based Approach: A New Paradigm in Translation Studies”, *META*, 43 (4), pp. 474-479.
- LAZARATON, A. (2005). “Quantitative Research Methods”, E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, London: Lawrence Erlbaum Associates, pp. 209–224.
- LEKI, D, CUMMING, A. y SILVA, T. (2008). *A Synthesis of Research on Second Language Writing in English*. Nueva York: Routledge/Taylor y Francis.

- LINNARUD, M. (1983). On lexis: the Swedish learner and the native speaker compared. In Cross Language Analysis and Second Language Acquisition, 2. Ed., K. Sajavaara, Jyväskylä.
- LLOBERA, M. (1995). "Una perspectiva sobre la competencia comunicativa y la didáctica de las lenguas extranjeras". En LLOBERA, M. (coord.), *Competencia comunicativa: documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Ed. Edelsa.
- LOEWEN, S. y GASS, S. (2009). "The use of statistics in L2 acquisition research", *Language Teaching*, 42, pp. 181-196.
- LONG, M. (1983). "Native speaker/non-native speaker conversation and the negotiation of comprehensible input", *Applied Linguistics*, 4 (2), pp. 126-41.
- LONG, M. (1985). "A role for instruction in second language acquisition." En Hyltenstam, K y Pieneman, M. (eds.), *Modelling and Assessing Second Language Acquisition*. Clevedon Avon: Multilingual Matters.
- LONG, M. H. (1997). "Focus on form in Task-Based Language Teaching", *The McGraw-Hill Companies*.
- LÓPEZ MORALES, H. (2010). "Los índices de 'riqueza léxica' y la enseñanza de lenguas". En J. de Santiago, H. Bongaerts, J.J. Iglesias y M. Seseña (coords.), *Actas del XXI Congreso Internacional de la ASELE*, Salamanca.
- LÓPEZ MORALES, H. (1994). *Métodos de Investigación Lingüística*. Biblioteca Filológica Salamanca: Ediciones Colegio de España.
- LOZANO, C. y MENDIKOETXEA, A. (2008). "Postverbal subjects at the interface in Spanish and Italian learners of L2 English: a corpus analysis." En G. Gilquin, S. Papp, M.B. Díez-Bedmar (eds). *Linking up contrastive and corpus learner research*. Amsterdam: John Benjamins, pp. 127-166.
- LOZANO, C. y MENDIKOETXEA, A. (2010). "Interface conditions on postverbal subjects: a corpus study of L2 English." *Bilingualism: Language and Cognition*, 13 (4), pp. 475-497.
- LUCHA, R.M. (2006). Lucha, RM. "¿Por qué enseñar a escribir utilizando técnicas de "focus on form"?" MarcoEle 2006  
<http://www.marcoele.com/num/2/escriturafocoenlaforma.php>)

- LUCHA, R. M. Y BARLOCHER, C. (2008). “¿Es posible evaluar objetivamente contenidos gramaticales en producciones escritas?”. R. M. Lucha y C. Baerlocher, *Comunicación en el XVIII Congreso Internacional ASELE*, Universidad de Alicante, Alicante.
- LUCHA, R. M. Y BARLOCHER, C. (2012). “El efecto del enfoque comunicativo en la adquisición del contraste de pasados en un contexto multilingüe”. En R. M. Lucha y C. Baerlocher, *Comunicación en el XXIII Congreso Internacional ASELE, Multilingüismo y enseñanza de ELE en contextos multiculturales*, Universidad de Girona, Girona.
- MACKEY, A. y GASS, S. (2005). *Second Language Research: Methodology and design*. Nueva Jersey: Lawrence Erlbaum Associates.
- MACWHINNEY, B. (2001). “From CHILDES to TalkBank”. En M. Almgren et al.(eds.), *Research on Child Language Acquisition*. Somerville, MA: Cascadilla, pp. 17-34.
- MACWHINNEY, B. (2005). “A Unified Model of Language Acquisition”. En J. F. Kroll y A. M. B. De Groot (eds.), *Handbook of Bilingualism*. Oxford University Press.
- MALVERN, D., RICHARDS, B. J., CHIPERE, N. y P. DURAN (2004). *Lexical diversity and language development: quantification and assessment*. Nueva York: Palgrave Macmillan.
- MALVERN, D., y RICHARDS, B. J. (2000). “Investigating accommodation in language proficiency interviews using a new measure of lexical diversity”, *Language Testing*, 19, pp. 85-104.
- MANCHÓN, R. M., ROCA DE LARIOS, J. y MURPHY, L. (2009). “The temporal dimension and problem-solving nature of foreign language composing. Implications for theory”. En R. M. Manchón (ed.), *Foreign Language Writing: Learning, Teaching and Research*. Clevedon: Multilingual Matters.
- MANCHÓN, R. M. (2013). “Writing”. En F. Grosjean y Li Ping (eds.), *The psycholinguistics of bilingualism*. Malden, MA: Wiley-Blackwell, pp. 100-115.
- MANCHÓN, R. M. (ed.) (2009). *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Multilingual Matters.
- MANCHÓN, R. M. (ed.) (2011). *Learning-to-Write and Writing-to-Learn in an Additional Language*. John Benjamins Publishing Company.

