

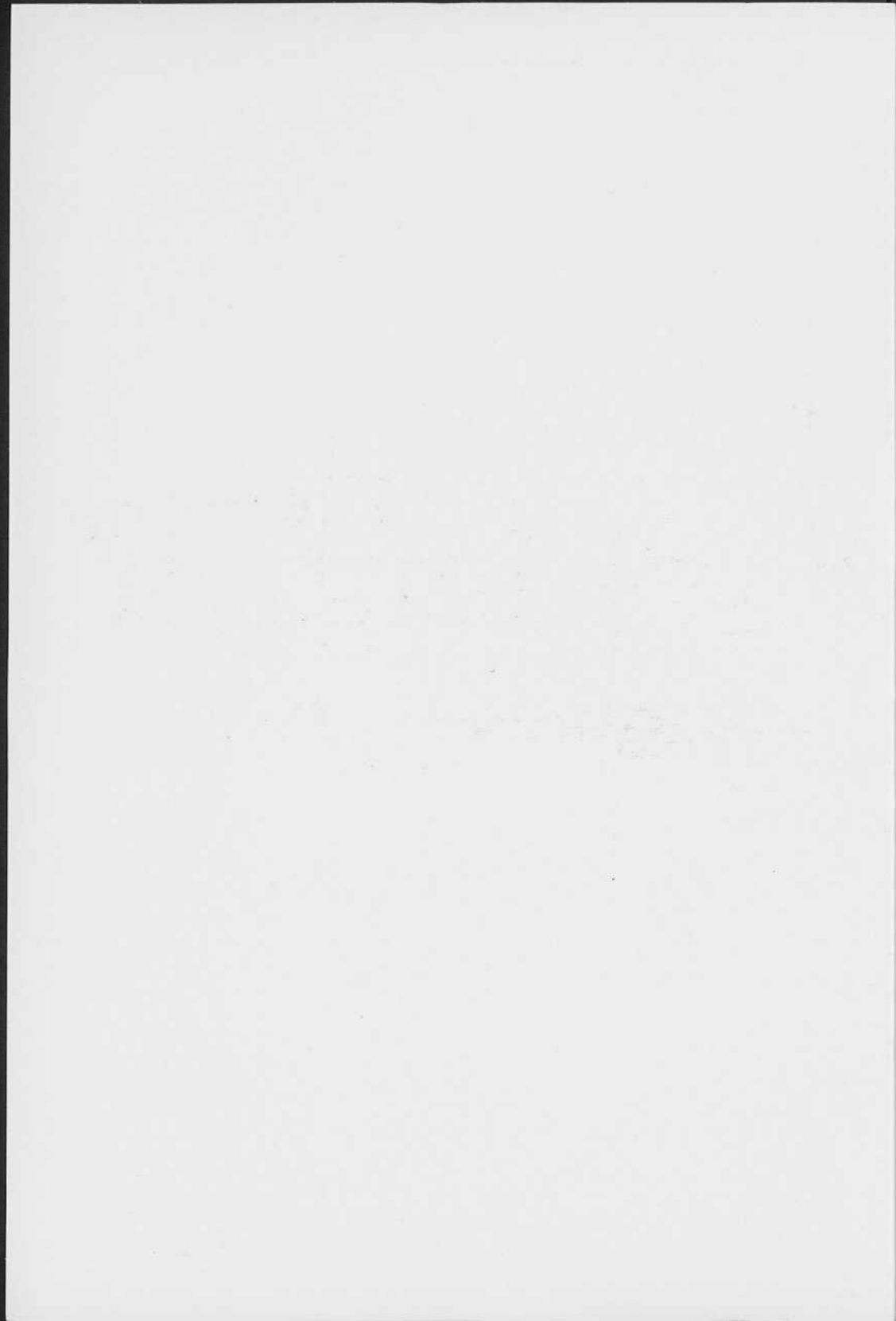
de Cadernos
Lingua

2008/2009

30/31
núm.

