

Etiquetario morfosintáctico del SLI para corpus de lengua gallega: aplicación al corpus paralelo TECTRA

José Luis Aguirre Moreno
Alberto Álvarez Lugrís
Xavier Gómez Guinovart
Seminario de Lingüística Informática
Universidade de Vigo
sli@uvigo.es

Resumen: En este artículo se presenta un etiquetario morfosintáctico completo y normalizado para etiquetar corpus lingüísticos de lengua gallega. La elaboración de este etiquetario, diseñado por el Seminario de Lingüística Informática (SLI) de la Universidad de Vigo siguiendo estrictamente las recomendaciones de EAGLES (Leech y Wilson, 1996), incluye la creación de un etiquetario intermedio que nos permite establecer una correspondencia entre la información gramatical para el gallego codificada en el CLUVI (Corpus Lingüístico de la Universidad de Vigo) y la que se encuentra codificada en el formato estándar de EAGLES en corpus de otras lenguas.

Palabras clave: lingüística de corpus, anotación morfosintáctica, lengua gallega.

Abstract: In this article we present a complete and normalized morphosyntactic tagset for the annotation of linguistic corpora in Galician. The elaboration of this tagset, designed by the Computational Linguistics Group (SLI) of the University of Vigo, following strictly the EAGLES recommendations (Leech and Wilson, 1996), includes the creation of an intermediate tagset that allows us to establish a correspondence between the grammatical information encoded for Galician in the CLUVI (Linguistic Corpus of the University of Vigo) and the information encoded in the EAGLES standard format in corpora of other languages.

Keywords: corpus linguistics, morphosyntactic tagging, Galician language.

1 Introducción¹

El CLUVI (“Corpus Lingüístico da Universidade de Vigo”) es un corpus textual de registros especializados de lengua gallega contemporánea oral y escrita. En su estado

¹ Este trabajo ha sido financiado por la Xunta de Galicia, dentro de los proyectos “Desenvolvemento e aplicación de técnicas de análise lingüístico-computacional de corpus orais e escritos para o procesamento do CLUVI (Corpus Lingüístico da Universidade de Vigo)” (PGIDT01PXI30203PR) y “Estudio e adquisición de recursos básicos de lingüística computacional do galego para a elaboración e mellora de aplicacións informáticas de tecnoloxía lingüística” (ref. PGIDT01TICC06E). Más información en <http://www.uvigo.es/webs/sli/>.

actual de desarrollo, los textos de la sección escrita del CLUVI pertenecen a cuatro registros especializados (de los ámbitos jurídico-administrativo, periodístico, informático y literario) y a tres “combinaciones” lingüísticas con el gallego (monolingüe gallego, traducción gallego-castellano y traducción gallego-inglés), y poseen una extensión total aproximada de 4 millones de palabras. Los textos del CLUVI se reparten en cuatro subcorpus de alrededor de 1 millón de palabras cada uno: el corpus paralelo TECTRA (CLUVI-1) de textos literarios gallego-inglés, el corpus paralelo LEGA (CLUVI-2) de textos jurídico-administrativos gallego-castellano, el corpus monolingüe XIGA (CLUVI-3) de textos sobre informática en gallego, y el corpus monolingüe MEGA (CLUVI-4) de lenguaje de los medios de

comunicación. Los objetivos del procesamiento del CLUVI incluyen su etiquetado morfosintáctico completo y la alineación de las equivalencias oracionales de los dos corpus paralelos, la extracción de información léxica, terminológica y fraseológica de los corpus etiquetados, y la transferencia de resultados para aplicaciones de traducción automática, extracción de información y resumen de documentos, recuperación de información en Internet, y correctores gramaticales para procesamiento de textos.

En esta exposición, presentaremos el etiquetario morfosintáctico utilizado en el SLI (Seminario de Lingüística Informática de la Universidad de Vigo) para la anotación lingüística del CLUVI, mostrando concretamente las soluciones adoptadas para la explotación del corpus paralelo TECTRA. En la exposición se incluirá una descripción del etiquetario morfosintáctico para la lengua gallega elaborado por el SLI conforme a las directrices de EAGLES (Leech y Wilson, 1996; Monachini y Calzolari, 1996, 1999). Asimismo, se presentarán las correspondencias entre el etiquetario SLI del gallego y el etiquetario intermedio propuesto por EAGLES (Leech y Wilson, 1996) como representación lingüísticamente neutral del conjunto de pares atributo-valor que describen la información lingüística codificada en las etiquetas.

2 El corpus TECTRA

El corpus TECTRA (Álvarez Lugrís, 2001) contiene catorce novelas en lengua inglesa con sus correspondientes traducciones al gallego, con un total de 1.127.044 palabras, de las cuales 551.878 corresponden a las catorce traducciones gallegas y 575.166 a los catorce originales ingleses. En los anexos de este trabajo puede verse la relación completa de originales y traducciones recopiladas en este corpus.

En una fase inicial, los textos del TECTRA fueron adquiridos mediante digitalización de las obras impresas, posteriormente revisados (para corregir los inevitables errores en el reconocimiento óptico de caracteres) y finalmente etiquetados estructuralmente en XML, con información sobre obra, autor/a, traductor/a, capítulo, sección, página, párrafo y oración.

El procesamiento lingüístico-computacional del TECTRA, dentro del proyecto CLUVI del

SLI, se realiza en tres etapas sucesivas: etiquetado, alineación y explotación. En la fase inicial de etiquetado, los textos del TECTRA se anotan morfosintácticamente mediante un sistema probabilístico de etiquetado basado en el programa informático de etiquetación y desambiguación Trigram's Tags (TnT), desarrollado en el Departamento de Lingüística Computacional de la Universidad de Saarland (en Saarbrücken, Alemania) para cuyo uso disponemos de una licencia de investigación. El etiquetario para los textos en inglés es el estándar de EAGLES (Leech y Wilson, 1996). Para los textos en gallego, se emplea el etiquetario morfosintáctico elaborado por el SLI conforme a las directrices de EAGLES y descrito en el siguiente apartado del trabajo. De manera paralela al etiquetado, se elabora un léxico computacional del gallego con las especificaciones morfosintácticas definidas en el etiquetario del SLI.

Con el fin de permitir desarrollos y aplicaciones basadas en las correspondencias gramaticales entre los textos gallegos e ingleses etiquetados, se ha establecido la correspondencia entre el etiquetario SLI para el gallego y el etiquetario EAGLES para el inglés mediante el etiquetario intermedio estándar propuesto también por EAGLES. Tanto la alineación como la explotación del TECTRA para la extracción de información léxica se benefician de la anotación morfosintáctica y de su correspondencia bilingüe. A continuación, centraremos la exposición de este trabajo en la descripción del etiquetario SLI para el gallego y del etiquetario intermedio.