- MANCHÓN, R. M. (ed.) (2012). *L2 Writing Development. Multiples Perspectives*. USA: de Gruyter Mouton.
- MANCHÓN, R. M. (ed.) (2014). *Research methods and approaches in Applied Linguistics*. John Benjamins. AILA.
- MANCHÓN, R.M. y ROCA DE LARIOS, J. (2011). "Writing to learn in FL contexts: Exploring learners' perceptions of the language learning potential of L2 writing". En R. M. Manchón (ed.), *Learning-to-write and writing-to-learn in an additional language*. Philadelphia/Amsterdam: John Benjamins, pp. 181-207.
- MANCILLA, R. L., POLAT, N. y AKCAY, A.O. (2015). "An Investigation of Native and Nonnative English Speakers' Levels of Written Syntactic Complexity in Asynchronous Online Discussions", *Applied Linguistics* (publicado online: 15 de abril de 2015).
- MARTÍN PERIS, E. (1993). "Propuestas de trabajo de la expresión escrita". Expolingua. En *MonográficosMarcoEle*, (8), 2009.
- MARTÍN PERIS, E. (1996). *Las actividades de aprendizaje en los manuales de E/LE*. Tesis doctoral. Universidad de Barcelona. Barcelona.
- MARTÍN PERIS, E. (2004). "¿Qué significa trabajar en clase con tareas comunicativas?". *RedELE* 0.
- MARTÍNEZ ARBELAIZ, M.A. (2004). "Índices de progreso en la producción escrita de estudiantes de español en situación de inmersión", *RAEL: Revista Electrónica de Lingüística Aplicada*, ISSN 1885-9089, (3), pp. 115-145.
- MATSUDA, P. K., COX, M., JORDAN, J. y ORTMEIER-HOOPER, C. (eds.) (2006). *Second-language writing in the composition classroom: A critical sourcebook*. Boston: Bedford/St. Martin's Press; Urbana, IL: National Council of Teachers of English.
- MATSUDA, P. K. y SILVA, T. (eds.) (2005). *Second language writing research: Perspectives on the process of knowledge construction*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MAYOR, J. (1994). "Adquisición de una Segunda Lengua",. En J. Mayor, ASELE, *Actas IV*.

- MCCLELLAND, B. W. y DONOVAN, T. R. (Eds.) 1985. *Perspectives on Research and scholarship in composition*. New York: Modern Language Association.
- MCLAUGHLIN, B. (1992). "Algunas consideraciones metodológicas sobre el modelo del monitor". En J. Muñoz Licerias (ed.), *La adquisición de las lenguas extranjeras*. Madrid: Visor Dis.
- MCWHINNEY, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Nueva Jersey: Lawrence Erlbaum Associates.
- MELLOR, A. (2011). "Essay Length, Lexical Diversity and Automatic Essay Scoring", *Memoirs of the Osaka Institute of Technology, Series B*, 55 (2), pp. 1-14.
- MORENO FERNÁNDEZ, Francisco. (2007). "Adquisición de Segundas Lenguas y Sociolingüística", *Revista de Educación*, (343) pp. 55-70.
- MUÑOZ, C. (ed.) (2000). *Segundas lenguas: Adquisición en el aula*. Barcelona: Editorial Ariel Lingüística, S.A.
- NAVÉS, T., MIRALPEIX, I. y M. L. CELAYA (2005). "Who Transfer More ... and What? Cross-linguistic Influence in Relation to School Grade and Language Dominance in EFL", *International Journal of Multilingualism*, 2 (2), pp. 113-134.
- NORRIS, J.M. y ORTEGA, L. (2001). "Does Type of Instruction Make a Difference? Substantive Findings From a Meta-analytic Review", *Language Learning*, 51, pp. 157-213.
- NORRIS, JM. y ORTEGA, L. (2009). "Towards an organic approach to investigating CAF in instructed SLA: The case of complexity", *Applied Linguistics*, 30 (4), pp. 555-578.
- NORTH, S. M. (1987). *The making of Knowledge in Composition: portrait of an emerging field*. Upper Montclair, Nueva Jersey: Boynton/Cook.
- NUNAN, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.
- NUNAN, D. (2006). "Task-based Language Teaching in the Asia Context: Defining 'Task'", *Asian EFL Journal*, 8 (3), pp. 12-18.

