

# ÍNDICE

LISTA DE FIGURAS Y TABLAS .....	XI
PREFACIO .....	XIII
<b>CAPÍTULO I. BREVE RECORRIDO HISTÓRICO</b> .....	1
1.1 Introducción .....	1
1.2 Las políticas europeas en materia de industrias de la lengua .....	2
1.2.1 Antecedentes .....	3
1.2.2 El VII Programa Marco .....	8
1.3 La investigación con corpus en lingüística computacional .....	10
1.3.1 La vertiente teórica .....	12
1.3.2 Recursos y técnicas .....	16
1.3.3 Aplicaciones .....	20
1.4 El florecimiento de los estudios aplicados .....	27
1.4.1 Elaboración de gramáticas y diccionarios .....	29
1.4.2 Enseñanza de lenguas .....	34
1.4.3 Otros usos .....	40
<b>CAPÍTULO II. METODOLOGÍA DE CORPUS PARA EL ESTABLECIMIENTO DE LA EQUIVALENCIA</b> .....	47
2.1 Introducción .....	47
2.2 El corpus como nuevo paradigma .....	49
2.2.1 La dicotomía “corpus-based” versus “corpus-driven” .....	50
2.2.2 La gramática léxica .....	55
2.2.2.1 Conceptos básicos .....	56
2.2.2.2 Reformulaciones y ampliaciones .....	59
2.3. Detección y análisis semi-automático de equivalentes .....	65
2.3.1 El proyecto Multilingual Corpora .....	66
2.3.1.1 E. Tognini-Bonelli .....	69
2.3.1.2 W. Teubert .....	73
2.3.2 Otros proyectos relacionados .....	75
2.3.2.1 <i>METIS</i> .....	77
2.3.2.2 <i>ASSIST</i> .....	79
<b>CAPÍTULO III. APORTACIONES A LOS ESTUDIOS DE TRADUCCIÓN E INTERPRETACIÓN</b> .....	83
3.1 Introducción .....	83
3.2 Aplicaciones didácticas .....	85
3.2.1 Enseñanza de la traducción .....	89

3.2.2 Enseñanza de la interpretación .....	95
3.3 Estudios descriptivos.....	98
3.3.1 Regularidades, leyes, universales .....	99
3.3.2 Caracterización de la lengua traducida.....	103
3.3.2.1 Simplificación.....	104
3.3.2.2 Explicitación.....	107
3.3.2.3 Normalización .....	111
3.3.2.4 Convergencia.....	115
3.3.2.5 Transferencia .....	117

#### **CAPÍTULO IV. INVESTIGACIÓN DE LA HIPÓTESIS DE UNIVERSALIDAD:**

<b>ORGANIZACIÓN DE LOS DATOS</b> .....	125
4.1 Introducción .....	125
4.2 Análisis de estudios previos.....	126
4.2.1 Limitaciones y propuesta de mejoras .....	126
4.2.2 Ubicación en la “cartografía” traductológica .....	132
4.3 Hipótesis y objetivos.....	134
4.4 Composición de los corpus .....	136
4.4.1 Textos traducidos.....	136
4.4.1.1 Traducciones semiprofesionales.....	137
4.4.1.2 Traducciones profesionales .....	138
4.4.2 Textos no traducidos.....	144
4.4.2.1 Segmentos LO .....	144
4.4.2.2 Colecciones digitales.....	145
4.4.2.3 Documentos electrónicos .....	149
4.4.3 Niveles de organización.....	157
4.4.3.1 Corpus compilados .....	158
4.4.3.2 Simetría y comparabilidad.....	161
4.5 Preprocesamiento .....	163
4.5.1 Conversión de formatos.....	163
4.5.2 Lematización y anotación.....	165

#### **CAPÍTULO V. INVESTIGACIÓN DE LA HIPÓTESIS DE UNIVERSALIDAD:**

<b>EXPERIMENTOS Y RESULTADOS</b> .....	171
5.1 Introducción .....	171
5.2 Técnicas estadísticas .....	172
5.2.1 Pruebas para comparar grupos muestrales.....	173
5.2.2 Pruebas para medir la similitud entre vectores, cadenas y documentos.....	176
5.3 Estudio A: Simplificación.....	177
5.3.1 Selección de los corpus.....	178
5.3.2 Descripción de los experimentos.....	179

5.3.3	Análisis y discusión de los resultados .....	183
5.4	Estudio B: Convergencia.....	189
5.4.1	Selección de los corpus.....	189
5.4.2	Descripción de los experimentos .....	190
5.4.3	Análisis y discusión de los resultados .....	193
5.4.3.1	Estilo.....	193
5.4.3.2	Sintaxis .....	198
5.5	Estudio C: Transferencia.....	199
5.5.1	Selección de los corpus.....	200
5.5.2	Descripción de los experimentos.....	201
5.5.3	Análisis y discusión de los resultados .....	207
5.5.3.1	Transferencia léxica.....	207
5.5.3.2	Interferencia pronominal .....	209
5.5.3.3	Interferencia sintáctica.....	211
<b>CAPÍTULO VI. CONCLUSIONES.....</b>		<b>215</b>
<b>BIBLIOGRAFÍA.....</b>		<b>225</b>
<b>APÉNDICES .....</b>		<b>263</b>
I.	EJEMPLO DE CODIFICACIÓN EN .XML REUTER .....	263
II.	RELACIÓN DE CORPUS EMPLEADOS.....	275
III.	MARCADORES DISCURSIVOS .....	278
IV.	INDICADORES DE CLÁUSULAS SUBORDINADAS NO FINITAS .....	280
V.	LISTA ANOTADA DE CANDIDATOS A COGNADOS.....	281