REAL ACADEMIA
GALEGA





R (05)/3

CADERNOS DE LINGUA

ANO 2008-2009

30-31

**R E A L
ACADEMIA
GALEGA**



Director: Manuel González González
Coordinador: Xosé Luís Regueira
Secretaria: María do Carme Pazos Balado

Consello de redacción:

Carlos Díaz Abreira, Xermán García Cancela, Antón López Dobao,
 Miguel Pérez Pereira, Modesto A. Rodríguez Neira

Comité científico:

R. Álvarez Blanco (Un. Santiago), J.A. Argente Giralt (Un. Aut. Barcelona), Takekazu Asaka (Un. Azabu), Michel Contini (Un. Stendhal de Grenoble), F. Fernández Rei (Un. Santiago), P. García Mouton (CSIC Madrid), Taina Hämäläinen (Un. Helsinki), Johannes Kabatek (Un. Tübingen), R. Lorenzo Vázquez (Un. Santiago), C. de Azevedo Maia (Un. Coimbra), David Mackenzie (Un. Corcaigh), Lorenzo Massobrio (Un. Torino), Michael Metzeltin (Un. Viena), A. Santamarina Fernández (Un. Santiago), M. Santos Rego (Un. Santiago)

Colaboracións e correspondencia:

Xosé Luís Regueira Fernández
 Paseo da Quinta, 23
 15897 Santiago de Compostela

Subscripción e intercambio: Dirixirse a

Editorial Galaxia, S. A.
 r/. Reconquista, 1 - 36201 Vigo

©, Real Academia Galega
 A Coruña, 2009

Edita: Real Academia Galega
 r/. Tabernas, 11 - 15001 A Coruña

I.S.S.N.: 1130-5924

Título clave: Cadernos de Lingua
 Título abreviado: Cad. Ling.

Depósito Legal: C-1.642-1992

Deseño: Sinmás Comunicación Visual S.L.
 Imprime: Galigraf Galicia S.L.

ÍNDICE

ARTIGOS

- X.-H. COSTAS GONZÁLEZ, *A deturpación da toponimia galega do Eo-Navia* 5
- S. DOMÍNGUEZ PORTELA, *Olla e mira*, dous marcadores discursivos en tres linguas: portugués, galego e español 27
- A. B. ESCOURIDO PERNAS, *A prosodia de Camelle: descrición acústica dunha fala* 75
- A. RODRÍGUEZ GUERRA, *A oración en galego medieval (AOGM): O banco de datos sintácticos medievais do ILG* 123
- E. DOMÍNGUEZ NOYA, FCO. M. BARCALA RODRÍGUEZ, M. A. MOLINERO, *Avaliación dun etiquetador automático estatístico para o galego actual: Xiada* 151

RECENSIONS

- Petit atlas lingüístic del domini catalá*, J. VENY (M. González González) 195
- Historia de la lengua gallega*, R. MARIÑO PAZ (E. Moscoso Mato) 199



THE UNIVERSITY OF CHICAGO
LIBRARY

1910

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

A DETURPACIÓN DA TOPONIMIA GALEGA DO EO-NAVIA

XOSÉ-HENRIQUE COSTAS GONZÁLEZ

Universidade de Vigo

RESUMO

No territorio asturiano do Eo-Navia estase efectuando un proceso acelerado de dobre substitución lingüística e cultural. Á imposición histórica do español únese agora a imposición política de usos lingüísticos e culturais da Asturias central. Iso ten un forte impacto na toponimia desta zona, e os topónimos galegos galegos autóctonos están a ser deturpados. Coa nova lexislación, as autoridades lingüísticas e culturais galegas, así coma a Xunta de Galicia, poderían axudar á restauración das formas toponímicas auténticas e autóctonas.

ABSTRACT

"The misrepresentation of Éo-Navia Galician Toponymy"

In the Asturian region of Eo-Navia an accelerated double process of linguistic and cultural substitution is taking place. The historical imposition of the Spanish language is joined to the political imposition of linguistic and cultural Central Asturian uses. This has a strong impact on the toponymy of this area, as the autochthonous Galician place names are being altered. With the new legislation, the Galician linguistic and cultural authorities and regional Government could help restore the authentic and native place names.

PALABRAS CHAVE

Toponimia, onomástica, lexislación lingüística, substitución lingüística.