3 Etiquetario morfosintáctico del SLI

3.1 Diseño del etiquetario

Para la creación del etiquetario del SLI, en lo que respecta a la identificación de los fenómenos gramaticales relevantes en lengua gallega, nos basamos principalmente en la descripción gramatical de Álvarez, Regueira y Monteagudo (1986). Al mismo tiempo, adoptamos las propuestas de EAGLES en lo relativo a las categorías gramaticales y rasgos morfosintácticos que conviene distinguir. Para ello, no nos limitamos a seguir las directrices generales de EAGLES, sino que aplicamos estrictamente el esquema de atributos y valores recomendado por Leech y Wilson (1996), adecuándolo al gallego de modo análogo a lo

realizado para otras lenguas, como el italiano y el alemán (Teufel, 1996).

Partiendo de la división en categorías principales, obligatorias según Leech y Wilson (1996), determinamos los rasgos morfosintácticos aplicables al gallego, prescindiendo de los que no lo son y añadiendo atributos o valores cuando resulta necesario. Así, no incluimos en el etiquetario el valor “neutro” del atributo “género” para los nombres comunes, ya que no es aplicable al gallego; mientras que debemos añadir el valor “antepretérito” al atributo “tiempo” de los verbos en modo indicativo, para recoger la forma verbal no compuesta del gallego que expresa lo que se suele denominar pluscuamperfecto: “eu cantara” (“yo había cantado”).

Otro aspecto fundamental en el diseño del etiquetario SLI del gallego es el establecimiento de las correspondencias con el etiquetario intermedio de EAGLES. El etiquetario intermedio es una representación lingüísticamente neutral que describe los rasgos lingüísticos (descritos en forma de pares atributo-valor) reflejados en un etiquetario, de modo que se puedan hacer corresponder fácilmente con las marcas de otro conjunto de etiquetas (Leech y Wilson, 1996). El etiquetario intermedio nos permite trabajar con etiquetas definidas según la terminología gramatical propia de la lengua gallega, y convertirlas automáticamente a los rasgos definidos en el estándar de EAGLES. Aplicado al corpus paralelo TECTRA, el etiquetario intermedio nos permite establecer de manera automática e inequívoca una correspondencia entre la información gramatical de los textos en gallego y la de los textos en inglés, lo que nos posibilita la explotación de estas correspondencias en procesos lingüístico-computacionales posteriores al etiquetado, como la alineación de los bitextos o la extracción automática de información léxica bilingüe contextual y fraseológica. De manera más general, la correspondencia del etiquetario del gallego con el etiquetario intermedio nos permite la reutilización de los textos etiquetados en aplicaciones adaptadas al estándar EAGLES.

3.2 Presentación del etiquetario

Para cada categoría se muestran tres tablas:

1. Tabla de relaciones jerárquicas entre subcategorías, en la que también se especifican

las restricciones en la aplicación de atributos y valores. Esta descripción está basada en las recomendaciones de EAGLES recogidas por Monachini y Calzolari (1999), si bien tratamos de mostrar la jerarquía entre las subcategorías y las restricciones de aparición de unos valores con otros de una manera unificada, en una sola tabla para toda la categoría, en vez de emplear una tabla por cada atributo. En cada columna está representado un atributo. En cada fila se representan las combinaciones de los distintos valores admitidos. La jerarquía entre subcategorías se muestra mediante la agrupación gráfica de los valores en la tabla. Cuando un valor no es aplicable, se deja la casilla correspondiente en blanco. Los atributos o valores específicos de la lengua gallega, añadidos al etiquetario, aparecen resaltados en negrita.

2. Relación de atributos y valores siguiendo la numeración de Leech y Wilson (1996), tal y como queda para su aplicación al gallego, incluyendo los rasgos que son opcionales en el estándar si los adoptamos para esta lengua. Se muestran entre paréntesis los atributos y valores que no se consideran aplicables o relevantes para el gallego y en negrita los específicos de esta lengua, añadidos al etiquetario.

3. Inventario completo de etiquetas para cada categoría, incluyendo una palabra de ejemplo, la etiqueta empleada en nuestra codificación, la descripción gramatical correspondiente a la etiqueta y, por último, la codificación de la etiqueta intermedia correspondiente según la numeración de atributos y valores expresada en la relación anterior. De nuevo, los dígitos de valores añadidos para el gallego se muestran en negrita.

A continuación, mostramos la relación completa de etiquetas del etiquetario SLI para corpus de lengua gallega, agrupadas por categorías siguiendo el orden y el formato de descripción estándar en las recomendaciones de EAGLES.

3.2.1 Categoría nombre

Cat = nombre			
Atributos			
	Tipo	Género	Número
Valores	Común	masc, fem, (neu)	sg, pl
	Propio	masc, fem	sg, pl

Tabla 1: Jerarquía subcategorías nominales

(i) Tipo: 1. Común. 2. Propio. (ii) Género: 1. Masculino. 2. Femenino. (3. Neutro.) (iii) Número: 1. Singular. 2. Plural. (iv) (Caso: 1. Nominativo...)
--

Figura 1: Atributos y valores nominales

Ejemplo	Etiqueta	Descripción	Etiqueta intermedia
<i>can</i>	NCMS	común m sg	N1110
<i>folia</i>	NCFS	común f sg	N1210
<i>homes</i>	NCMP	común m p	N1120
<i>mulleres</i>	NCFP	común f p	N1220
<i>Aldán</i>	NPMS	propio m sg	N2110
<i>Antía</i>	NPFS	propio f sg	N2210
<i>Ancaras</i>	NPMP	propio m p	N2120
<i>Burgas</i>	NFPF	propio f p	N2220

Tabla 2: Inventario etiquetas nominales

3.2.2 Categoría verbo

Cat = verbo							
Atributos							
	Fin	Modo/ forma	Tiempo	Pers	Nm	Gn	
Valores	Finito	Ind.	pres, imperf, fut, pret, antepret	1,2, 2c,3	sg, pl		
		Subj	pres, imperf, fut	1,2, 2c,3	sg, pl		
		Imper		2, 2c	sg, pl		
		Cond		1,2, 2c,3	sg, pl		
	No finito	Info					
		Inf conj			1,2, 2c,3	sg, pl	
		Part				sg, pl	m, f
		Ger					
		(Supino)					

Tabla 3: Jerarquía subcategorías verbales

(i) Persona: 1. Primera. 2. Segunda. 3. Tercera. 4. Cortesía. (ii) Género: 1. Masculino. 2. Femenino. (3. Neutro.) (iii) Número: 1. Singular. 2. Plural. (iv) Finitud: 1. Finito. 2. No finito. (v) Forma verbal / Modo: 1. Indicativo. 2. Subjuntivo. 3. Imperativo. 4. Condicional. 5. Infinitivo. 6. Participio. 7. Gerundio. (8. Supino). 9. Infinitivo conjugado. (vi) Tiempo: 1. Presente. 2. Imperfecto. 3. Futuro. 4. Pretérito. 5. Antepretérito. (vii) Voz: 1. Activa. 2. Pasiva. (viii) (Estatus: 1. Principal. 2. Auxiliar.)

