

APLICACIÓN DO ETIQUETARIO MORFOSINTÁCTICO DO SLI Ó CORPUS DE TRADUCCIONS TECTRA

José Luis Aguirre Moreno
Alberto Álvarez Lugrís
Xavier Gómez Guinovart
Seminario de Lingüística Informática
Universidade de Vigo

Resumo

Neste artigo preséntase un etiquetario morfosintáctico completo e normalizado para etiquetar corpus lingüísticos de lingua galega. A elaboración deste etiquetario, deseñado polo Seminario de Lingüística Informática (*sli*) da Universidade de Vigo seguindo estrictamente as recomendacións de eagles (Leech e Wilson 1996), adaptadas por vez primeira ó galego, inclúe a creación dun etiquetario intermedio que nos permite establecer unha correspondencia entre a información gramatical para o galego codificada no *cluvi* (Corpus Lingüístico da Universidade de Vigo) e a que se atopa codificada no formato estándar de *eagles* en corpus doutras linguas.

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.

¹ Este traballo foi financiado pola Xunta de Galicia, dentro dos proxectos "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) e "Estudio e adquisición de recursos básicos de lingüística computacional do galego para a elaboración e mellora de aplicacións informáticas de tecnoloxía lingüística" (ref. PGIDT01TICC06E); e mais polo Ministerio de Ciencia y Tecnología (*mcyt*) e o Fondo Europeo de Desenvolvemento Rexional (*feder*), dentro do proxecto "Procesamiento lingüístico-computacional del Corpus Lingüístico de la Universidad de Vigo (*cluvi*)" (ref. BFF2002-01385). Máis información en <http://webs.uvigo.es/sli>.

1. Introducción¹

O *cluvi* (“Corpus Lingüístico da Universidade de Vigo”) é un corpus textual de rexistros especializados de lingua galega contemporánea oral e escrita. No seu estado actual de desenvolvemento, os textos da sección escrita do *cluvi* pertencen a catro rexistros especializados (dos ámbitos xurídico-administrativo, periodístico, informático e literario) e a tres “combinacións” lingüísticas co galego (monolingüe galego, traducción galego-castelán e traducción inglés-galego), e posúen unha extensión total aproximada de 4 millóns de palabras. Os textos do *cluvi* repártense en catro subcorpus de ó redor de 1 millón de palabras cada un: o corpus paralelo TECTRA (*cluvi-1*) de textos literarios inglés-galego, o corpus paralelo LEGA (*cluvi-2*) de textos xurídico-administrativos galego-castelán, o corpus monolingüe XIGA (*cluvi-3*) de textos sobre informática en galego e o corpus monolingüe MEGA (*cluvi-4*) de linguaxe dos medios de comunicación. Os obxectivos de procesamento do *cluvi* inclúen a súa etiquetaxe morfosintáctica completa e o aliñamento das equivalencias oracionais dos dous corpus paralelos, a extracción de información léxica, terminolóxica e fraseolóxica dos corpus etiquetados e a transferencia de resultados para aplicacións de traducción automática, extracción de información e resumo de documentos, recuperación de información en Internet e correctores gramaticais para procesamento de textos. Está en fase de elaboración a ampliación do *cluvi* con textos paralelos portugués-galego.

Nesta exposición, presentáremo-lo etiquetario morfosintáctico empregado no *sli* (Seminario de Lingüística Informática da Universidade de Vigo) para a anotación lingüística do *cluvi*, amosando en concreto as solucións adoptadas para a explotación do corpus paralelo TECTRA. Na exposición inclúese unha descrición do etiquetario morfosintáctico para a lingua galega elaborado polo *sli* consonte as directrices estándares europeas elaboradas por EAGLES (*Expert Advisory Group on Language Engineering Standards*) (Leech e Wilson, 1996; Monachini e Calzolari, 1996, 1999), adaptadas por vez primeira ó galego. Preséntanse así mesmo as correspondencias entre o etiquetario *sli* do galego e o etiquetario intermedio proposto por EAGLES (Leech e Wilson, 1996) como representación lingüísticamente neutral do conxunto de pares atributo-valor que describen a información lingüística codificada nas etiquetas.

2. O corpus TECTRA

O corpus TECTRA (Álvarez Lugrís, 2001) contén catorce novelas en lingua inglesa coas súas correspondentes traducións ó galego, cun total de 1.127.044 palabras, das cales 551.878 corresponden ás catorce traducións galegas e 575.166 ós catorce orixinais ingleses. Nos apéndices deste traballo

pódese ve-la relación completa de orixinais e traducións recompiladas neste corpus. Nunha fase inicial, os textos de TECTRA adquiríronse mediante dixitalización das obras impresas, posteriormente revisáronse (para corrixi-los inevitables erros do recoñecemento óptico de caracteres) e finalmente etiquetáronse estruturalmente en *xml*, con información sobre a obra, autor ou autora, traductor ou traductora, capítulo, sección, páxina, parágrafo e oración.

O procesamento lingüístico computacional de TECTRA, dentro do proxecto *cluvi* do *sli*, presenta tres vertentes diferenciadas: etiquetaxe, aliñamento e explotación. Na etiquetaxe dos textos da sección galega do TECTRA empregámo-lo estándar *xml* e o etiquetario morfosintáctico elaborado polo *sli* consonte as directrices de EAGLES e descrito no seguinte apartado do traballo. O sistema probabilístico de etiquetaxe e desambiguación empregado no *cluvi*, desenvolvido conxuntamente polo *sli* e Imaxin Software, utiliza un léxico computacional do galego que contén as especificacións morfosintácticas definidas no etiquetario do *sli*. Os textos en inglés son etiquetados cun etiquetario morfosintáctico da lingua inglesa compatible co de EAGLES e ferramentas doutros grupos de investigación, como o programa informático de etiquetaxe e desambiguación Trigram's Tags (TnT), desenvolvido no Departamento de Lingüística Computacional da Universidade de Saarland (en Saarbrücken, Alemaña) para o que o *sli* dispón dunha licencia de investigación.

Co fin de permitir desenvolvementos e aplicacións baseadas nas correspondencias gramaticais entre os textos galegos e ingleses etiquetados, estableceuse a correspondencia entre o etiquetario *sli* para o galego e o etiquetario EAGLES para o inglés mediante o etiquetario intermedio estándar proposto tamén por EAGLES. Tanto o aliñamento como a explotación do TECTRA para a extracción de información léxica beneficianse da anotación morfosintáctica e da súa correspondencia bilingüe.

O aliñamento dos bitextos almacénase en formato *tmx*, por ser o estándar para a codificación en *xml* de memorias de tradución e de corpus paralelos con independencia da aplicación utilizada (Melby, 2000). A consulta pública dos corpus paralelos do CLUVI², a través da web do *sli*, permite consultar e explotar as equivalencias bilingües do TECTRA con finalidades académicas de investigación e docencia, e tamén como ferramenta para a tradución³. Cómpre salientar o feito de que o galego non conta aínda cun dicionario bilingüe inglés-galego-inglés apto para o labor de tradución profesional, o que fai do *cluvi* unha ferramenta de consulta imprescindible neste contexto. A continuación, centrarémola exposición deste traballo na descrición do etiquetario *sli* para o galego e do etiquetario intermedio.

²Os corpus paralelos do CLUVI están dispoñibles en <http://sli.uvigo.es/CLUVI/>.

³Para os diversos problemas relativos á divulgación de corpus a través da web, véxase Santos (1999).

3. Etiquetario morfosintáctico do *sli*

3.1. Deseño do etiquetario

Para a creación do etiquetario do *sli*, no que respecta á identificación dos fenómenos gramaticais relevantes en lingua galega, baseámonos principalmente na descrición gramatical de Álvarez, Regueira e Monteagudo (1986) e Álvarez e Xove (2002). Ó mesmo tempo, adoptámo-las propostas de EAGLES tocante ás categorías gramaticais e ós trazos morfosintácticos que convén distinguir. Para isto, non nos limitamos a segui-las directrices xerais de EAGLES, senón que aplicamos estrictamente o esquema de atributos e valores recomendado por Leech e Wilson (1996), axeitándoo para o galego de modo análogo ó que se ten feito para outras linguas, como o italiano e o alemán (Teufel 1996).