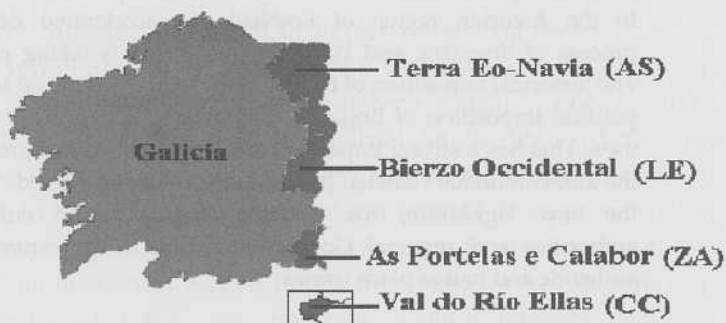
KEY WORDS

Toponymy, onomastics, linguistic legislation, language shift.

1. *Limiar*

Se analizamos a toponimia dos concellos asturianos situados entre os ríos Eo e Navia, podemos recoñecer doadamente moitísimos topónimos que nos son familiares por galegos ou por portugueses do norte. *Boiro* haino en Ibias, *Ferrol* en Castropol, *Xinzo* en Tapia, *Paramios* na Veiga, *A Lama* en Taramundi, *Vilameá* en San Martín de Ozcos e mesmo hai dous *Hío*, un en San Martín e outro en Pezós. Ata hai un río *Valcarce* en Ibias distinto do río Valcarce que se une ao Burbia en Vilafranca do Bierzo. En terras onde se fala a mesma lingua o habitual é a repetición formal de moitos topónimos: *Mourelle*, *Pacios*, *A Grova*, *Salgueira*, *Outeiro*, *Barreiros*, *Xestoso*, *Santalla*, *Ferreira*, *Quintela*, *Vilela*, *Valiñas*, *Vilarello*, *Penouta*, *Sestelo* etc.

Así e todo, ao contrario da Galicia administrativa, onde desde 1983 a toponimia ten como forma única a galega (Lei de Normalización Lingüística, art. 10), nos territorios de lingua e cultura galegas arraianas polo leste asistimos, por parte da maioría da oficialidade, a uns usos toponímicos absolutamente deturpadores e aberrantes, herdeiros do máis escuro e ignorante centralismo.



Mapa 1

Territorios de lingua autóctona galega (elaboración propia)

2. A legalidade toponímica no galego estremeiro e valego

2.1. Castela-León

En Castela e León o galego está recoñecido no Estatuto de Autonomía de 1999 (Lei orgánica 4/83), concretamente no seu art. 4.2., onde se di que: "Gozarán de respeto y protección la *lengua gallega* y las modalidades lingüísticas en los lugares en que habitualmente se utilicen".

Pero o art. 24.1 da Lei 1/98, de 4 de xuño, de réxime local de Castela-León, en relación coa toponimia oficial dispón que "la denominación de los municipios habrá de ser en lengua castellana".

Se se recoñece a existencia do galego, deberíanse respectar as formas toponímicas das zonas onde se fala, e máxime cando o art. 10.2.g) da *Carta Europea das Linguas Rexionais e Minoritarias* (CELRM) documento asinado e ratificado por España, e que forma parte do bloque de constitucionalidade da lexislación española, di literalmente que:

As partes comprométense ao emprego ou adopción, no seu caso conxuntamente coa denominación na lingua oficial, das formas tradicionais e correctas da toponimia nas linguas rexionais ou minoritarias.

O rango lexislativo desta CELRM é, por suposto, moi superior a calquera norma ou disposición autonómica. Porén, agás o deputado socialista e académico correspondente da RAG Felipe Lubián, ninguén no Parlamento de Castela-León está disposto a defender que *Barxas* poida ser *Barxas* e non *Barjas* (castelán *Bárcenas*), *A Teixeira* poida ser *A Teixeira* e non *La Tejera* etc. Repárese que a deturpación toponímica é incoherente; trádúcese algúns topónimos máis ou menos transparentes (*Villa de Canes*, *Herrería*, *Villar de Acero* etc., por *Vila de Cais*, *Ferrería*, *Vilar de Aceiro* etc.), adáptanse outros (*Quilós*, *Barjas*, *Borrenes*, *Veguellina* etc., por *Quilous*, *Barxas*, *Borrés*, *A Veigueliña* etc.), pero a maioría conservan a súa forma tradicional galega inalterada:

Porto, Pías, Sobrado, Vilela, Valouta etc. Se Castela-León fose coherente co art. 24.1 da Lei 1/98 de réxime local debería impor en castelán tamén estes topónimos e así sairían cousas como **Puerto, *Pilas, *Alcornocal, *Villilla, *Vallota* etc.

