

Índice

Prólogo.....11

Presentación15

Parte I **Los corpus y la lingüística de corpus**

1. La lingüística de corpus.....23

2. Corpus textuales31

3. Parámetros clasificatorios de los corpus41

3.1 Modalidad.....42

3.2 Temática.....43

3.3 Época.....45

3.4 Temporalidad45

3.5 Magnitud46

3.6 Evolución48

3.7 Distribución.....49

3.8 Número de ediciones50

3.9 Número de lenguas.....53

3.10 Tipo de edición53

3.11 Muestras.....53

3.12 Marcaje.....55

4. Corpus de lectura59

Parte II

Diseño de la estructura del corpus y post-edición de los textos

5. Fases en la construcción de un corpus	63
6. Estructura y ejes principales	67
6.1 Eje temporal	72
6.1.1 Límites.....	73
6.1.2 Periodización.....	74
6.1.3 El eje diacrónico en los corpus históricos informatizados iberorrománicos	83
6.2 Eje diatópico.....	84
6.2.1 Organización	84
6.2.2 El eje diatópico en los corpus históricos informatizados iberorrománicos	91
6.3 Eje tipológico.....	93
6.3.1 Clasificaciones tipológicas para textos antiguos.....	103
6.3.1.1 Propuestas de clasificación	103
6.3.1.2 Postulados teóricos.....	105
6.3.2 Propuesta de indicadores para una clasificación tipológica.....	107
6.3.3 Nivel de formalidad y grado de oralidad	117
6.3.3.1 Nivel de formalidad.....	118
6.3.3.2 Grado de oralidad.....	119
6.3.4 El eje tipológico en los corpus históricos informatizados iberorrománicos.....	123
6.4 Traducciones	126
7. Composición del corpus	129
7.1 Representatividad y equilibrio	129
7.1.1 Representatividad cualitativa	137
7.1.2 Representatividad cuantitativa.....	138
7.1.2.1 Equilibrio externo.....	139
7.1.2.2 Equilibrio interno	139

7.1.3	Cantidad de textos que deben configurar un corpus	142
7.1.4	Tamaño de las muestras	145
7.2	Corpus provisional versus corpus definitivo.....	147
7.3	Selección de las obras	149
7.4	Selección de los documentos.....	153
7.5	Filiación de los documentos (Metadatos)	156
7.5.1	Título	157
7.5.2	Autor.....	157
7.5.3	Fecha.....	159
7.5.4	Tipo textual.....	162
7.5.5	Dialecto	162
7.6	Principios básicos en la composición de un corpus.....	162
8.	Preparación de los textos	165
8.1	Preparación de los ficheros.....	166
8.1.1	Codificación de los caracteres	166
8.1.2	Formato de los ficheros.....	166
8.1.3	Nombre de los ficheros.....	166
8.2	Edición textual	167
8.2.1	Tipo de edición	168
8.2.1.1	Edición facsímil	170
8.2.1.2	Edición diplomática o paleográfica	170
8.2.1.3	Edición diplomático-interpretativa.....	171
8.2.1.4	Edición normalizada	172
8.2.1.5	Edición crítica	172
8.2.1.6	Multiedición	176
8.2.2	Normas de edición.....	178
8.2.2.1	Unidades léxicas	181
8.3	Edición filológica digital.....	187
8.3.1	Marcaje de las obras.....	193
8.3.2	SGML	197

8.3.3	XML.....	198
8.3.4	Estructura de las etiquetas.....	198
8.3.5	Definición del tipo de documento (DTD).....	200
8.3.6	TEI.....	206
8.3.6.1	La estructura de un documento TEI.....	207
8.3.6.2	Cabecera (Header).....	209
8.3.6.3	Cuerpo (Body).....	211
8.3.6.4	La edición filológica digital en el sistema TEI.....	223
8.4	Edición lingüística.....	224
8.4.1	Lematización.....	224
8.4.1.1	Lexicón.....	225
8.4.1.2	Lematización y categorización.....	227
8.4.1.3	Análisis morfológico.....	227
8.4.1.4	Desambiguación.....	228
8.4.2	Estandarización de las categorías (EAGLES).....	230

Parte III

Bases científicas de la investigación a partir de corpus

9.	Elementos base en la investigación científica.....	237
10.	Método comparativo.....	243
11.	Bases estadísticas en la investigación con corpus.....	247
11.1	Análisis.....	248
11.2	Las variables.....	250
11.3	El muestreo.....	252
12.	El valor de la estadística.....	255
13.	Epílogo.....	259
	Referencias bibliográficas.....	261