Figura 2: Atributos y valores verbales

Para facilitar la consulta de la Tabla 4, dividimos el inventario de etiquetas verbales en grupos, según el tiempo y modo verbal:

Indicativo presente			
<i>collo</i>	VIPRS1	1s	V10111110
<i>colles</i>	VIPRS2	2s	V20111110
<i>colle</i>	VIPRS2C	2scor	V40111110
<i>colle</i>	VIPRS3	3s	V30111110
<i>collemos</i>	VIPRP1	1p	V10211110
<i>colledes</i>	VIPRP2	2p	V20211110
<i>collen</i>	VIPRP2C	2pcor	V40211110
<i>collen</i>	VIPRP3	3p	V30211110

Indicativo copretérito (= imperfect)			
<i>collia</i>	VICPS1	1s	V10111210
<i>collias</i>	VICPS2	2s	V20111210
<i>collia</i>	VICPS2C	2scor	V40111210
<i>collia</i>	VICPS3	3s	V30111210
<i>colliamos</i>	VICPP1	1p	V10211210
<i>colliades</i>	VICPP2	2p	V20211210
<i>collían</i>	VICPP2C	2pcor	V40211210
<i>collían</i>	VICPP3	3p	V30211210

Indicativo pretérito (= past)			
<i>collín</i>	VIPES1	1s	V10111410
<i>colliches</i>	VIPES2	2s	V20111410
<i>colleu</i>	VIPES2C	2scor	V40111410
<i>colleu</i>	VIPES3	3s	V30111410
<i>collemos</i>	VIPEP1	1p	V10211410
<i>collestes</i>	VIPEP2	2p	V20211410
<i>colleron</i>	VIPEP2C	2pcor	V40211410
<i>colleron</i>	VIPEP3	3p	V30211410

Indicativo antepretérito			
<i>collera</i>	VIAPS1	1s	V10111510
<i>colleras</i>	VIAPS2	2s	V20111510
<i>collera</i>	VIAPS2C	2scor	V40111510
<i>collera</i>	VIAPS3	3s	V30111510
<i>colleramos</i>	VIAPP1	1p	V10211510
<i>collerades</i>	VIAPP2	2p	V20211510
<i>colleran</i>	VIAPP2C	2pcor	V40211510
<i>colleran</i>	VIAPP3	3p	V30211510

<i>collerdes</i>	VSFUP2	2p	V20212310
<i>colleren</i>	VSFUP2C	2pcor	V40212310
<i>colleren</i>	VSFUP3	3p	V30212310

Imperativo			
<i>coller</i>	VIMPS2	2s	V20113010
<i>colla</i>	VIMPS2C	2scor	V40113010
<i>collede</i>	VIMPP2	2p	V20213010
<i>collan</i>	VIMPP2C	2pcor	V40213010

Indicativo futuro			
<i>collerei</i>	VIFUS1	1s	V10111310
<i>collerás</i>	VIFUS2	2s	V20111310
<i>collerá</i>	VIFUS2C	2scor	V40111310
<i>collerá</i>	VIFUS3	3s	V30111310
<i>colleremos</i>	VIFUP1	1p	V10211310
<i>colleredes</i>	VIFUP2	2p	V20211310
<i>collerán</i>	VIFUP2C	2pcor	V40211310
<i>collerán</i>	VIFUP3	3p	V30211310

Infinitivo conjugado			
<i>coller</i>	VINFCS1	1s	V10129010
<i>colleres</i>	VINFCS2	2s	V20129010
<i>coller</i>	VINFCS2C	2scor	V40129010
<i>coller</i>	VINFCS3	3s	V30129010
<i>collermos</i>	VINFCP1	1p	V10229010
<i>collerdes</i>	VINFCP2	2p	V20229010
<i>colleren</i>	VINFCP2C	2pcor	V40229010
<i>colleren</i>	VINFCP3	3p	V30229010

Indicativo pospretérito (= conditional)			
<i>collería</i>	VIPPS1	1s	V10114010
<i>collerías</i>	VIPPS2	2s	V20114010
<i>collería</i>	VIPPS2C	2scor	V40114010
<i>collería</i>	VIPPS3	3s	V30114010
<i>colleríamos</i>	VIPPP1	1p	V10214010
<i>colleríades</i>	VIPPP2	2p	V20214010
<i>collerían</i>	VIPPP2C	2pcor	V40214010
<i>collerían</i>	VIPPP3	3p	V30214010

Formas no finitas			
<i>coller</i>	VINF	inf	V00025010
<i>collendo</i>	VXER	xer	V00027010
<i>collido</i>	VPARMS	part m sg	V01126010
<i>collida</i>	VPARFS	part f sg	V02126010
<i>collidos</i>	VPARMP	part m pl	V01226010
<i>collidas</i>	VPARFP	part f pl	V02226010

Subjuntivo presente			
<i>colla</i>	VSPRS1	1s	V10112110
<i>collas</i>	VSPRS2	2s	V20112110
<i>colla</i>	VSPRS2C	2scor	V40112110
<i>colla</i>	VSPRS3	3s	V30112110
<i>collamos</i>	VSPRP1	1p	V10212110
<i>collades</i>	VSPRP2	2p	V20212110
<i>collan</i>	VSPRP2C	2pcor	V40212110
<i>collan</i>	VSPRP3	3p	V30212110

Subjuntivo pretérito (= subjunctive imperfect)			
<i>collese</i>	VSPES1	1s	V10112210
<i>colleses</i>	VSPES2	2s	V20112210
<i>collese</i>	VSPES2C	2scor	V40112210
<i>collese</i>	VSPES3	3s	V30112210
<i>collesemos</i>	VSPEP1	1p	V10212210
<i>collesedes</i>	VSPEP2	2p	V20212210
<i>collesen</i>	VSPEP2C	2pcor	V40212210
<i>collesen</i>	VSPEP3	3p	V30212210