Partindo da división en categorías principais, obrigatorias segundo Leech e Wilson (1996), determinámo-los trazos morfosintácticos aplicables ó galego, prescindindo dos que non o son e engadindo atributos ou valores cando foi preciso. Deste xeito, non incluímos no etiquetario o valor “neuro” do atributo “xénero” para os nomes comúns, xa que non é aplicable ó galego; mentres que debemos engadirlle o valor “antepretérito” ó atributo “tempo” dos verbos en modo indicativo para recolle-la forma verbal non composta do galego que expresa o que se adoita denominar pluscuamperfecto: “eu cantara”.

Outro aspecto fundamental do deseño do etiquetario do galego é o establecemento das correspondencias co etiquetario intermedio de EAGLES. O etiquetario intermedio é unha representación lingüisticamente neutral que describe os trazos lingüísticos (descritos en forma de pares atributo-valor) recollidos nun etiquetario, de xeito que se poidan facer corresponder doadamente coas marcas doutro conxunto de etiquetas (Leech e Wilson, 1996). O etiquetario intermedio permítenos traballar con etiquetas definidas segundo a terminoloxía gramatical propia da lingua galega e convertelas automaticamente ós trazos definidos no estándar de EAGLES. Aplicado ó corpus paralelo TECTRA, o etiquetario intermedio permítenos establecer de xeito automático a inequívoca correspondencia entre a información gramatical dos textos en galego e a dos textos en inglés, o que posibilita a explotación destas correspondencias en procesos lingüístico-computacionais posteriores á etiquetaxe, como o aliñamento dos bitextos ou a extracción automática de información léxica bilingüe contextual e fraseolóxica. De forma máis xeral, a correspondencia do etiquetario galego co etiquetario intermedio permítenos reutiliza-los textos etiquetados en aplicacións adaptadas ó estándar EAGLES.

3.2. Presentación do etiquetario

Para cada categoría amósanse tres táboas:

1. Táboa de relacións xerárquicas entre subcategorías, na que tamén se especifican as restricións na aplicación de atributos e valores. Esta descrición está baseada nas recomendacións de EAGLES recollidas por Monachini e Calzolari (1999), aínda que tratamos de amosa-la xerarquía entre as subcategorías e as restricións de aparición duns valores con outros de forma unificada, nunha soa táboa para toda a categoría, no canto de empregarmos unha táboa por cada atributo. En cada columna representamos un atributo. En cada ringleira representámolas combinacións dos distintos valores admitidos. A xerarquía entre subcategorías amósase mediante a agrupación gráfica dos valores na táboa. Cando un valor non é aplicable, déixase a cela correspondente en branco. Os atributos ou valores específicos da lingua galega, engadidos ó etiquetario, aparecen destacados en negra.
2. Relación de atributos e valores seguindo a numeración de Leech e Wilson (1996), tal e como queda para a súa aplicación ó galego. Inclúense os trazos que son opcionais no estándar se os adoptamos para a nosa lingua. Amósanse entre parénteses os atributos e valores que non se consideran aplicables para o galego e en negra os específicos desta lingua, engadidos ó etiquetario.
3. Inventario completo de etiquetas para cada categoría. Inclúese unha palabra como exemplo, a etiqueta empregada na nosa codificación, a descrición gramatical correspondente á etiqueta e, por último, a codificación da etiqueta intermedia correspondente segundo a numeración de atributos e valores expresada na relación anterior. Unha vez máis, os díxitos dos valores engadidos para o galego amósanse en negra.

A continuación amosámo-la relación completa de etiquetas do etiquetario *sli* para corpus de lingua galega, agrupadas por categorías seguindo a orde e o formato da descrición estándar das recomendacións de EAGLES.

3.2.1. Categoría nome

Cat = nome				
Atributos				
	Tipo	Xénero	Número	Forma
Valores	común	masc, fem, (neu)	sg, pl	plena, dimin.
	propio	masc, fem	sg	plena, dimin.
		masc, fem	pl	plena

Táboa 1: Xerarquía de subcategorías nominais

- (i) Tipo: 1. Común. 2. Propio.
(ii) Xénero: 1. Masculino. 2. Feminino. (3. Neutro.)
(iii) Número: 1. Singular. 2. Plural.
(iv) (Caso: 1. Nominativo...)
(v) **Forma: 1. Plena. 2. Diminutivo.**

Figura 1: Atributos e valores nominais

Exemplo	Etiqueta	Descrición	Etiqueta intermedia
<i>can</i>	NCMS	nome común masculino singular	N11101
<i>folla</i>	NCFS	nome común feminino singular	N12101
<i>homes</i>	NCMP	nome común masculino plural	N11201
<i>mulleres</i>	NCFP	nome común feminino plural	N12201
<i>canciño</i>	NCDMS	nome común diminutivo masculino singular	N11102
<i>folliña</i>	NCDFS	nome común diminutivo feminino singular	N12102
<i>homiños</i>	NCDMP	nome común diminutivo masculino plural	N11202
<i>mulleriñas</i>	NCDFP	nome común diminutivo feminino plural	N12202
<i>Aldán</i>	NPMS	nome propio masculino singular	N21101
<i>Antía</i>	NPFS	nome propio feminino singular	N22101
<i>Ancares</i>	NPMP	nome propio masculino plural	N21201
<i>Burgas</i>	NPPF	nome propio feminino plural	N22201
<i>Pedriño</i>	NPDMS	nome propio diminutivo masculino singular	N21102
<i>Carmiña</i>	NPDFS	nome propio diminutivo feminino singular	N22102

Táboa 2: Inventario de etiquetas nominais

3.2.2. Categoría verbo

Cat = verbo							
Atributos							
	Fin	Modo/ forma	Tempo	Pers	Nm	Xn	
Valores	finito	indicativo	pres, imperf, fut, pret, antepret	1,2, 2c,3	sg, pl		
		subx	pres, imperf, fut	1,2, 2c,3	sg, pl		
		imper		2, 2c	sg, pl		
		cond		1,2, 2c,3	sg, pl		
	non finito	infinitivo					
		inf conx			1,2, 2c,3	sg, pl	
		participio				sg, pl	m, f
		xerundio					
		xer conx (supino)			1,2	pl	

Táboa 3: Xerarquía de subcategorías verbais

- (i) Persoa: 1. Primeira, 2. Segunda, 3. Terceira. **4. Cortesía.**
(ii) Xénero: 1. Masculino, 2. Feminino. (3. Neutro.)
(iii) Número: 1. Singular, 2. Plural.
(iv) Finitude: 1. Finito, 2. Non finito.
(v) Forma verbal / Modo: 1. Indicativo, 2. Subxuntivo, 3. Imperativo, 4. Condicional, 5. Infinitivo, 6. Participio, 7. Xerundio, (8. Supino). **9. Infinitivo conxugado, A. Xerundio conxugado⁴.**
(vi) Tempo: 1. Presente, 2. Imperfecto, 3. Futuro, 4. Pretérito, **5. Antepretérito.**
(vii) Voz: 1. Activa, 2. Pasiva.
(viii) (Status: 1. Principal, 2. Auxiliar.)

Figura 2: Atributos e valores verbais

Para facilita-la consulta da Táboa 4, dividímo-lo inventario de etiquetas verbais en grupos, segundo o tempo e modo verbais:

Indicativo presente			
<i>collo</i>	VIPRS1	indicativo presente primeira singular	V10111110
<i>colles</i>	VIPRS2	indicativo presente segunda singular	V20111110
<i>colle</i>	VIPRS2C	indicativo presente segunda singular de cortesía	V40111110
<i>colle</i>	VIPRS3	indicativo presente terceira singular	V30111110
<i>collemos</i>	VIPRP1	indicativo presente primeira plural	V10211110
<i>colledes</i>	VIPRP2	indicativo presente segunda plural	V20211110
<i>collen</i>	VIPRP2C	indicativo presente segunda plural de cortesía	V40211110
<i>collen</i>	VIPRP3	indicativo presente terceira plural	V30211110

Indicativo copretérito			
<i>collia</i>	VICPS1	indicativo copretérito primeira singular	V10111210
<i>collias</i>	VICPS2	indicativo copretérito segunda singular	V20111210
<i>collia</i>	VICPS2C	indicativo copretérito segunda singular de cortesía	V40111210
<i>collia</i>	VICPS3	indicativo copretérito terceira singular	V30111210
<i>colliamos</i>	VICPP1	indicativo copretérito primeira plural	V10211210
<i>colliadés</i>	VICPP2	indicativo copretérito segunda plural	V20211210
<i>collian</i>	VICPP2C	indicativo copretérito segunda plural de cortesía	V40211210
<i>collian</i>	VICPP3	indicativo copretérito terceira plural	V30211210