2.2. Estremadura

Máis indefensa está aínda a toponimia dos tres concellos cacerreños de lingua histórica galega. No Estatuto de Autonomía de Estremadura nada se di verbo do respecto ou recoñecemento doutras realidades lingüísticas distintas do castelán, enténdase o galego do Val do Elas e o portugués de varios concellos fronteirizos. Unicamente en 2000 houbo unha declaración da “Fala” (*sic*) como Ben de Interese Cultural, sen que isto supoña na práctica nada de nada. A Junta cala ante a demanda de oficialidade toponímica bilingüe. A nivel local, Valverde ten toda a toponimia “urbana” e municipal en aparente castelán (*Calle Levadiña, Calle del Forno* etc), San Martiño de Trebello todo bilingüe e As Elas todo monolingüe en lagarteiro ou galego das Elas.

No informe elaborado en 2004 pola Comisión de expertos do Consello de Europa sobre a aplicación en España da *Carta Europea das Linguas Rexionais ou Minoritarias* (CELRM) o galego de Estremadura non aparecía por ningures, mentres que si se citaban os casos dos galegos de Asturias e de Castela-León. Por iso diciamos que estaba menos protexido, porque nin sequera os expertos do Consello de Europa sabían da súa existencia. No novo informe elaborado en 2008 xa se recoñece a existencia deste enclave galegófono en Estremadura, a instancias da Secretaría Xeral de Política Lingüística e de organizacións cívicas como a Mesa pola Normalización Lingüística, que foron quen informaron da súa existencia aos expertos europeos. Neste novo informe reclámaselles ás autoridades españolas e extremeñas que dean noticias sobre a aplicación da CELRM neste enclave.

3. A situación toponímica en Asturias

3.1. Disposicións legais

A diferenza de Castela-León ou Estremadura, onde o galego continúa padecendo en distinto grao a presión lingüística histórica do castelán, en Asturias o galego vive hoxe unha dobre imposición: a histórica do castelán e a máis recente dun emergente asturiano.

O vixente Estatuto Autonomía de Asturias de 1983 nada di da existencia do galego, e si unicamente do "bable", especificando no seu artigo 10.1.15 que este gozará de "fomento y protección en todas as súas variantes". Enténdese no Estatuto asturiano que o galego de Asturias é simple variante do "bable" ou nin sequera se considera a súa existencia?

Na Lei 1/1998 de 23 de marzo, de uso e promoción do Bable/Asturiano, dise no seu artigo 2º e nunha disposición adicional especifícase que todo o que se lle aplique ao bable/asturiano tamén lle será aplicado ao gallego/asturiano nas zonas onde é modalidade lingüística propia¹, sen que ninguén aclare por agora cales son esas zonas, quizais porque interesa a indefinición.

Nesta mesma lei, o art. 15 referíase á toponimia do seguinte xeito:

Los topónimos de la Comunidad Autónoma del Principado de Asturias *tendrán la denominación oficial en su forma tradicional*. Cuando un topónimo tenga uso generalizado en su forma tradicional y en castellano, la denominación podrá ser bilingüe.

¹ Art. 2.1: "El régimen de protección, respeto, tutela y desarrollo establecido en esta Ley para el bable/asturiano se extenderá, mediante regulación especial al gallego-asturiano en las zonas en las que tiene carácter de modalidad lingüística propia."

Disp. Adicional: "El gallego-asturiano tendrá un tratamiento similar al asturiano en lo que se refiere a protección, respeto, enseñanza, uso y tutela en su ámbito territorial."