Subjuntivo futuro			
<i>coller</i>	VSFUS1	1s	V10112310
<i>colleres</i>	VSFUS2	2s	V20112310
<i>coller</i>	VSFUS2C	2scor	V40112310
<i>coller</i>	VSFUS3	3s	V30112310
<i>collermos</i>	VSFUP1	1p	V10212310

Tabla 4: Inventario etiquetas verbales²

3.2.3 Categoría adjetivo

Cat = adjetivo				
Valores	Atributos			
	Grado	Gén	Núm	Forma
	positivo, superlativo, comparativo	m, f	sg, pl	
positivo	m, f	sg		apocopado

Tabla 5: Jerarquía subcategorías adjetivo

² En la tercera columna de la Tabla 4, la abreviatura “1s” indica “primera persona singular”; “1p”, “primera persona plural”; “2scor”, “segunda persona singular de cortesía” (“vostede”); y “2pcor”, “segunda persona plural de cortesía” (“vostedes”).

(i) Grado: 1. Positivo. 2. Comparativo. 3. Superlativo.
(ii) Género: 1. Masculino. 2. Femenino. (3. Neutro.)
(iii) Número: 1. Singular. 2. Plural.
(iv) (Caso: 1. Nominativo ...)
(vi) Forma: 1. Apocopado.

Figura 3: Atributos y valores adjetivo

<i>novo</i>	AXMS	adj ms	AJ11100
<i>nova</i>	AXFS	adj fs	AJ12100
<i>novos</i>	AXMP	adj mp	AJ11200
<i>novas</i>	AXFP	adj fp	AJ12200
<i>novísimo</i>	AXSMS	adj sup ms	AJ31100
<i>novísima</i>	AXSFS	adj sup fs	AJ32100
<i>novísimos</i>	AXSMP	adj sup mp	AJ31200
<i>novísimas</i>	AXSFP	adj sup fp	AJ32200
<i>mellor</i>	AXCMS	adj cp ms	AJ21100
<i>mellor</i>	AXCFS	adj cp fs	AJ22100
<i>mellores</i>	AXCMP	adj cp mp	AJ21200
<i>mellores</i>	AXCFP	adj cp fp	AJ22200
<i>gran</i>	AXAPMS	adj apo ms	AJ11101
<i>gran</i>	AXAPFS	adj apo fs	AJ12101

Tabla 6: Inventario etiquetas adjetivo³

3.2.4 Categoría pronombre

La Tabla 7 que se incluye a continuación, con la jerarquía de atributos y valores pronominales, no incluye la información relativa a los pronombres personales que, por su complejidad, se muestra por separado en la Tabla 8. Para facilitar la consulta de la Tabla 8 que se muestra más adelante, con las restricciones de aplicación entre atributos y valores de los pronombres personales, se incluye una columna adicional con las formas que se describen.

Cat = pronombre							
Atributos							
	Tipo-pron	Caso	Persona	Género	Número	Posesivo ⁴	Cortesía
Valores	Dem		1,2,3	m, f, n	sg, pl		
	Indef			m, f	sg, pl		
	Poses		1,2,3	m, f	sg, pl	sg, pl	cortés
	Interr			m, f	sg, pl		
	Relat ⁵			m, f	sg, pl		

Tabla 7: Jerarquía subcat. pronominales

(i) Persona: 1. Primera. 2. Segunda. 3. Tercera.
(ii) Género: 1. Masculino. 2. Femenino. 3. Neutro.
(iii) Número: 1. Singular. 2. Plural.
(iv) Posesivo: 1. Singular. 2. Plural.
(v) Caso: 1. Nominativo. (2. Genitivo.) 3. Dativo. 4. Acusativo. (5. No genitivo.) 6. Oblicuo.
(vi) Categoría: 1. Pronombre. (2. Determinante ⁶) (3. Ambos.)
(vii) Tipo-pron.: 1. Demostrativo 2. Indefinido 3. Posesivo. 4. Interr./Relat. 5. Personal / reflexivo.
(viii) (Tipo-det: 1. Demostrativo ...)
(ix) Tipo especial de pron: 1. Personal. 2. Reflexivo. (3. Recíproco.) 4. Exclusivo.
(x) Tipo qu-: 1. Interrogativo. 2. Relativo. 3. Exclamativo.
(xi) Cortesía: 1. Cortés 2. Familiar.
(xii) Tonicidad: 1. Tónico. 2. Átono.

Figura 4: Atributos y valores pronominales

³ En el grado comparativo y en las formas apocopadas de los adjetivos, optamos por distinguir los valores de género aunque la distinción no se refleje morfológicamente.

⁴ Atributo que se refiere a la persona poseedora: “miña” (sg)/ “nosa” (pl) (“mía / nuestra”).

⁵ Optamos por distinguir en el etiquetario entre interrogativo y relativo a pesar de las coincidencias morfológicas.

⁶ Optamos por no emplear la distinción entre pronombre y determinante, siguiendo la descripción gramatical del gallego en la que nos basamos (Álvarez, Regueira y Monteagudo, 1986) para la elaboración del etiquetario.

Cat = pronombre, tipo-pron = personal / reflexivo												
Atributos												
Valores	Tónico	Tonicidad	Caso	Tipo-pron especial	Pers.	Gén.	Núm.	Cortésia	Formas			
		Nom	Obl	Nom	Obl	pers	1		sg		<i>eu</i>	
							2		sg	fam.	<i>ti</i>	
							1	2		pl		<i>nós</i> <i>vós</i>
							2		sg	cor-tés	<i>vostede</i> <i>vostedes</i>	
	3						m f n	sg pl		<i>el</i> <i>ela</i> <i>elo</i> ⁷		
			pers, exclu-sivo		2	m f	pl		<i>vosoutros</i> <i>vosoutras</i>			
	Átono	Obl	Dat	Ac	pers	1		sg		<i>me</i> <i>nos</i>		
						2		pl		<i>vos</i>		
						2		sg	fam	<i>che</i>		
2							sg	cor-tés	<i>lle</i> <i>lles</i>			
3							sg		<i>lle</i> <i>lles</i>			
		pers		2		sg	fam	<i>te</i>				
		pers		2	m f	sg pl	cor-tés	<i>o, a</i> <i>os, as</i>				
		pers		3	m f	sg pl		<i>o, a</i> <i>os, as</i>				
		refl		2		sg	cor-tés	<i>se</i>				
		refl		3		sg		<i>se</i>				

Tabla 8: Jerarquía subcat. pron. pers.