Indicativo pretérito			
<i>collin</i>	VIPES1	indicativo pretérito primeira singular	V10111410
<i>colliches</i>	VIPES2	indicativo pretérito segunda singular	V20111410
<i>colleu</i>	VIPES2C	indicativo pretérito segunda singular de cortesía	V40111410
<i>colleu</i>	VIPES3	indicativo pretérito terceira singular	V30111410
<i>collemos</i>	VIPEP1	indicativo pretérito primeira plural	V10211410
<i>collestes</i>	VIPEP2	indicativo pretérito segunda plural	V20211410
<i>colleron</i>	VIPEP2C	indicativo pretérito segunda plural de cortesía	V40211410
<i>colleron</i>	VIPEP3	indicativo pretérito terceira plural	V30211410

Indicativo antepretérito			
<i>collera</i>	VIAPS1	indicativo antepretérito primeira singular	V10111510
<i>colleras</i>	VIAPS2	indicativo antepretérito segunda singular	V20111510
<i>collera</i>	VIAPS2C	indicativo antepretérito segunda singular de cortesía	V40111510
<i>collera</i>	VIAPS3	indicativo antepretérito terceira singular	V30111510
<i>colleramos</i>	VIAPP1	indicativo antepretérito primeira plural	V10211510
<i>collerades</i>	VIAPP2	indicativo antepretérito segunda plural	V20211510
<i>colleran</i>	VIAPP2C	indicativo antepretérito segunda plural de cortesía	V40211510
<i>colleran</i>	VIAPP3	indicativo antepretérito terceira plural	V30211510

⁴Forma verbal pouco frecuente en galego, mais viva nalgunhas zonas para a primeira e segunda persoa do plural (Álvarez e Xove 2002: 319).

Indicativo futuro			
<i>collerei</i>	VIFUS1	indicativo futuro primeira singular	V10111310
<i>collerás</i>	VIFUS2	indicativo futuro segunda singular	V20111310
<i>collerá</i>	VIFUS2C	indicativo futuro segunda singular de cortesia	V40111310
<i>collerá</i>	VIFUS3	indicativo futuro terceira singular	V30111310
<i>colleremos</i>	VIFUP1	indicativo futuro primeira plural	V10211310
<i>colleredes</i>	VIFUP2	indicativo futuro segunda plural	V20211310
<i>collerán</i>	VIFUP2C	indicativo futuro segunda plural de cortesia	V40211310
<i>collerán</i>	VIFUP3	indicativo futuro terceira plural	V30211310

Indicativo pospretérito			
<i>colleria</i>	VIPPS1	indicativo pospretérito primeira singular	V10114010
<i>collerías</i>	VIPPS2	indicativo pospretérito segunda singular	V20114010
<i>colleria</i>	VIPPS2C	indicativo pospretérito segunda singular de cortesia	V40114010
<i>colleria</i>	VIPPS3	indicativo pospretérito terceira singular	V30114010
<i>colleríamos</i>	VIPPP1	indicativo pospretérito primeira plural	V10214010
<i>colleríades</i>	VIPPP2	indicativo pospretérito segunda plural	V20214010
<i>collerían</i>	VIPPP2C	indicativo pospretérito segunda plural de cortesia	V40214010
<i>collerían</i>	VIPPP3	indicativo pospretérito terceira plural	V30214010

Subxuntivo presente			
<i>colla</i>	VSPRS1	subxuntivo presente primeira singular	V10112110
<i>collas</i>	VSPRS2	subxuntivo presente segunda singular	V20112110
<i>colla</i>	VSPRS2C	subxuntivo presente segunda singular de cortesia	V40112110
<i>colla</i>	VSPRS3	subxuntivo presente terceira singular	V30112110
<i>collamos</i>	VSPRP1	subxuntivo presente primeira plural	V10212110
<i>collades</i>	VSPRP2	subxuntivo presente segunda plural	V20212110
<i>collan</i>	VSPRP2C	subxuntivo presente segunda plural de cortesia	V40212110
<i>collan</i>	VSPRP3	subxuntivo presente terceira plural	V30212110

Subxuntivo pretérito			
<i>collese</i>	VSPE1	subxuntivo pretérito primeira singular	V10112210
<i>collees</i>	VSPE2	subxuntivo pretérito segunda singular	V20112210
<i>collese</i>	VSPE2C	subxuntivo pretérito segunda singular de cortesia	V40112210
<i>collese</i>	VSPE3	subxuntivo pretérito terceira singular	V30112210
<i>collesemos</i>	VSPEP1	subxuntivo pretérito primeira plural	V10212210
<i>collestedes</i>	VSPEP2	subxuntivo pretérito segunda plural	V20212210
<i>collesen</i>	VSPEP2C	subxuntivo pretérito segunda plural de cortesia	V40212210
<i>collesen</i>	VSPEP3	subxuntivo pretérito terceira plural	V30212210

Subxuntivo futuro			
<i>coller</i>	VSFUS1	subxuntivo futuro primeira singular	V10112310
<i>colleres</i>	VSFUS2	subxuntivo futuro segunda singular	V20112310
<i>coller</i>	VSFUS2C	subxuntivo futuro segunda singular de cortesia	V40112310
<i>coller</i>	VSFUS3	subxuntivo futuro terceira singular	V30112310
<i>collermos</i>	VSFUP1	subxuntivo futuro primeira plural	V10212310
<i>collerdes</i>	VSFUP2	subxuntivo futuro segunda plural	V20212310
<i>colleren</i>	VSFUP2C	subxuntivo futuro segunda plural de cortesia	V40212310
<i>colleren</i>	VSFUP3	subxuntivo futuro terceira plural	V30212310

Imperativo			
<i>colle</i>	VIMPS2	imperativo segunda singular	V20113010
<i>colla</i>	VIMPS2C	imperativo segunda singular de cortesia	V40113010
<i>collede</i>	VIMPP2	imperativo segunda plural	V20213010
<i>collan</i>	VIMPP2C	imperativo segunda plural de cortesia	V40213010

Formas non finitas			
<i>coller</i>	VINFCS1	infinitivo conxugado primeira singular	V10129010
<i>colleres</i>	VINFCS2	infinitivo conxugado segunda singular	V20129010
<i>coller</i>	VINFCS2C	infinitivo conxugado segunda singular de cortesía	V40129010
<i>coller</i>	VINFCS3	infinitivo conxugado terceira singular	V30129010
<i>collermos</i>	VINFCP1	infinitivo conxugado primeira plural	V10229010
<i>collerdes</i>	VINFCP2	infinitivo conxugado segunda plural	V20229010
<i>colleren</i>	VINFCP2C	infinitivo conxugado segunda plural de cortesía	V40229010
<i>colleren</i>	VINFCP3	infinitivo conxugado terceira plural	V30229010
<i>coller</i>	VINF	infinitivo	V00025010
<i>collendo</i>	VXER	xerundio	V00027010
<i>colléndomos</i>	VXERCP1	xerundio conxugado primeira plural	V1022A010
<i>colléndodes</i>	VXERCP2	xerundio conxugado segunda plural	V2022A010
<i>collido</i>	VPARMS	participio masculino singular	V01126010
<i>collida</i>	VPARFS	participio feminino singular	V02126010
<i>collidos</i>	VPARMP	participio masculino plural	V01226010
<i>collidas</i>	VPARFP	participio feminino plural	V02226010

Táboa 4: Inventario de etiquetas verbais

3.2.3. Categoría adxectivo

Cat = adxectivo				
Atributos				
	Grao	Xén	Núm	Forma
Valores	positivo	m, f	sg, pl	plena, diminutivo
	positivo	m, f	sg	apocopado
	superlativo, comparativo	m, f	sg, pl	plena

Táboa 5: Xerarquía de subcategorías do adxectivo

(i) Grao: 1. Positivo. 2. Comparativo. 3. Superlativo.

(ii) Xénero: 1. Masculino. 2. Feminino. (3. Neutro.)

(iii) Número: 1. Singular. 2. Plural.

(iv) (Caso: 1. Nominativo ...)