- ORTEGA, L. (2004). "Aproximaciones cognitivo-interaccionistas al aprendizaje de segundas lenguas mediante tareas", *ELIA* 5, pp. 15-40.
- ORTEGA, L. y CARSON, J. G. (2010). "Multicompetence, social context, and L2 writing research praxis". En T. Silva y P. Matsuda (eds.), *Practicing theory in second language writing*. West Lafayette, IN: Parlor Press, pp. 48-71.
- ORTEGA, L. y IBERRI-SHEA, G. (2005). "Longitudinal research in second language acquisition: Recent trends and future directions", *Annual Review of Applied Linguistics*, 25, pp. 26-45.
- PALLOTTI, G. (2015). "A simple view of linguistic complexity". *Second Language Research*, Vol. 31 (1), pp. 117-134.
- PASTOR, S. (1994). "El desarrollo de la competencia escrita del alumno de español". En S. Pastor, *ASELE*, Actas V.   
[http://cvc.cervantes.es/Ensenanza/biblioteca\\_ele/asele/pdf/05/05\\_0245.pdf](http://cvc.cervantes.es/Ensenanza/biblioteca_ele/asele/pdf/05/05_0245.pdf) (consultado en línea 15 de abril de 2014).
- PAVESI, M. (1984). "The acquisition of relative clauses in a formal and in an informal setting: further evidences in support of the markedness hypothesis". En Singleton y D. Little (eds.), *Language learning in formal and informal contexts*. IRAAL: Dublin.
- PEDROSA, I., et al. (2015). "Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar?" *Universitas Psychologica*, 14 (1), pp. 15-24. Disponible en <http://dx.doi.org/10.11144/Javeriana.upsy14-1.pbad>.
- PIAGET, J. (1979). *The Development of Thought*. Nueva York: Viking.
- POLIO, C. (2003). "Research on Second Language Writing: An overview of what we investigate and how". En B. KROLL (ed.), (2003), *Exploring the Dynamics of Second Language Writing*. Cambridge: Cambridge University Press.
- POLIO, C. (2012). "How to research second language writing". En S. Gass y A. Mackey, (eds.), *Second language research methods*. Nueva York: Routledge, pp. 139-157.
- POLIO, Ch. (2001). "Research Methodology in L2 Writing Assessment." En T. Silva y P. Kei Matsuda (eds.), *On Second Language Writing*. Mahwah, NJ: Lawrence Erlbaum Associates.

- RAIMES, A. (1983). "Tradition and revolution in ESL teaching", *TESOL Quarterly*, 17 pp. 535-552.
- RAIMES, A. (1985). "What unskilled writers do as they write", *TESOL Quarterly*, 19, pp. 229-258.
- RAIMES, A. (1987). "Language proficiency, writing ability, and composition strategies: A study of ESL college student writers", *Language Learning*, 37 (3), pp. 439-467.
- RAMALLO, F.F. (1999). "Informática y sociolingüística cuantitativa", *Revista Española de Lingüística Aplicada*. Vol. Extra, pp. 263-290
- READ, J. (2000). *Assessing vocabulary*. Cambridge: CUP.
- REID, J. (1990). "Responding to different topic types: a quantitative analysis from a contrastive rhetoric perspective". En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press, pp. 191-210.
- RICHARDS, B. J. y MALVERN, D. D. (1999). "The application of a new measure of lexical diversity to preschool children". En B.J. Richards y D.D. Malverns, 8<sup>th</sup> International Congress for the Study of Child Language, The University of Basque Country, San Sebastian-Donostia, España.
- RINNERT, C. y KOBAYASHI, H. (2009). "Situated writing practices in foreign language settings: the role of previous experience and instruction. En R. Manchón (ed.), *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Multilingual Matters.
- RISAGER, K. (2001). "La enseñanza de idiomas y el proceso de la integración europea". En M. Byram y M. Fleming (eds), *Perspectivas interculturales en el aprendizaje de idiomas*. Madrid: Cambridge University Press.
- ROBINSON, P. (2011). "Task-Based Language Learning: A Review of Issues", *Language Learning* 61: Suppl. 1, pp. 1-36.
- ROBINSON, P. y ELLIS, N. C. (eds.) (2008). *Handbook of Cognitive Linguistics and Second Language Acquisition*. Nueva York: Routledge.