Para simplificar el manejo de la Tabla 9, dividimos el inventario de etiquetas pronominales en grupos, según el tipo de pronombre y sus características:

⁷ Forma antigua y en desuso, pero que recogemos en el etiquetario por aparecer en gallego contemporáneo en algunas frases hechas (“estar a elo”) (Álvarez, Regueira y Monteagudo, 1986, páginas 160-161).

⁸ En el etiquetario, sólo incluimos el valor “reflexivo” para las formas pronominales “si” y “se”.

Pronombres demostrativos			
<i>este</i>	PDMS1	m sg 1 ^a	PD111001100000
<i>ese</i>	PDMS2	m sg 2 ^a	PD211001100000
<i>aquel</i>	PDMS3	m sg 3 ^a	PD311001100000
<i>estes</i>	PDMP1	m pl 1 ^a	PD112001100000
<i>eses</i>	PDMP2	m pl 2 ^a	PD212001100000
<i>aqueles</i>	PDMP3	m pl 3 ^a	PD312001100000
<i>esta</i>	PDFS1	f sg 1 ^a	PD121001100000
<i>esa</i>	PDFS2	f sg 2 ^a	PD221001100000
<i>aquela</i>	PDFS3	f sg 3 ^a	PD321001100000
<i>estas</i>	PDFP1	f pl 1 ^a	PD122001100000
<i>esas</i>	PDFP2	f pl 2 ^a	PD222001100000
<i>aquelas</i>	PDFP3	f pl 3 ^a	PD322001100000
<i>isto</i>	PDNS1	neu sg 1 ^a	PD131001100000
<i>iso</i>	PDNS2	neu sg 2 ^a	PD231001100000
<i>aquilo</i>	PDNS3	neu sg 3 ^a	PD331001100000

Pronombres posesivos ⁹			
<i>meu</i>	PVS1MS	Pr: sg 1 ^a Pd: m sg	PD111101300000
<i>miña</i>	PVS1FS	Pr: sg 1 ^a Pd: f sg	PD121101300000
<i>meus</i>	PVS1MP	Pr: sg 1 ^a Pd: m pl	PD112101300000
<i>miñas</i>	PVS1FP	Pr: sg 1 ^a Pd: f pl	PD122101300000
<i>teu</i>	PVS2MS	Pr: sg 2 ^a Pd: m sg	PD211101300000
<i>túa</i>	PVS2FS	Pr: sg 2 ^a Pd: f sg	PD221101300000
<i>teus</i>	PVS2MP	Pr: sg 2 ^a Pd: m pl	PD212101300000
<i>túas</i>	PVS2FP	Pr: sg 2 ^a Pd: f pl	PD222101300000
<i>seu</i>	PVS3MS	Pr: sg 3 ^a Pd: m sg ¹⁰	PD311101300000
<i>seu</i>	PVP3MS	Pr: pl 3 ^a Pd: m pl ¹¹	PD311201300000
<i>seu</i>	PVS2CMS	Pr: sg 2 ^a cort Pd: m sg ¹²	PD211101300010
<i>seu</i>	PVP2CMS	Pr: pl 2 ^a cort Pd: m sg ¹³	PD111201300010
<i>súa</i>	PVS3FS	Pr: sg 3 ^a Pd: f sg ¹⁴	PD321101300000

⁹ En la tercera columna de la Tabla, para los pronombres posesivos, la abreviatura “Pr” se refiere al concepto de “Poseedor”, y “Pd” a lo “Poseído”.

¹⁰ Ej.: “Antía fixo o seu traballo” (“Antía hizo su trabajo”).

¹¹ Ej.: “As formigas afanábanse no seu traballo” (“Las hormigas se afanaban en su trabajo”).

¹² Ej.: “Dígame o seu enderezo, por favor” (“Dígame su dirección, por favor”).

¹³ Ej.: “Suban ó seu coche” (“Suban a su coche”).

<i>súa</i>	PVP3FS	Pr: pl 3 ^a Pd: f sg ¹⁵	PD321201300000
<i>súa</i>	PVS2CFS	Pr: sg 2 ^a cort Pd: f sg ¹⁶	PD221101300010
<i>súa</i>	PVP2CFS	Pr: pl 2 ^a cort Pd: f sg ¹⁷	PD221201300010
<i>seus</i>	PVS3MP	Pr: sg 3 ^a Pd: m pl ¹⁸	PD312101300000
<i>seus</i>	PVP3MP	Pr: pl 3 ^a Pd: m pl ¹⁹	PD312201300000
<i>seus</i>	PVS2CMP	Pr: sg 2 ^a cort Pd: m pl ²⁰	PD212101300010
<i>seus</i>	PVP2CFP	Pr: pl 2 ^a cort Pd: m pl ²¹	PD212201300010
<i>súas</i>	PVS3FP	Pr: sg 3 ^a Pd: f pl ²²	PD322101300000
<i>súas</i>	PVP3FP	Pr: pl 3 ^a Pd: f pl ²³	PD322201300000
<i>súas</i>	PVS2CFP	Pr: sg 2 ^a cort Pd: f pl ²⁴	PD222101300010
<i>súas</i>	PVP2CFP	Pr: pl 2 ^a cort Pd: f pl ²⁵	PD222201300010
<i>noso</i>	PVP1MS	Pr: pl 1 ^a Pd: m sg	PD111201300000
<i>nosa</i>	PVP1FS	Pr: pl 1 ^a Pd: f sg	PD121201300000
<i>nosos</i>	PVP1MP	Pr: pl 1 ^a Pd: m pl	PD112201300000

¹⁴ Ej.: “Aldán está na súa casa” (“Aldán está en su casa”).

¹⁵ Ej.: “As formigas afanábanse na súa tarefa” (“Las formigas se afanaban en su tarea”).

¹⁶ Ej.: “Agarde un intre que o levo á súa casa” (“Espere un instante que lo llevo a su casa”).

¹⁷ Ej.: “Váianse á súa casa” (“Váyanse a su casa”).

¹⁸ Ej.: “Dálle estes libros a Lúa: son seus” (“Dale estos libros a Lúa: son suyos”).

¹⁹ Ej.: “Eles non poden levar eses libros: non son seus” (“Ellos no pueden llevar esos libros: no son suyos”).