(v) **Forma: 1. Plena. 2. Diminutivo. 3. Apocopado.**

Figura 3: Atributos e valores do adxectivo

<i>novo</i>	AXMS	adxectivo masculino singular	AJ11101
<i>nova</i>	AXFS	adxectivo feminino singular	AJ12101
<i>novos</i>	AXMP	adxectivo masculino plural	AJ11201
<i>novas</i>	AXFP	adxectivo feminino plural	AJ12201
<i>noviño</i>	AXDMS	adxectivo diminutivo masculino singular	AJ11102
<i>noviña</i>	AXDFS	adxectivo diminutivo feminino singular	AJ12102
<i>noviños</i>	AXDMP	adxectivo diminutivo masculino plural	AJ11202
<i>noviñas</i>	AXDFP	adxectivo diminutivo feminino plural	AJ12202
<i>novísimo</i>	AXSMS	adxectivo superlativo masculino singular	AJ31101
<i>novísima</i>	AXSFS	adxectivo superlativo feminino singular	AJ32101

<i>novísimos</i>	AXSMP	adxectivo superlativo masculino plural	AJ31201
<i>novísimas</i>	AXSFP	adxectivo superlativo feminino plural	AJ32201
<i>mellor</i>	AXCMS	adxectivo comparativo masculino singular	AJ21101
<i>mellor</i>	AXCFS	adxectivo comparativo feminino singular	AJ22101
<i>mellores</i>	AXCMP	adxectivo comparativo masculino plural	AJ21201
<i>mellores</i>	AXCFP	adxectivo comparativo feminino plural	AJ22201
<i>gran</i>	AXAPMS	adxectivo apocopado masculino singular	AJ11103
<i>gran</i>	AXAPFS	adxectivo apocopado feminino singular	AJ12103

Táboa 6: Inventario de etiquetas do adxectivo⁵

3.2.4. Categoría pronome

A Táboa 7 que se inclúe a continuación, coa xerarquía de atributos e valores pronominais, non inclúe a información relativa ós pronomes persoais que, pola súa complexidade, se amosa por separado na Táboa 8. Para facilita-la consulta da Táboa 8 que se amosa máis adiante, coas restricións de aplicación entre atributos e valores dos pronomes persoais, inclúese unha columna adicional coas formas que se describen.

Cat = pronome									
Atributos									
	Tipo-pron	Caso	Persoa	Xénero	Número	Poseutivo ⁶	Cortesía	Grao	Forma
Valores	dem		1,2,3	m, f, n	sg, pl				
	indef			m, f	sg, pl			positivo	plena, diminutivo
								superlativo	plena
	poses		1,2,3	m, f	sg, pl	sg, pl	cortés		
	interr			m,f	sg, pl				
rel ⁷			m, f	sg, pl					

Táboa 7: Xerarquía de subcategorías pronominais

- (i) Persoa: 1. Primeira. 2. Segunda. 3. Terceira.
(ii) Xénero: 1. Masculino. 2. Feminino. 3. Neutro.
(iii) Número: 1. Singular. 2. Plural.
(iv) Poseutivo: 1. Singular. 2. Plural.
(v) Caso: 1. Nominativo. (2. Xenitivo.) 3. Dativo. 4. Acusativo. (5. Non xenitivo.) 6. Oblicuo.
(vi) Categoría: 1. Pronome. (2. Determinante⁸.) (3. Ámbolos dous.)
(vii) Tipo-pron.: 1. Demostrativo 2. Indefinido 3. Poseutivo. 4. Interr./Relat. 5. Persoal / reflexivo.
(viii) (Tipo-det: 1. Demostrativo ...)
(ix) Tipo especial de pron: 1. Persoal. 2. Reflexivo. (3. Recíproco.) 4. Exclusivo.
(x) Tipo qu-: 1. Interrogativo. 2. Relativo. 3. Exclamativo.
(xi) Cortesía: 1. Cortés 2. Familiar.
(xii) Tonicidade: 1. Tónico. 2. Átono.
(xiii) Grao: 1. Positivo. 2. Superlativo.
(xiv) Forma: 1. Plena. 2. Diminutivo.

Figura 4: Atributos e valores pronominais

Cat = pronome, tipo-pron = persoal / reflexivo								
Atributos								
Valores	Tonicidade	Caso	Tipo-pron especial	Pers.	Xén.	Núm.	Cortesía	Formas
		Tónico	Nom	pers	1		sg	
				2		sg	fam.	<i>ti</i>
				1				<i>nós</i>
				2		pl		<i>vós</i>
				2		sg	cor-tés	<i>vostede</i>
						pl		<i>vostedes</i>
				3	m	sg		<i>el</i>
					f	pl		<i>ela</i>
				2	m	pl		<i>vosoutros</i>
					f			<i>vosoutras</i>
				1	m	sg		<i>min</i>
				3	n			<i>elo⁹</i>
				3		sg		<i>si</i>
						pl		
				1		sg		<i>me</i>
						pl		<i>nos</i>
				2		pl		<i>vos</i>
				2		sg	fam	<i>che</i>
				2		sg	cor-tés	<i>lle</i>
						pl		<i>lles</i>
				3		sg		<i>lle</i>
						pl		<i>lles</i>
				2		sg	fam	<i>te</i>
				2	m	sg	cor-tés	<i>o, a</i>
					f	pl		<i>os, as</i>
				3	m	sg		<i>o, a</i>
					f	pl		<i>os, as</i>
				2		sg	cor-tés	<i>se</i>
						pl		
				3		sg		<i>se</i>
						pl		

Táboa 8: Xerarquía de subcategorías do pronome persoal

⁵No grao comparativo e nas formas apocopadas dos adxectivos, optamos por distingui-los valores de xénero aínda que a distinción non se reflicta morfoloxicamente.

⁶Atributo que se refire á persoa posuidora: "miña" (sg)/ "nosa" (pl).

⁷Optamos por distinguir no etiquetario entre interrogativo e relativo malia as coincidencias morfolóxicas.

⁸Optamos por non empregar-la distinción entre pronome e determinante, seguindo as descrições gramaticais do galego nas que nos baseamos para a elaboración do etiquetario.

⁹Forma antiga e en desuso, pero que recolleemos no etiquetario por aparecer en galego contemporáneo nalgúñas frases feitas ("estar a elo") (Álvarez, Regueira e Monteagudo 1986:160-161; *rag/ilg* 1995: 91).

¹⁰No etiquetario só incluímo-lo valor "reflexivo" para as formas pronominais "si" e "se".

Para simplificar a consulta da Tábua 9, dividímo-lo inventario de etiquetas pronominais en grupos, segundo o tipo de pronome e as súas características:

Pronomes demostrativos			
<i>este</i>	PDMS1	pronome demostrativo masculino singular primeira	PD11100110000000
<i>ese</i>	PDMS2	pronome demostrativo masculino singular segunda	PD21100110000000
<i>aquel</i>	PDMS3	pronome demostrativo masculino singular terceira	PD31100110000000
<i>estes</i>	PDMP1	pronome demostrativo masculino plural primeira	PD11200110000000
<i>eses</i>	PDMP2	pronome demostrativo masculino plural segunda	PD21200110000000
<i>aqueles</i>	PDMP3	pronome demostrativo masculino plural terceira	PD31200110000000
<i>esta</i>	PDFS1	pronome demostrativo feminino singular primeira	PD12100110000000
<i>esa</i>	PDFS2	pronome demostrativo feminino singular segunda	PD22100110000000
<i>aquela</i>	PDFS3	pronome demostrativo feminino singular terceira	PD32100110000000
<i>estas</i>	PDFP1	pronome demostrativo feminino plural primeira	PD12200110000000
<i>esas</i>	PDFP2	pronome demostrativo feminino plural segunda	PD22200110000000
<i>aquelas</i>	PDFP3	pronome demostrativo feminino plural terceira	PD32200110000000
<i>isto</i>	PDNS1	pronome demostrativo neutro singular primeira	PD13100110000000
<i>iso</i>	PDNS2	pronome demostrativo neutro singular segunda	PD23100110000000
<i>aquilo</i>	PDNS3	pronome demostrativo neutro singular terceira	PD33100110000000