- ROCA, J., MANCHÓN, R. y MURPHY, L. (2007). “Componentes básicos y evolutivos del proceso de formulación en la escritura de textos en lengua materna y lengua extranjera”, *RESLA*, 20, pp. 159-183.
- ROWE, A. (1990). “An overview of second language writing process research”. En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press, pp. 37-56.
- SÁNCHEZ LOBATO, J. y SANTOS GARGALLO, I. (2004). *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*. Madrid: SGEL.
- SANTOS, T. (1988). “Professor's reactions to the academic writing of non-native-speaking students”, *TESOL Quarterly*, 22, pp. 69-90.
- SANZ, C. y MORGAN-SHORT, K (2004). “Positive evidence versus explicit rule presentation and explicit negative feedback: a computer – assisted study”. *Language Learning*, 54 (1), pp. 35-78.
- SASAKI, M. (2009). “Changes in English as a Foreign Language Students’ Writing Over 3.5 years: A Sociocognitive Account ”. En R. Manchón (ed.), *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Multilingual Matters.
- SAVILLE-TROIKE, M. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- SCHOONEN, R., SNELLINGS, P., STEVENSON, M. y VAN GELDEREN, A. (2009). “Towards a Blueprint of the Foreign Language Writer: The Linguistic and Cognitive Demands of Foreign Language Writing ”. En R. Manchón (ed.), *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Multilingual Matters.
- SCHUMANN, J. H. (1978). *The Pidgination Process: A Model for Second Language Acquisition*. Rowley, MA: Newbury House.
- SCHUMANN, J. H. (1986). “Research of the Acculturation Model for Second Language Acquisition”, *Journal of Multilingual and Multicultural Development*.
- SHAVELSON, R y STERN, P. (1981). “Research on teachers’ pedagogical thought, judgements, decisions and behaviour.”, *Review of Educational Research*, 51,

- SILVA, T. (1990). "Second Language composition instruction: Developments, issues, and direction in ESL. En B. Kroll (ed.), *Second language writing: Research insights for the classroom*. Nueva York: Cambridge University Press, pp. 11-23.
- SILVA, T. y MATSUDA, P. K. (eds.) (2010). *Practicing theory in second language writing*. West Lafayette, IN: Parlor Press.
- SINCLAIR, J. (2001). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, J. (2004). "Developing Linguistic Corpora: a Guide to Good Practice. Corpus and Text - Basic Principles". (Consultado en línea 27 de noviembre de 2104). <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>
- SKEHAN, P. (1996). "Framework for the Implementation of Task- based Instruction", *Applied Linguistics*, Oxford University Pres, 17 (1), pp. 38-66.
- STRÖMQVIST, S. et al. (2002). "Toward a crosslinguistic comparison of lexical quanta in speech and writing", *Written Language and Literacy*, 5, pp. 45 – 67.
- SWAIN, M. (2000). "The output hypothesis and beyond: Mediating acquisition through collaborative dialogue". En James P. Lantolf (ed.), *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- SWAIN, M. (1998). "Focus on form through conscious reflection". En C. Doughty y J. Williams (eds), *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- SWAIN, M. y LAPKIN, S. (2001). "Focus on form through collaborative dialogue: Exploring task effects". En M. Bygate et. al. (eds.), *Researching Pedagogic Tasks Second Language Learning, Teaching and Testing*. Essex, Longman.
- TARDY, C. M. (2011). "The Future of Genre in Second Language Writing: A North American Perspective", *Journal of Second Language Writing*, 20 (1), pp. 1–5.
- TEIJEIRA RODRÍGUEZ, M. C., VAN ESCH, K. y DE HAAN, P. (2005). "La coherencia y la cohesión en textos escritos por estudiantes neerlandeses de español como LE" *Estudios de Lingüística Aplicada*, 41, pp. 67-100.