De acuerdo con los procedimientos que reglamentariamente se determinen, corresponde al Consejo de Gobierno, previo *dictamen de la Junta de Toponimia del Principado* de Asturias, y sin perjuicio de las competencias municipales y estatales, determinar los topónimos de la Comunidad Autónoma.

Porén, a pesar desta fermosa declaración de boas intencións, a realidade é teimosa e demostra con feitos unha e outra vez que, polo menos desde 1986, vén tendo lugar no Eo-Navia unha belixerante campaña da clase política e académica asturiana (con algunhas excepcións) contra todo o que soe a galego no occidente asturiano: cursos só de lingua asturiana en territorio galegófono; cursos de verán anuais de asturiano da Academia de la Llingua Asturiana en Ibias, en territorio galegófono; forte apoio institucional, económico e mediático a asociacións socioculturais pro-asturianistas e antigalegas como Xeira; introdución da gaita asturiana en lugar da tradicional da zona, a galega; protestas reiteradas pola recepción da TVG en todo o occidente astur; cubrición do sinal da TVG en todo o Eo-Navia co sinal da nova TV autonómica asturiana etc. As acusacións de “inxerencia de Galicia en terra asturiana” ou “imperialismo galego” ante calquera actividade, declaración ou mobilización pola lingua están á orde do día na prensa asturiana e na boca dos líderes políticos de todos os partidos, maiormente os asturianistas.

Na zona lingüística e culturalmente galega do occidente de Asturias estase operando un fenómeno de transculturización acelerada. A lingua que se quere potenciar -minimamente-, desde os poderes políticos, mediáticos e académicos non é o autóctono galego eonaviego, senón o asturiano estándar ou, en moita menor medida, un galego eonaviego vulgarizante ortografado segundo as directrices da ALLA e da Consejería de Cultura.

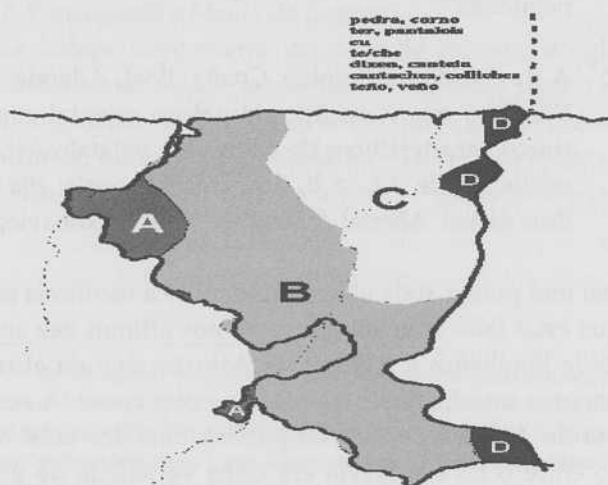
Desde os gobernos galegos nunca se lle prestou demasiada atención a este glotoculturicidio aplicando a máxima da “non inxerencia” nas competencias doutras comunidades autónomas. O caso máis sangrante foi o do último presidente de Galicia, Pérez Touriño, quen reiteradamente excluiu este tema da

axenda de conversas que frecuentemente mantivo ao longo destes últimos anos co seu homólogo asturiano, Álvarez Areces.

Os medios de comunicación galegos tampouco se fixeron nunca moito eco desta realidade, por iso a opinión pública galega viviu allea a ela, e unicamente unha produtora madrileña (Pascana Films) realizou dúas reportaxes: *Fronteiras* e *A terceira póla*, ambas as dúas dirixidas por Rubén Pardiñas, que recibiron duras críticas desde os medios e os partidos asturianos.