Pronomes posesivos			
<i>meu</i>	PVS1MS	pronome posesivo. Posuidor(a): singular primeira. Posuído(a): masculino singular	PD11110130000000
<i>miña</i>	PVS1FS	pronome posesivo. Posuidor(a): singular primeira. Posuído(a): feminino singular	PD12110130000000
<i>meus</i>	PVS1MP	pronome posesivo. Posuidor(a): singular primeira. Posuído(a): masculino plural	PD11210130000000
<i>miñas</i>	PVS1FP	pronome posesivo. Posuidor(a): singular primeira. Posuído(a): feminino plural	PD12210130000000
<i>teu</i>	PVS2MS	pronome posesivo. Posuidor(a): singular segunda. Posuído(a): masculino singular	PD21110130000000
<i>tiúa</i>	PVS2FS	pronome posesivo. Posuidor(a): singular segunda. Posuído(a): feminino singular	PD22110130000000
<i>teus</i>	PVS2MP	pronome posesivo. Posuidor(a): singular segunda. Posuído(a): masculino plural	PD21210130000000
<i>tiúas</i>	PVS2FP	pronome posesivo. Posuidor(a): singular segunda. Posuído(a): feminino plural	PD22210130000000
<i>seu</i>	PVS3MS	pronome posesivo. Posuidor(a): singular terceira. Posuído(a): masculino singular	PD31110130000000
<i>seu</i>	PVP3MS	pronome posesivo. Posuidor(a): plural terceira. Posuído(a): masculino plural	PD31120130000000
<i>seu</i>	PVS2CMS	pronome posesivo. Posuidor(a): singular segunda de cortesia. Posuído(a): masculino singular	PD21110130001000
<i>seu</i>	PVP2CMS	pronome posesivo. Posuidor(a): plural segunda de cortesia. Posuído(a): masculino singular	PD11120130001000
<i>súa</i>	PVS3FS	pronome posesivo. Posuidor(a): singular terceira. Posuído(a): feminino singular	PD32110130000000
<i>súa</i>	PVP3FS	pronome posesivo. Posuidor(a): plural terceira. Posuído(a): feminino singular	PD32120130000000
<i>súa</i>	PVS2CFS	pronome posesivo. Posuidor(a): singular segunda de cortesia. Posuído(a): feminino singular	PD22110130001000
<i>súa</i>	PVP2CFS	pronome posesivo. Posuidor(a): plural segunda de cortesia. Posuído(a): feminino singular	PD22120130001000
<i>seus</i>	PVS3MP	pronome posesivo. Posuidor(a): singular terceira. Posuído(a): masculino plural	PD31210130000000
<i>seus</i>	PVP3MP	pronome posesivo. Posuidor(a): plural terceira. Posuído(a): masculino plural	PD31220130000000
<i>seus</i>	PVS2CMP	pronome posesivo. Posuidor(a): singular segunda de cortesia. Posuído(a): masculino plural	PD21210130001000
<i>seus</i>	PVP2CFP	pronome posesivo. Posuidor(a): plural segunda de cortesia. Posuído(a): masculino plural	PD21220130001000

<i>súas</i>	PVS3FP	pronome posesivo. Posuidor(a): singular terceira. Posuído(a): feminino plural	PD32210130000000
<i>súas</i>	PVP3FP	pronome posesivo. Posuidor(a): plural terceira. Posuído(a): feminino plural	PD32220130000000
<i>súas</i>	PVS2CFP	pronome posesivo. Posuidor(a): singular segunda de cortesia. Posuído(a): feminino plural	PD22210130001000
<i>súas</i>	PVP2CFP	pronome posesivo. Posuidor(a): plural segunda de cortesia. Posuído(a): feminino plural	PD22220130001000
<i>noso</i>	PVP1MS	pronome posesivo. Posuidor(a): plural primeira. Posuído(a): masculino singular	PD11120130000000
<i>nosa</i>	PVP1FS	pronome posesivo. Posuidor(a): plural primeira. Posuído(a): feminino singular	PD12120130000000
<i>nosos</i>	PVP1MP	pronome posesivo. Posuidor(a): plural primeira. Posuído(a): masculino plural	PD11220130000000
<i>nosas</i>	PVP1FP	pronome posesivo. Posuidor(a): plural primeira. Posuído(a): feminino plural	PD12220130000000
<i>voso</i>	PVP2MS	pronome posesivo. Posuidor(a): plural segunda. Posuído(a): masculino singular	PD21120130000000
<i>vosa</i>	PVP2FS	pronome posesivo. Posuidor(a): plural segunda. Posuído(a): feminino singular	PD22120130000000
<i>vosos</i>	PVP2MP	pronome posesivo. Posuidor(a): plural segunda. Posuído(a): masculino plural	PD21220130000000
<i>vosas</i>	PVP2FP	pronome posesivo. Posuidor(a): plural segunda. Posuído(a): feminino plural	PD22220130000000

Pronomes indefinidos

<i>algo</i>	PI	pronome indefinido (invariable)	PD00000322000011
<i>algún</i>	PIMS	pronome indefinido masculino singular	PD01100120000011
<i>algunha</i>	PIFS	pronome indefinido feminino singular	PD02100120000011
<i>algúns</i>	PIMP	pronome indefinido masculino plural	PD01200120000011
<i>algunhas</i>	PIFP	pronome indefinido feminino plural	PD02200120000011
<i>pouquiño</i>	PIDMS	pronome indefinido diminutivo masculino singular	PD01100120000012
<i>pouquiña</i>	PIDFS	pronome indefinido diminutivo feminino singular	PD02100120000012
<i>pouquiños</i>	PIDMP	pronome indefinido diminutivo masculino plural	PD01200120000012
<i>pouquiñas</i>	PIDFP	pronome indefinido diminutivo feminino plural	PD02200120000012
<i>pouquisimo</i>	PISMS	pronome indefinido superlativo masculino singular	PD01100120000021
<i>pouquisima</i>	PISFS	pronome indefinido superlativo feminino singular	PD02100120000021
<i>pouquisimos</i>	PISMP	pronome indefinido superlativo masculino plural	PD01200120000021
<i>pouquisimas</i>	PISFP	pronome indefinido superlativo feminino plural	PD02200120000021

Pronomes relativos

<i>quen</i>	PR	pronome relativo (invariable)	PD00000140020000
<i>cal (o)</i>	PRMS	pronome relativo masculino singular	PD01100140020000
<i>cal (a)</i>	PRFS	pronome relativo feminino singular	PD02100140020000
<i>cales (os)</i>	PRMP	pronome relativo masculino plural	PD01200140020000
<i>cales (as)</i>	PRFP	pronome relativo feminino plural	PD02200140020000

Pronomes interrogativos

<i>quen</i>	PIN	pronome interrogativo (invariable)	PD00000140010000
<i>cal (o)</i>	PINMS	pronome interrogativo masculino singular	PD01100140010000
<i>cal (a)</i>	PINFS	pronome interrogativo feminino singular	PD02100140010000
<i>cales (os)</i>	PINMP	pronome interrogativo masculino plural	PD01200140010000
<i>cales (as)</i>	PINFP	pronome interrogativo feminino plural	PD02200140010000

Pronomes pessoais tónicos: formas rectas			
<i>eu</i>	PPS1	pronome pessoal tónico recto singular primeira	PD10101150100100
<i>ti</i>	PPS2	pronome pessoal tónico recto singular segunda	PD20101150100100
<i>vostede</i>	PPS2C	pronome pessoal tónico recto singular segunda de cortesia	PD20101150101100
<i>el</i>	PPMS3	pronome pessoal tónico recto masculino singular terceira	PD31101150100100
<i>ela</i>	PPFS3	pronome pessoal tónico recto feminino singular terceira	PD32101150100100
<i>nós</i>	PPP1	pronome pessoal tónico recto plural primeira	PD10201150100100
<i>nosoutros</i>	PPMP1E	pronome pessoal tónico recto masculino plural primeira exclusivo	PD11201150400100
<i>nosoutras</i>	PPFP1E	pronome pessoal tónico recto feminino plural primeira exclusivo	PD12201150400100
<i>vós</i>	PPP2	pronome pessoal tónico recto plural segunda	PD20201150100100
<i>vosoutros</i>	PPMP2E	pronome pessoal tónico recto masculino plural segunda exclusivo	PD21201150400100
<i>vosoutras</i>	PPFP2E	pronome pessoal tónico recto feminino plural segunda exclusivo	PD22201150400100
<i>vostedes</i>	PPP2C	pronome pessoal tónico recto plural segunda de cortesia	PD20201150101100
<i>eles</i>	PPMP3	pronome pessoal tónico recto masculino plural terceira	PD31201150100100
<i>elas</i>	PPFP3	pronome pessoal tónico recto feminino plural terceira	PD32201150100100