²⁰ Ej.: “Leve vostede estes libros: son seus” (“Llévese usted estos libros: son suyos”).

²¹ Ej.: “Leven vostedes estes libros: son seus” (“Llévense ustedes estos libros: son suyos”).

²² Ej.: “Dálle estas cartas a Aldán: son súas” (“Dale estas cartas a Aldán: son suyas”).

²³ Ej.: “Dálle estas cartas a teus pais: son súas” (“Dale estas cartas a tus padres: son suyas”).

²⁴ Ej.: “Leve vostede estas cartas: son súas” (“Llévese usted estas cartas: son suyas”).

²⁵ Ej.: “Leven vostedes estas cartas: son súas” (“Llévense ustedes estas cartas: son suyas”).

<i>nosas</i>	PVP1FP	Pr: pl 1 ^a Pd: f pl	PD122201300000
<i>voso</i>	PVP2MS	Pr: pl 2 ^a Pd: m sg	PD211201300000
<i>vosa</i>	PVP2FS	Pr: pl 2 ^a Pd: f sg	PD221201300000
<i>vosos</i>	PVP2MP	Pr: pl 2 ^a Pd: m pl	PD212201300000
<i>vosas</i>	PVP2FP	Pr: pl 2 ^a Pd: f pl	PD222201300000

Pronombres indefinidos

<i>algo</i>	PI	(invar.)	PD000003220000
<i>algún</i>	PIMS	m sg	PD011001200000
<i>algunha</i>	PIFS	f sg	PD021001200000
<i>algúns</i>	PIMP	m pl	PD012001200000
<i>algunhas</i>	PIFP	f pl	PD022001200000

Pronombres relativos

<i>quen</i>	PR	(invar.)	PD000001400200
<i>cal (o)</i>	PRMS	m sg	PD011001400200
<i>cal (a)</i>	PRFS	f sg	PD021001400200
<i>cales (os)</i>	PRMP	m sg	PD012001400200
<i>cales (as)</i>	PRFP	m sg	PD022001400200

Pronombres interrogativos

<i>quen</i>	PIN	(invar.)	PD000001400100
<i>cal (o)</i>	PINMS	m sg	PD011001400100
<i>cal (a)</i>	PINFS	f sg	PD021001400100
<i>cales (os)</i>	PINMP	m sg	PD012001400100
<i>cales (as)</i>	PINFP	m sg	PD022001400100

Pronombres personales tónicos: formas rectas

<i>eu</i>	PPS1	sg 1 ^a	PD101011501001
<i>ti</i>	PPS2	sg 2 ^a	PD201011501001
<i>vostede</i>	PPS2C	sg 2 ^a cort	PD201011501011
<i>el</i>	PPMS3	m sg 3 ^a	PD311011501001
<i>ela</i>	PPFS3	f sg 3 ^a	PD321011501001
<i>elo</i>	PPNS3	neu sg 3 ^a	PD331011501001
<i>nós</i>	PPP1	pl 1 ^a	PD102011501001
<i>nosoutros</i>	PPMP1E	m pl 1 ^a excl	PD112011504001
<i>nosoutras</i>	PPFP1E	f pl 1 ^a excl	PD122011504001
<i>vós</i>	PPP2	pl 2 ^a	PD202011501001
<i>vosoutros</i>	PPMP2E	m pl 2 ^a excl	PD212011504001
<i>vosoutras</i>	PPFP2E	f pl 2 ^a excl	PD222011504001
<i>vostedes</i>	PPP2C	pl 2 ^a cort	PD202011501011
<i>eles</i>	PPMP3	m pl 3 ^a	PD312011501001
<i>elas</i>	PPFP3	m pl 3 ^a	PD322011501001

Pronombres personales tónicos: formas oblicuas			
<i>min</i>	PPS1O	sg 1ª obl	PD101061501001
<i>ti</i>	PPS2O	sg 2ª obl	PD201061501001
<i>vostede</i>	PPS2CO	sg 2ª cort obl	PD201061501011
<i>el</i>	PPMS3O	m sg 3ª obl	PD311061501001
<i>ela</i>	PPFS3O	f sg 3ª obl	PD321061501001
<i>elo</i>	PPNS3O	neu sg 3ª obl	PD331061501001
<i>nós</i>	PPP1O	pl 1ª obl	PD102061501001
<i>nosoutros</i>	PPMP1EO	m pl 1ª excl obl	PD112061504001
<i>nosoutras</i>	PPFP1EO	f pl 1ª excl obl	PD122061504001
<i>vós</i>	PPP2O	pl 2ª obl	PD202061501001
<i>vosoutros</i>	PPMP2EO	m pl 2ª excl obl	PD212061504001
<i>vosoutras</i>	PPFP2EO	f pl 2ª excl obl	PD222061504001
<i>vostedes</i>	PPP2CO	pl 2ª cort obl	PD202061501011
<i>eles</i>	PPMP3O	m pl 3ª obl	PD312061501001
<i>elas</i>	PPFP3O	f pl 3ª obl	PD322061501001
<i>si</i>	PPOR	obl refl	PD300061502001

<i>vos</i>	PPP2A	át pl 2ª	PD2020[3 4]1501002
<i>lles</i>	PPP2CAD	át pl 2ª cort dat	PD202031501012
<i>lles</i>	PPP3AD	át pl 3ª dat	PD302031501002
<i>os</i>	PPMP3AA	át m pl 3ª acus	PD312041501002
<i>as</i>	PPFP3AA	át f pl 3ª acus	PD322041501002
<i>se</i>	PPP3AR	át pl 3ª refl	PD302001502002

Tabla 9: Inventario etiquetas pronominales

3.2.5 Categoría numeral

En el etiquetario gallego tratamos los numerales como un tipo de pronombres. Sin embargo, al hacer la correspondencia con el etiquetario intermedio lo consideramos una categoría independiente, siguiendo las recomendaciones de EAGLES.