- TORRAS, M.R. y CELAYA, M.L. (2001). "Age-related differences in the development of written production. An empirical study of EFL school learners". En R. M. Manchón (ed.), *Writing in the L2 Classroom: Issues in Research and in Pedagogy*. Special issue of *International Journal of English Studies* 1 (2), pp. 103-126.
- TRÉVISE, A. (1993). "Acquisition/apprentissage/enseignement d'une langue 2: modes d'observation, modes d'intervention". *Études de linguistique appliquée*, (92), pp. 38-50.
- VALLEJO, P. M. (2008). *Estadística aplicada a las ciencias sociales*. Madrid: Universidad Pontificia Comillas.
- VAN EK, J. (1986). *Objectives for Foreign Language Learning* (Vol I.). Estrasburgo: Council of Europe.
- VAN ESCH, K., DE HAAN, P., FRISSEN, L., GONZÁLEZ SANTERO, I. y DE LA TORRE MIRANDA, A. (2006). "Evolución en la competencia escrita de estudiantes de español como lengua extranjera", *Estudios de Lingüística Aplicada*, 43, pp. 55-76.
- VAN ESCH, K., P. DE HAAN y M. NAS (2004). "El desarrollo de la escritura en inglés y español como lenguas extranjeras", *Estudios de Lingüística Aplicada*, 39, pp. 53-79.
- VERMEER, A. (2000). "Coming to grips with lexical richness in spontaneous speech data", *Language Testing*, 17, pp. 65-83.
- VERSPoor, M. y SMISKOVA, H. (2012). "Foreign language writing development from a dynamic usage based perspective". En R. Manchón, (ed.), *L2 Writing Development. Multiple Perspectives*. USA: de Gruyter Mouton.
- VERSPoor, SCHMID, M.S. y XU, X. (2012). "A dynamic usage based perspective on L2 writing development", *Journal of Second Language Writing*, 21 (3), pp. 239-263.
- VYGOTSKY, L.S. (1978). *Mind in Society: the Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- WIDDOWSON, H.G. (1995). "Conocimiento de la lengua y habilidad para usarla." En Llobera, M. (coord.), *Competencia comunicativa: documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Ed. Edelsa.

- WILLIAMS, J. (2008). "Focus on form and L2 writing instruction". *The Korean Language in America*.
- WILLIAMS, J. (2012). "The role(s) of writing and writing instruction in L2 development", *Journal of Second Language Writing*, 21, pp. 321-331.
- WILLIAMS, M. y BURDEN, R. L. (1999). *Psicología para profesores de idiomas*. Cambridge: Cambridge University Press. Madrid: Cambridge University Press. Col. Cambridge de Didáctica de lenguas.
- WOLFE-QUINTERO, K., INAGAKI, S. y H-Y KIM (1998). Second language development in writing: Measures of fluency, accuracy, and complexity (Report No. 17). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- WOODS, A., FLETCHER, P. y HUGHES, A. (1986). *Statistics in Language Studies*, Cambridge University Press.
- WRIGHT, H. H., SILVERMAN, S. W. y M. NEWHOFF (2003). "Measures of lexical diversity in aphasia", *Aphasiology*, 17 (5), pp. 443-452.
- WRIGHT, T. (1987). "Instructional task and discorsal outcome in the L2 classroom." En C. Candlin & D. Murphy (eds.), *Language learning tasks*. London: Prentice Hall.
- ZAMEL, V. (1982). "Writing: The Process of Discovering Meaning", *TESOL Quarterly*, 16, pp. 195-209.
- ZAMEL, V. (1983). "The composing processes of advanced ESL students: Six case studies", *TESOL Quarterly*, 17, pp. 165-187.
- ZAMEL, V. (1987). "Recent Research on Writing Pedagogy", *TESOL Quarterly*, 21, pp. 697-715.
- ZANÓN, J. (1988). "Comunicación y discurso en la clase de lengua", *Cable*, (3), pp. 22-32.
- ZANÓN, J.(1988). "Del aprendizaje a la adquisición de la segunda lengua", *Cable*, (2), pp. 47-52.



## **ANEXOS**

### **ANEXO 1**

**PLANTILLA PARA ESCRIBIR EL TEXTO 'F'**

### **ANEXO 2**

**PLANTILLA PARA ESCRIBIR EL TEXTO '1D'**