Pronomes pessoais tónicos: formas oblicuas			
<i>min</i>	PPS1O	pronome pessoal tónico oblicuo singular primeira	PD10106150100100
<i>ti</i>	PPS2O	pronome pessoal tónico oblicuo singular segunda	PD20106150100100
<i>vostede</i>	PPS2CO	pronome pessoal tónico oblicuo singular segunda de cortesia	PD20106150101100
<i>el</i>	PPMS3O	pronome pessoal tónico oblicuo masculino singular terceira	PD31106150100100
<i>ela</i>	PPFS3O	pronome pessoal tónico oblicuo feminino singular terceira	PD32106150100100
<i>elo</i>	PPNS3O	pronome pessoal tónico oblicuo neutro singular terceira	PD33106150100100
<i>nós</i>	PPP1O	pronome pessoal tónico oblicuo plural primeira	PD10206150100100
<i>nosoutros</i>	PPMP1EO	pronome pessoal tónico oblicuo masculino plural primeira exclusivo	PD11206150400100
<i>nosoutras</i>	PPFP1EO	pronome pessoal tónico oblicuo feminino plural primeira exclusivo	PD12206150400100
<i>vós</i>	PPP2O	pronome pessoal tónico oblicuo plural segunda	PD20206150100100
<i>vosoutros</i>	PPMP2EO	pronome pessoal tónico oblicuo masculino plural segunda exclusivo	PD21206150400100
<i>vosoutras</i>	PPFP2EO	pronome pessoal tónico oblicuo feminino plural segunda exclusivo	PD22206150400100
<i>vostedes</i>	PPP2CO	pronome pessoal tónico oblicuo plural segunda de cortesia	PD20206150101100
<i>eles</i>	PPMP3O	pronome pessoal tónico oblicuo masculino plural terceira	PD31206150100100
<i>elas</i>	PPFP3O	pronome pessoal tónico oblicuo feminino plural terceira	PD32206150100100
<i>si</i>	PPOR	pronome pessoal tónico oblicuo reflexivo	PD30006150200100

Pronomes persoais átonos			
<i>me</i>	PPS1A	pronome persoal átono singular primeira	PD1010[3 4]150100200
<i>che</i>	PPS2AD	pronome persoal átono singular segunda dativo	PD20103150102200
<i>te</i>	PPS2AA	pronome persoal átono singular segunda acusativo	PD20104150102200
<i>lle</i>	PPS2CAD	pronome persoal átono singular segunda de cortesía dativo	PD20103150101200
<i>lle</i>	PPS3AD	pronome persoal átono singular terceira dativo	PD30103150100200
<i>o</i>	PPMS3AA	pronome persoal átono masculino singular terceira acusativo	PD31104150100200
<i>o</i>	PPMS2CAA	pronome persoal átono masculino singular segunda de cortesía acusativo	PD21104150101200
<i>a</i>	PPFS3AA	pronome persoal átono feminino singular terceira acusativo	PD32104150100200
<i>a</i>	PPFS2CAA	pronome persoal átono feminino singular segunda de cortesía acusativo	PD22104150101200
<i>se</i>	PPS3AR	pronome persoal átono singular terceira reflexivo	PD30100150200200
<i>se</i>	PPS2CAR	pronome persoal átono singular segunda de cortesía reflexivo	PD20000150201200
<i>nos</i>	PPP1A	pronome persoal átono plural primeira	PD1020[3 4]150100200
<i>vos</i>	PPP2A	pronome persoal átono plural segunda	PD2020[3 4]150100200
<i>lles</i>	PPP2CAD	pronome persoal átono plural segunda de cortesía dativo	PD20203150101200
<i>lles</i>	PPP3AD	pronome persoal átono plural terceira dativo	PD30203150100200
<i>os</i>	PPMP3AA	pronome persoal átono masculino plural terceira acusativo	PD31204150100200
<i>os</i>	PPMP2CAA	pronome persoal átono masculino plural segunda de cortesía acusativo	PD21204150101200
<i>as</i>	PPFP3AA	pronome persoal átono feminino plural terceira acusativo	PD32204150100200
<i>as</i>	PPFP2CAA	pronome persoal átono feminino plural segunda de cortesía acusativo	PD22204150101200
<i>se</i>	PPP3AR	pronome persoal átono plural terceira reflexivo	PD30200150200200
<i>se</i>	PPP2CAR	pronome persoal átono plural segunda de cortesía reflexivo	PD20200150201200

Táboa 9: Inventario de etiquetas pronominais

3.2.5. Categoría numeral

No etiquetario galego tratámo-los numerais como un tipo de pronome. Con todo, ó facérmola correspondencia co etiquetario intermedio considerámolo como unha categoría independente, seguindo as recomendacións de EAGLES.

Cat = numeral				
Atributos				
Valores	Función	Tipo	Xénero	Número
		Pronome	Cardinal	m, f
	Ordinal		m, f	sg, pl

Táboa 10: Xerarquía de subcategorías do numeral

(i) Tipo: 1. Cardinal. 2. Ordinal.
(ii) Xénero: 1. Masculino. 2. Feminino.
(iii) Número: 1. Singular. 2. Plural.
(iv) (Caso: 1. Nominativo...)
(v) Función: 1. Pronome. (2. Determinante.) (3. Adxectivo.)

Figura 5: Atributos e valores do numeral

<i>un</i>	PNCMS	pronome numeral cardinal masculino singular	NU11101
<i>unha</i>	PNCFS	pronome numeral cardinal feminino singular	NU12101
<i>dous</i>	PNCMP	pronome numeral cardinal masculino plural	NU11201
<i>dúas</i>	PNCFP	pronome numeral cardinal feminino plural	NU12201
<i>tres</i>	PNC	pronome numeral cardinal (invariable)	NU10001
<i>primeiro</i>	PNOMS	pronome numeral ordinal masculino singular	NU21101
<i>primeira</i>	PNOFS	pronome numeral ordinal feminino singular	NU22101
<i>primeiros</i>	PNOMP	pronome numeral ordinal masculino plural	NU21201
<i>primeiras</i>	PNOPF	pronome numeral ordinal feminino plural	NU22201

Táboa 11: Inventario de etiquetas do numeral

3.2.6. Categoría artigo

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

Táboa 12: Xerarquía de subcategorías do artigo

- (i) Tipo de artigo: 1. Definido. 2. Indefinido.
(ii) Xénero: 1. Masculino. 2. Feminino.
(iii) Número: 1. Singular. 2. Plural.
(iv) (Caso: 1. Nominativo...)

Figura 6: Atributos e valores do artigo

<i>o</i>	ARDMS	artigo determinado masculino singular	AT1110
<i>a</i>	ARDFS	artigo determinado feminino singular	AT1210
<i>os</i>	ARDMP	artigo determinado masculino plural	AT1120
<i>as</i>	ARDFP	artigo determinado feminino plural	AT1220
<i>un</i>	ARIMS	artigo indeterminado masculino singular	AT2110
<i>unha</i>	ARIFS	artigo indeterminado feminino singular	AT2210
<i>uns</i>	ARIMS	artigo indeterminado masculino plural	AT2120
<i>unhas</i>	ARIMP	artigo indeterminado feminino plural	AT2220

Táboa 13: Inventario de etiquetas do artigo

3.2.7. Categoría adverbio

Cat = adverbio		
Atributos		
Valores	Grao	Forma
	Positivo	plena, diminutivo
	Comparativo	plena
	Superlativo	plena

Táboa 14: Xerarquía de subcategorías do adverbio

- (i) Grao: 1. Positivo. 2. Comparativo. 3. Superlativo.
(ii) **Forma: 1. Plena. 2. Diminutivo.**

Figura 7: Atributos e valores do adverbio

<i>aquí</i>	ADV	adverbio	AV11
<i>cerquiña</i>	ADVD	adverbio diminutivo	AV12
<i>mellor</i>	ADVC	adverbio comparativo	AV21
<i>cerquisima</i>	ADVS	adverbio superlativo	AV31

Táboa 15: Inventario de etiquetas do adverbio

3.2.8. Outras categorías gramaticais

<i>con</i>	PREP	preposición	AP1
<i>e</i>	CONX	conxunción	C0
<i>ah</i>	INTERX	interxección	I
<i>non</i>	NEG	partícula negativa	U2

Táboa 16: Inventario de etiquetas doutras categorías

3.2.9. Categorías residuais

Incluimos a seguir, na Táboa 15, unha lista de etiquetas de fenómenos lingüísticos denominados por EAGLES “residuais”, que non se refiren propiamente a categorías morfosintácticas, pero que resultan de grande utilidade na anotación de corpus textuais de lingua real.