Cat = numeral				
Atributos				
Valores	Función	Tipo	Género	Número
	Pronombre		Cardinal	m, f
		Ordinal	m, f	sg, pl

Tabla 10: Jerarquía subcategorías numeral

Pronombres personales átonos			
<i>me</i>	PPS1A	át sg 1ª	PD1010[3 4]1501002
<i>che</i>	PPS2AD	át sg 2ª dat	PD201031501022
<i>te</i>	PPS2AA	át sg 2ª acus	PD201041501022
<i>lle</i>	PPS2CAD	át sg 2ª cort dat	PD201031501012
<i>lle</i>	PPS3AD	át sg 3ª dat	PD301031501002
<i>o</i>	PPMS3AA	át m sg 3ª acus	PD311041501002
<i>o</i>	PPMS2CAA	át m sg 2ª cort acus	PD211041501012
<i>a</i>	PPFS3AA	át f sg 3ª acus	PD321041501002
<i>a</i>	PPFS2CAA	át f sg 2ª cort acus	PD221041501012
<i>se</i>	PPS3AR	át sg 3ª refl	PD301001502002
<i>se</i>	PPS2CAR	át sg 2ª cort refl	PD200001502012
<i>nos</i>	PPP1A	át pl 1ª	PD1020[3 4]1501002

- (i) Tipo: 1. Cardinal. 2. Ordinal.
(ii) Género: 1. Masculino. 2. Femenino.
(iii) Número: 1. Singular. 2. Plural.
(iv) (Caso: 1. Nominativo...)
(v) Función: 1. Pronombre. (2. Determinante.) (3. Adjetivo.)

Figura 5: Atributos y valores numeral

<i>un</i>	PNCMS	card m sg	NU11101
<i>unha</i>	PNCFS	card f sg	NU12101
<i>dous</i>	PNCMP	card m pl	NU11201
<i>dúas</i>	PNCFP	card f pl	NU12201
<i>tres</i>	PNC	card (invar.)	NU10001
<i>primeiro</i>	PNOMS	ord m sg	NU21101
<i>primeira</i>	PNOFS	ord f sg	NU22101
<i>primeiros</i>	PNOMP	ord m pl	NU21201
<i>primeiras</i>	PNOFP	ord f pl	NU22201

Tabla 11: Inventario etiquetas numeral

3.2.6 Categoría artículo

Cat = artículo			
Atributos			
Valores	Tipo	Género	Número
	Definido	m, f, (n)	sg, pl
	Indefinido	m, f, (n)	sg, pl

Tabla 12: Jerarquía subcategorías artículo

- (i) Tipo de artículo: 1. Definido. 2. Indefinido.
 (ii) Género: 1. Masculino. 2. Femenino.
 (iii) Número: 1. Singular. 2. Plural.
 (iv) (Caso: 1. Nominativo...)

Figura 6: Atributos y valores artículo

<i>o</i>	ARDMS	det m sg	AT1110
<i>a</i>	ARDFS	det f sg	AT1210
<i>os</i>	ARDMP	det m pl	AT1120
<i>as</i>	ARDFP	det f pl	AT1220
<i>un</i>	ARIMS	indet m sg	AT2110
<i>unha</i>	ARIFS	indet f sg	AT2210
<i>unos</i>	ARIMS	indet m pl	AT2120
<i>unhas</i>	ARIMP	indet f pl	AT2220

Tabla 13: Inventario etiquetas artículo

3.2.7 Otras categorías gramaticales

aquí	ADV	Adverbio	AV0
con	PREP	Preposición	AP1
e	CONX	Conjunción	C0
ah	INTERX	Interjección	I
non	NEG	Partícula negativa	U2

Tabla 14: Inventario etiquetas otras categorías

3.2.8 Categorías residuales

Incluimos a continuación, en la Tabla 15, una lista de etiquetas de fenómenos lingüísticos denominados por EAGLES “residuales”, que no se refieren propiamente a categorías morfosintácticas, pero que resultan de gran utilidad en la anotación de corpus textuales de lengua real.

pizza	REX	Palabra extranjera	R100
x = 2h.p	RFO	Fórmula	R200
@	RSI	Símbolo	R300
ONU	RAC	Acrónimo	R400
D.	RAB	Abreviatura	R500
???	RSC	Sin clasificar	R600

Tabla 15: Inventario etiquetas cat. residuales

3.2.9 Etiquetas para puntuación

Incluimos aquí también, a modo de apéndice de este capítulo, las etiquetas empleadas en el CLUVI para la puntuación, con la intención de proporcionar un etiquetario completo útil para la lingüística corpus del gallego.

.	PUNTO	Punto y seguido
.	PUNTOAP	Punto y aparte
,	COMA	Coma
;	PCOMA	Punto y coma
(PARINI	Abrir paréntesis
)	PARFIN	Cerrar paréntesis
¿	INTERRINI	Abrir interrogación
?	INTERRFIN	Cerrar interrogación
:	DOSPUN	Dos puntos
:	DOSPUNAP	Dos puntos y aparte
"	COMIDOB	Comillas dobles
'	COMISIM	Comillas simples
-	GUION	Guión
	GUIONLARGO	Guión bajo (subrayado)
¡	EXCLAINI	Abrir exclamación
!	EXCLAFIN	Cerrar exclamación
[CORCHINI	Abrir corchete
]	CORCHFIN	Cerrar corchete

Tabla 16: Inventario etiquetas puntuación

4 Conclusiones

En este artículo presentamos un etiquetario morfosintáctico completo y normalizado (de acuerdo con las recomendaciones de EAGLES) para etiquetar corpus lingüísticos de lengua gallega. Con la presentación del etiquetario SLI para el gallego, y de las correspondencias entre el etiquetario SLI del gallego y el etiquetario intermedio propuesto por EAGLES, pretendemos contribuir al avance de la investigación y desarrollo en las áreas de la lingüística de corpus y de las tecnologías lingüísticas de la lengua gallega.

Bibliografía

Aguirre Moreno, J.L., N. Andión y X. Gómez Guinovart. 2001. Aspectos ortográficos, léxicos y morfosintácticos del etiquetado lingüístico de un corpus de informática en lengua gallega. *Procesamiento del Lenguaje Natural*, 27: 13-19.

Álvarez, R., X.L. Regueira y H. Monteagudo. 1986. *Gramática galega*. Galaxia, Vigo.

Álvarez Lugrís, A. 2001. *Estilística comparada da tradución: Proposta metodolóxica e aplicación práctica ó estudio do corpus TECTRA de traducións do inglés ó galego*. Universidade de Vigo, Vigo.

Leech, G. y A. Wilson. 1996. *Recommendations for the Morphosyntactic Annotation of Corpora*. EAGLES Guidelines. Disponible en: <http://www.ilc.pi.cnr.it/EAGLES96/annotate/annotate.html>.

Monachini, M. y N. Calzolari, coord. 1996. *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora*. EAGLES Guidelines. Disponible en: <http://www.ilc.pi.cnr.it/EAGLES96/morphsyn/morphsyn.html>.

Monachini, M. y N. Calzolari. 1999. Standardization in the Lexicon. En H. van Halteren (ed), *Syntactic Wordclass Tagging*. Kluwer, Dordrecht.