## LISTA DE FIGURAS Y TABLAS

Figura 1. Búsqueda para “reach a compromise” .....	82
Figura 2. Interfaz de consulta del corpus LTC.....	87
Figura 3. Compilación de corpus mediante descriptores <i>A</i> y <i>B</i> .....	152
Figura 4. Diseño de corpus en forma de espejo.....	162
Figura 5. Simetría del modelo en espejo.....	163
Figura 6. Porcentaje de oraciones complejas de los corpus T y NT.....	184
Figura 7. Longitud oracional media de los corpus T y NT.....	185
Figura 8. Profundidad del árbol sintáctico de los corpus de T y NT.....	185
Figura 9. Riqueza léxica de los corpus T y NT.....	186
Figura 10. Carga informativa de los corpus T y NT.....	186
Figura 11. N.º de acepciones por palabra de los corpus T y NT.....	187
Figura 12. Densidad léxica y riqueza léxica de los corpus T y NT.....	194
Figura 13. Longitud oracional media de los corpus T y NT.....	194
Figura 14. Proporción de oraciones simples y complejas en los corpus T y NT.....	194
Figura 15. N.º total de pronombres de los corpus T y NT.....	209
Tabla 1. Características de los enfoques <i>corpus-based</i> frente a los <i>corpus-driven</i> .....	53
Tabla 2. Metodología para el establecimiento de unidades semánticas comparables (Tognini-Bonelli, 1996).....	70
Tabla 3. Metodología para el establecimiento de unidades semánticas comparables (Tognini-Bonelli, 2001).....	72
Tabla 4. Documentación de los textos meta.....	138
Tabla 5. Documentación de los textos de origen.....	138
Tabla 6. Pares de lenguas de las memorias de traducción.....	144
Tabla 7. Codificación de documento comparable técnico en español.....	155
Tabla 8. Codificación de documento comparable biomédico en inglés.....	155
Tabla 9. Codificación de documento comparable biomédico en español.....	156
Tabla 10. Codificación de documento de español general.....	157
Tabla 11. Fragmento de MT en .txt.....	164
Tabla 12. Algoritmo de separación (segmentos LO).....	165
Tabla 13. Machine Phrase Tagger 4.6. (inglés).....	167
Tabla 14. Machine Syntax (español).....	168
Tabla 15. Ejemplo de codificación en .xml Fdg.....	169
Tabla 16. Ejemplo de codificación en .xml Reuter.....	170
Tabla 17. Densidad léxica (frecuencia por subsecciones).....	174
Tabla 18. Frecuencias observadas.....	175
Tabla 19. Frecuencias esperadas.....	175

Tabla 20. $X^2$ por parámetros .....	175
Tabla 21. Corpus para simplificación .....	178
Tabla 22. Matriz de simplificación: valores medios para T .....	184
Tabla 23. Matriz de simplificación: valores medios para NT.....	184
Tabla 24. Matriz de simplificación: valores $t$ por rasgos en T y NT .....	188
Tabla 25. Matriz de simplificación: valores $t$ para cada par de corpus comparado .....	188
Tabla 26. Corpus para convergencia.....	190
Tabla 27. Proporción de marcadores discursivos en los corpus T y NT ....	195
Tabla 28. Matriz de convergencia: valores $X^2$ para cada par de corpus T:..	195
Tabla 29. Matriz de convergencia: valores $X^2$ para cada par de corpus NT .....	195
Tabla 30. Matriz estilística: valores $t$ para los corpus CEMOP y CEMOE.	196
Tabla 31. Matriz estilística: valores $t$ para los corpus CEMOE y CETO ....	196
Tabla 32. Matriz estilística: valores $t$ para los corpus CEMOP y CETO ....	197
Tabla 33. Matriz estilística: valores $t$ para los corpus CEMTP y CEMTE..	197
Tabla 34. Matriz estilística: valores $t$ para los corpus CEMTE y CETT .....	197
Tabla 35. Matriz estilística: valores $t$ para los corpus CEMTP y CETT .....	198
Tabla 36. Resultados de la matriz de distancia sintáctica entre vectores ....	199
Tabla 37. Corpus para transferencia léxica.....	203
Tabla 38. Corpus para transferencia sintáctica .....	206
Tabla 39. N.º de sustantivos extraídos de cada corpus .....	207
Tabla 40. Porcentaje estimado de cognados .....	208
Tabla 41. Interferencia pronominal en los corpus de español biomédico T y NT (profesionales) .....	210
Tabla 42. Interferencia pronominal en los corpus de español biomédico T y NT (semiprofesionales).....	210
Tabla 43. Interferencia pronominal en los corpus de español técnico T y NT .....	210
Tabla 44. Interferencia pronominal en los corpus de español especializado T y NT .....	210
Tabla 45. Experimento n.º 1: distancia sintáctica entre vectores.....	212
Tabla 46. Experimento n.º 2: distancia sintáctica entre vectores.....	213