<i>pizza</i>	RES	palabra estranxeira	R100
<i>x = 2h.p</i>	RFO	fórmula	R200
@	RSI	símbolo	R300
<i>ONU</i>	RAC	acrónimo	R400
<i>D.</i>	RAB	abreviatura	R500
???	RSC	sen clasificar	R600

Táboa 17: Inventario de etiquetas das categorías residuais

3.2.10. Etiquetas para puntuación

Incluimos aquí tamén, a modo de apéndice deste capítulo, as etiquetas empregadas no *chuvi* para a puntuación, coa intención de proporcionar un etiquetario completo útil para a lingüística de corpus do galego.

.	PUNTO	punto e seguido
.	PUNTOAP	punto e aparte
,	COMA	coma
;	PCOMA	punto e coma
(PARINI	abrir parénteses
)	PARFIN	pechar parénteses
¿	INTERRINI	abrir interrogación
?	INTERRFIN	pechar interrogación
:	DOUSPUN	dous puntos
:	DOUSPUNAP	dous puntos e aparte
"	COMIDOB	comiñas dobres
'	COMISIM	comiñas simples
-	GUION	guión
-	GUIONLONGO	guión longo (raia)
¡	EXCLAINI	abrir exclamación
!	EXCLAFIN	pechar exclamación
[CORCHINI	abrir corchete
]	CORCHFIN	pechar corchete

Táboa 18: Inventario de etiquetas de puntuación

3.3. Etiquetas compostas

Hai algunhas características do galego que requiren un tratamento específico na anotación morfosintáctica de corpus nesta lingua. Por unha banda, a “fusión” de dúas palabras nunha mesma palabra ortográfica, propia das contraccións, dos enclíticos e da segunda forma do artigo. No esquema de codificación do *sli*, este conxunto de fenómenos reciben unha etiqueta “composta”, formada pola etiqueta da primeira palabra seguida das etiquetas correspondente(s) á(s) palabra(s) ligada(s), todas elas separadas polo signo “_”. Deste xeito, a contracción da preposición “en” e o pronome persoal masculino singular de 3ª persoa “el” (isto é, “nel”) recibe a etiqueta “PREP_PPMS3”; o verbo con enclítico “díxome” leva a etiqueta composta “VIPES3_PPS1A”, formada coas anotación para “verbo indicativo pretérito terceira singular” e “pronome persoal átono singular primeira”; a “acercóuselle”, con dous enclíticos, correspóndelle a etiqueta “VIPES3_PPS3AR_PPS3AD”, composta coas anotacións para “verbo indicativo pretérito terceira singular”, “pronome persoal singular átono terceira reflexivo” e “pronome persoal átono singular terceira dativo”, mentres que a “díxomo”, con dous enclíticos (“me” e “o”) en alófono, se lle asigna a etiqueta composta “VIPES3_PPS1A_PPMS3AA”, coas etiquetas correspondentes a “verbo indicativo pretérito terceira singular”, “pronome persoal átono singular primeira” e mais “pronome persoal átono masculino singular terceira acusativo”.

No tocante ás segundas formas do artigo, emprégase o mesmo sistema para a composición das etiquetas compostas, diferenciando a segunda forma do artigo mediante a adición da anotación “-2” na súa etiqueta. Así, “bebelo” (“beber”+“o” artigo) codifícase coa etiqueta composta VINF_ARDMS-2, formada coas etiquetas de “verbo infinitivo” e de “artigo determinado masculino singular segunda forma”. As segundas formas do artigo tamén poden unirse con guión a un pronome enclítico, como en “gústalle-lo” (“gusta”+“lles”+“o”), que recibiría a marca “VIPRS3_PPP3AD_ARDMS-2”, composta polas etiquetas de “verbo indicativo presente terceira singular”, “pronome persoal átono plural terceira dativo” e “artigo determinado masculino singular segunda forma”. Utilízase o mesmo sistema de anotación coas segundas e terceiras formas dos pronomes persoais átonos acusativos de terceira persoa, de xeito que “bebelo” (“beber”+“o” pronome) codifícase como VINF_PPMS3AA-2 (“verbo infinitivo”+“pronome persoal átono masculino singular terceira acusativo na forma segunda”), e “colleuna” (“colleu”+“a”) como VIPES3_PPFS3AA-3 (“verbo indicativo pretérito terceira singular”+“pronome persoal átono feminino singular terceira acusativo na forma terceira”).

Por outra banda, outro trazo do galego que esixe un tratamento específico na anotación morfosintáctica son as locucións, particularmente as locucións

prepositivas, conxuntivas e adverbiais. Na súa codificación óptase por unha solución na liña do proposto por Sampson (1995), isto é, cando unha palabra forma parte dunha locución, asígnaselle a etiqueta correspondente á súa categoría, seguida do signo “_”, da etiqueta correspondente á categoría da locución, e dun número de dúas cifras onde o primeiro é o número de palabras que integran a locución e o segundo, o número correspondente ó lugar que ocupa a palabra dentro da locución. Por exemplo, na locución prepositiva “cara a”, a “cara” asígnaselle a etiqueta composta NCFS_PREP21, mentres que a “a” lle corresponde a etiqueta “PREP_PREP22”.

3.4. Etiquetaxe e aliñamento

Os textos etiquetados no *cluvi* respectan o estándar *xml* e inclúen, no caso dos corpus paralelos, información morfosintáctica e información sobre equivalencias de tradución. A especificación *cluvi-tmx* para os documentos da sección paralela do *cluvi* é unha versión modificada do tipo de documento definido mediante o estándar *tmx*¹¹. A definición do tipo de documento (*dtd*) válido no formato *cluvi-tmx*, que se pode consultar na Figura 7, inclúe a información morfosintáctica nun elemento “<ling>”, ausente do *tmx* e empregado aquí para etiquetar tódalas palabras e signos de puntuación dos elementos “<seg>” da estrutura *tmx* orixinal.

```

        <!-- DTD do formato CLUVI-TMX>
<!ELEMENT cluvi-tmx (header, body) >
<!ATTLIST cluvi-tmx
  version CDATA #REQUIRED >
<!ELEMENT header (#PCDATA)>
<!ATTLIST header
  creationtool CDATA #REQUIRED
  creationtoolversion CDATA #REQUIRED
  segtype (block|paragraph|sentence|phrase) #REQUIRED
  o-tmf CDATA #REQUIRED
  adminlang CDATA #REQUIRED
  srclang CDATA #REQUIRED
  datatype CDATA #REQUIRED >
<!ELEMENT body (tu*) >
<!ELEMENT tu (tuv+) >
<!ELEMENT tuv (seg) >
<!ATTLIST tuv
  lang CDATA #REQUIRED>
<!ELEMENT seg (ling+)>
<!ELEMENT ling (mor, ort)>
<!ELEMENT mor EMPTY>
<!ATTLIST mor
  cat (ARDFP|ARDFS|ARDMP|ARDMS...) #REQUIRED
  lema CDATA #REQUIRED
  lema2 CDATA #IMPLIED>
<!ELEMENT ort (#PCDATA)>

```

Figura 8: *dtd* do *cluvi-tmx*

¹¹ Pódese consultar a especificación *tmx* en <http://www.lisa.org/tmx/>.