Teufel, S. 1996. *ELM-DE: EAGLES Specifications for German Morphosyntax*. EAGLES Guidelines. Disponible en: http://www.ilc.pi.cnr.it/EAGLES96/elm_de/elm_de.html.

A Anexo 1: TECTRA (sección inglés)

A continuación se incluye una lista con los títulos de los textos originales ingleses que forman parte del corpus TECTRA, con el nombre de su autor/a, año de publicación y tamaño en número de palabras.

Original inglés, autor/a (año pub.)	Tamaño
<i>The Pearl</i> , John Steinbeck (1945)	26.476
<i>Animal Farm</i> , George Orwell (1945)	30.533
<i>To the Lighthouse</i> , Virginia Wolf (1927)	70.836
<i>The Call of the Wild</i> , Jack London (1903)	31.960

<i>Extracts from Adam's Diary</i> , Mark Twain (1893)	4.596
<i>Eve's Diary</i> , Mark Twain (1906)	7.036
<i>Spanish Galicia</i> , Aubrey F.G. Bell (1922)	40.543
<i>The Golem</i> , Isaac B. Singer (1982)	12.834
<i>Nine Stories</i> , J.D. Salinger (1948)	55.917
<i>The Catcher in the Rye</i> , J.D. Salinger (1945)	75.329
<i>A Portrait of the Artist as a Young Man</i> , James Joyce (1916)	83.641
<i>Lord of the Flies</i> , William Golding (1954)	62.052
<i>The Third Man</i> , Graham Greene (1950)	31.793
<i>A Sentimental Journey</i> , Laurence Sterne (1768)	42.620
Total TECTRA - sección inglés	575.166

Tabla 16: TECTRA (sección inglés)

B Anexo 2: TECTRA (sección gallego)

En este segundo anexo recogemos la lista de los títulos de las traducciones al gallego que forman parte del corpus TECTRA, con el nombre de su traductor/a, año de publicación y tamaño en número de palabras.

Traducción gallego, traductor/a (año)	Tamaño
<i>A perla</i> , Benigno F. Salgado (1990)	24.907
<i>A revolta dos animais</i> , X. Antón L. Dobao (1992)	26.215
<i>Cara ó faro</i> , Manuela Palacios & Xavier Castro (1993)	69.015
<i>A chamada da selva</i> , Gonzalo Navaza (1982)	29.053
<i>Retallos do diario de Adán</i> , Benigno F. Salgado (1992)	4.513
<i>Diario de Eva</i> , B. F. Salgado (1991)	6.771
<i>Galicia vista por un inglés</i> , X. M. Gómez Clemente (1994)	45.554
<i>O Golem</i> , Anxo Romero Louro (1989)	12.027
<i>Nove contos</i> , X. Antón L. Dobao (1994)	55.579
<i>O vixía no centeo</i> , X. Ramón F. Rodríguez (1990)	74.757
<i>Retrato do artista cando novo</i> , Vicente Araguas (1994)	82.398
<i>O señor das moscas</i> , X. M. Gómez Clemente (1993)	62.732
<i>O terceiro home</i> , M ^a Dolores M. Torres (1994)	30.833
<i>Unha viaxe sentimental</i> , Manuel Outeiriño (1992)	42.524
Total TECTRA - sección gallego	551.878

Tabla 18: TECTRA (sección gallego)

C Anexo 3: Referencias bibliográficas del corpus TECTRA

Por último, incluimos las referencias bibliográficas correspondientes a las novelas (originales en inglés y traducciones al gallego) incluidas en el corpus paralelo TECTRA:

Bell, A.F.G. 1922. *Spanish Galicia*. John Lane The Bodley Head, Londres.

Bell, A.F.G. 1994. *Galicia vista por un inglés*. Galaxia, Vigo.

Golding, W. 1954 (1962). *Lord of the Flies*. Faber & Faber, Londres.

Golding, W. 1993. *O Señor das moscas*. Sotelo Blanco, Santiago de Compostela.

Greene, G. 1950 (1974, 3ª ed.). *The Third Man*. Heinemann, Londres.

Greene, G. 1994. *O terceiro home*. Galaxia, Vigo.

Joyce, J. 1916 (1986, 11ª ed.). *A Portrait of the Artist as a Young Man*. Grafton Books, Londres.

Joyce, J. 1994. *Retrato do artista cando novo*. Laiovento, Santiago de Compostela.

London, J. 1903 (1975, 16ª ed.). *The Call of the Wild*. Heinemann, Londres.

London, J. 1982 (1983, 2ª ed.). *A chamada da selva*. Xerais, Vigo.

Orwell, G. 1945 (1987, 58ª ed.). *Animal Farm*. Penguin, Londres.

Orwell, G. 1992. *A revolta dos animais*. Positivas, Santiago de Compostela.

Salinger, J.D. 1948 (1986). *For Esmé, with Love and Squalor*. Penguin, Nueva York.

Salinger, J.D. 1994. *Nove contos*. Sotelo Blanco, Santiago de Compostela.

Salinger, J.D. 1951. *The Catcher in the Rye*. Penguin, Londres.

Salinger, J.D. 1990 (1992, 4ª ed.). *O vixía no centeo*. Xerais, Vigo.

Singer, I. B. 1982. *The Golem*. Penguin, Londres.

Singer, I. B. 1989. *O Golem*. Xerais, Vigo.

Steinbeck, J. 1945 (1986, 21ª ed.). *The Pearl*. Penguin, Nueva York.

Steinbeck, J. 1990 (1993, 5ª ed.). *A perla*. Galaxia, Vigo.

Sterne, L. 1768 (1995). *A Sentimental Journey Through France and Italy*. Wordsworth, Hertfordshire.

Sterne, L. 1992. *Unha viaxe sentimental por Francia e Italia*. Sotelo Blanco, Santiago de Compostela.

Twain, M. 1893 (1993). *Extracts from Adam's Diary*. Courage Books, Filadelfia.

Twain, M. 1992. *Retallos do diario de Adán*. Positivas, Santiago de Compostela.

Twain, M. 1906 (1993). *Eve's Diary*. Courage Books, Filadelfia.

Twain, M. 1991. *Diario de Eva*. Positivas, Santiago de Compostela.

Wolf, V. 1927 (1977, 18ª ed.). *To the Lighthouse*. Grafton Books, Londres.

Wolf, V. 1993. *Cara ó faro*. Sotelo Blanco, Santiago de Compostela.