Finalmente, pódese comprobar a aplicación do formato *cluvi*-TMX no corpus *cluvi* na Figura 8, onde se reproduce un fragmento do TECTRA que contén unha unidade de tradución etiquetada morfosintacticamente e aliñada. No exemplo, elimínase a información morfosintáctica relativa á variante en lingua inglesa da unidade de tradución para facilitar a súa comprensión.

```

<tu> <tuv lang="en"> <seg>In the town they tell the story of
the great pearl.</seg> </tuv> <tuv lang="gl"> <seg><ling>
<mor lema="en" lema2="o" cat="PREP_ARDFS"/> <ort>Na</ort>
</ling> <ling> <mor lema="cidade" cat="NCFS"/>
<ort>cidade</ort> </ling> <ling> <mor lema="contar"
lema2="se" cat="VIPRS3_PPS3AR"/> <ort>cóntase</ort> </ling>
<ling> <mor lema="o" cat="ARDFS"/> <ort>a</ort> </ling>
<ling> <mor lema="historia" cat="NCFS"/> <ort>historia</ort>
</ling> <ling> <mor lema="de" lema2="o" cat="PREP_ARDFS"/>
<ort>da</ort> </ling> <ling> <mor lema="grande"
cat="AXAPFS"/> <ort>gran</ort> </ling> <ling> <mor
lema="perla" cat="NCFS"/> <ort>perla</ort> </ling> <ling>
<mor lema="." cat="PUNTO"/> <ort>.</ort> </ling> </seg>
</tuv> </tu>

```

Figura 9: Exemplo do TECTRA

4. Conclusións

Neste artigo presentamos un etiquetario morfosintáctico completo e normalizado (consonte ós estándares europeos definidos por EAGLES) para etiquetar corpus lingüísticos de lingua galega. Coa presentación do etiquetario *s/li* para o galego, e das correspondencias entre o etiquetario *s/li* do galego e o etiquetario intermedio proposto por EAGLES, pretendemos contribuír ó avance da investigación e desenvolvemento nas áreas da lingüística de corpus e das tecnoloxías lingüísticas da lingua galega.

REFERENCIAS BIBLIOGRÁFICAS

- AGUIRRE MORENO, J.L., A. ÁLVAREZ LUGRÍS, I. BRAGADO, L. CASTRO, X. GÓMEZ GUINOVART, S. GONZÁLEZ LOPO, A. LÓPEZ LÓPEZ, J.R. PICHEL, E. SACAU e L. SANTOS. 2003. "Alinhamento e etiquetagem de corpora paralelos no *CLUVI* (Corpus Lingüístico da Universidade de Vigo)". En Almeida, J.J. (ed.), *Actas do Workshop CP3A, Corpora Paralelos: Aplicações e Algoritmos Associados*. Braga: Universidade de Braga.
- AGUIRRE MORENO, J.L., A. ÁLVAREZ LUGRÍS e X. GÓMEZ GUINOVART. 2002. "Etiquetario morfosintáctico del *SLI* para corpus de lengua gallega: aplicación al corpus paralelo TECTRA". En *Procesamiento del Lenguaje Natural*, 28, pp. 23-34.
- AGUIRRE MORENO, J.L., N. ANDIÓN e X. GÓMEZ GUINOVART. 2001. "Aspectos ortográficos, léxicos y morfosintácticos del etiquetado lingüístico de un corpus de informática en lengua gallega". En *Procesamiento del Lenguaje Natural*, 27, pp. 13-19.
- ÁLVAREZ, R., X.L. REGUEIRA e H. MONTEAGUDO. 1986. *Gramática galega*. Vigo: Galaxia.
- ÁLVAREZ, R. e X. XOVE. 2002. *Gramática da lingua galega*. Vigo: Galaxia.
- ÁLVAREZ LUGRÍS, A. 2001. *Estilística comparada da traducción: Proposta metodolóxica e aplicación práctica ó estudio do corpus tectra de traduccions do inglés ó galego*. Vigo: Universidade de Vigo.
- LEECH, G. e A. WILSON. 1996. *Recommendations for the Morphosyntactic Annotation of Corpora. eagles Guidelines*. Disponible en: <http://www.ilc.pi.cnr.it/eagles96/annotate/annotate.html>.
- MELBY, A.K. 2000. "Sharing of translation memory databases derived from aligned parallel text". En J. VÉRONIS (ed.), *Parallel Text Processing. Alignement an Use of Translation Corpora*, pp. 347-368. Dordrecht: Kluwer.
- MONACHINI, M. e N. CALZOLARI (coords.) 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. e N. CALZOLARI. 1999. "Standardization in the Lexicon". En H. VAN HALTEREN (ed.), *Syntactic Wordclass Tagging*, pp. 149-174. Dordrecht: Kluwer.
- RAG/ILG. 1982. *Normas ortográficas e morfolóxicas do idioma galego*. RAG/ILG: Vigo (13ª edición revisada: 1995).
- SAMPSON, G. 1995. *English for the Computer*. Oxford: Oxford University Press.

SANTOS, D. 1999. “Disponibilización de corpora de texto através da WWW”. *Linguística computacional: investigación fundamental e aplicacións*, pp. 323-335. Lisboa: Edições Colibri/Associação Portuguesa de Linguística.

TEUFEL, S. 1996. *ELM-DE: eagles Specifications for German Morphosyntax. eagles Guidelines*. Disponível en: http://www.ilc.pi.cnr.it/eagles96/elm_de/elm_de.html.

ANEXO 1: TECTRA (SECCIÓN INGLÉS)

A continuación preséntase unha lista cos títulos dos textos orixinais ingleses que forman parte do corpus TECTRA, co nome da persoa autora, ano de publicación e tamaño en número de palabras.

Orixinal inglés, autor/a (ano de publicación)	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

Táboa 17: TECTRA (sección inglés)

ANEXO 2: TECTRA (SECCIÓN GALEGO)

Neste segundo anexo recollémo-la lista dos títulos das traducións ó galego que forman parte do corpus TECTRA, co nome da persoa traductora, ano de publicación e tamaño en número de palabras.

Tradución galega, traductor/a (ano de publicación)	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 galego	551.878

Táboa 18: TECTRA (sección galego)

ANEXO 3: REFERENCIAS BIBLIOGRÁFICAS DO CORPUS TECTRA

Por último, incluímo-las referencias bibliográficas correspondentes ás novelas (orixinais en inglés e traducións ó galego) incluídas no corpus paralelo TECTRA:

- BELL, A.F.G. 1922. *Spanish Galicia*. John Lane The Bodley Head, Londres.
1994. *Galicia vista por un inglés*. Galaxia, Vigo.
- GOLDING, W. 1954 (1962). *Lord of the Flies*. Faber & Faber, Londres.
1993. *O Señor das moscas*. Sotelo Blanco, Santiago de Compostela.
- GREENE, G. 1950 (1974, 3ª ed.). *The Third Man*. Heinemann, Londres.
1994. *O terceiro home*. Galaxia, Vigo.
- JOYCE, J. 1916 (1986, 11ª ed.). *A Portrait of the Artist as a Young Man*.
Grafton Books, Londres.
1994. *Retrato do artista cando novo*. Laiovento, Santiago de Compostela.
- LONDON, J. 1903 (1975, 16ª ed.). *The Call of the Wild*. Heinemann,
Londres.
1982 (1983, 2ª ed.). *A chamada da selva*. Xerais, Vigo.
- ORWELL, G. 1945 (1987, 58ª ed.). *Animal Farm*. Penguin, Londres.
1992. *A revolta dos animais*. Positivas, Santiago de Compostela.
- SALINGER, J.D. 1948 (1986). *For Esmé, with Love and Squalor*. Penguin,
Nova York.
1994. *Nove contos*. Sotelo Blanco, Santiago de Compostela.
- SALINGER, J.D. 1951. *The Catcher in the Rye*. Penguin, Londres.
1990 (1992, 4ª ed.). *O vixia no centeo*. Xerais, Vigo.
- SINGER, I. B. 1982. *The Golem*. Penguin, Londres.
1989. *O Golem*. Xerais, Vigo.
- STEINBECK, J. 1945 (1986, 21ª ed.). *The Pearl*. Penguin, Nueva York.
1990 (1993, 5ª ed.). *A perla*. Galaxia, Vigo.
- STERNE, L. 1768 (1995). *A Sentimental Journey Through France and Italy*.
Wordsworth, Hertfordshire.
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.
1992. *Retallos do diario de Adán*. Positivas, Santiago de Compostela.
- TWAIN, M. 1906 (1993). *Eve's Diary*. Courage Books, Filadelfia.
1991. *Diario de Eva*. Positivas, Santiago de Compostela.
- WOLF, V. 1927 (1977, 18ª ed.). *To the Lighthouse*. Grafton Books, Londres.
1993. *Cara ó faro*. Sotelo Blanco, Santiago de Compostela.