3.2. As variedades galegas no Eo-Navia e a súa consideración social

En xeral, son galegas as falas que quedan ao occidente dunha liña imaxinaria que inclúe aquelas que non ditongan os Ö, Ë tónicos latinos (*pedra, vella, nova, corno*), onde se produciu a caída do -N- latino (*ter, corazois, chea*), onde teñen o pronome persoal *eu* e distinguen *chel/te*, onde teñen formas verbais como *teño, teña, veño, veña, cantaches, colliches, durmiches, cantei(n), collín, durmín, dixer, houben...* e outros moitos trazos máis, pero dentro desta grande área do Eo-Navia fálanse tres modalidades lingüísticas galegas (V. mapa 2):



Mapa 2

Variedades lingüísticas galegas en Asturias: en gris forte (A) as falas de tipo *taramundés*, cun grao de identidade estrutural co

galego oriental do 100%; en gris claro (B) as falas de tipo *asturnegueirés*, cunha proximidade do 92%; en branco (C) as falas de tipo naviego, cunha coincidencia do 85%; en gris escuro (D) falas galegas de transición ao asturiano occidental (elaboración a partir dos datos de Babarro 2003 recollidos posteriormente en Asociación Abertal 2007:49-64)

- a) A empregada en Santiso de Abres e Taramundi, e nos lugares da Trapa (Santalla) e Os Coutos e Valdeferreiros (Ibias), galego oriental normal que Abertal (2007: 57-64) denomina “de tipo taramundés”.
- b) A usada nos concellos de Castropol, A Veiga, Tapia, Eilao, Vilanova, Santalla, San Martín, Pezós, Grandas e Ibias, mais tamén no concello lugués de Negueira de Muñiz e nos lugares de Porto de Baxo (Ribadeo), Pántaras (A Fonsagrada) e A Peliceira (Navia de Suarna), con presenza do artigo masculino *el*, conservación do -L- intervocálico etc: *ela non mole nese tou molín*. Abertal denomínaas “falas de tipo asturnegueirés”.
- c) A da Caridá (El Franco), Coaña, Boal, Allande, Navia e Villallón, que é un tipo de galego oriental con algúns trazos característicos do asturiano: palatalización do L-, evolución de -LL- > ll, fundamentalmente: *ella levou el lleite da mai*. Abertal chámalles “falas baixonaviegas”.

Ata hai moi pouco, toda a iberorromanística recoñecía sen dubidar que estas falas eran galegas, pero nos últimos dez anos a oficialidade lingüística e cultural de Asturias deu en afirmar con argumentos extralingüísticos que son outra cousa. A actual presidenta da ALLA recoñecía en publicacións dos anos 80 que o falado entre o Eo e o Navia era unha variedade do galego de Lugo (Cano González 1980: 43-44, 1987: 151-152). Vinte anos máis tarde, a ALLA, a súa presidenta e a oficialidade mediática e política asturianas manteñen que as falas usadas entre Navia e

Eo son un *continuum* lingüístico entre galego e asturiano que conforma unha nova lingua (Cano González 1992: 652; ALLA 2006), por suposto unha nova lingua asturiana, pola razón de que os seus falantes son asturianos, e, xa que logo, unha nova lingua tutelada pola Academia de la Llingua Asturiana. Deste xeito, no seo da ALLA funciona nominalmente unha Secretaría Llingüística para el Navia-Eo. Este é o único caso no mundo en que unha academia tutela dúas linguas. E o sorprendente é que a ALLA e o Parlamento asturiano lle chamen a esa nova lingua “Fala”, de maneira que para toda a nomenclatura oficial en Asturias fálanse a “lingua asturiana” e a “fala”², agora xa nin o invento legalizado de “gallego-asturiano”. É decer, segundo as autoridades asturianas, en Asturias hai castelanfalantes, asturianofalantes e “falofalantes”.

Se o chamado en Asturias “gallego-asturiano” ou “fala” fose unha nova lingua, entón poderíamos doadamente deducir que Asturias sería tetralingüe (galego –en Santiso de Abres, Taramundi e outros puntos–, galego-asturiano –no resto do Eo-Navia–, asturiano e castelán) e Galicia sería trilingüe (galego, castelán e galego-asturiano en Negueira de Muñiz e puntos de Ribadeo, A Fonsagrada e Navia de Suarna).

Mais, independentemente de como lle chamen, do glotónimo, o fundamental é que non se toman medidas para a normalización lingüística no Eo-Navia, senón que o que se pretende é asturianizar ou adormecer o problema. Hai moitos centros de ensino do Eo-Navia onde non se imparte “gallego-asturiano”

² O 15 de decembro de 2005 o Parlamento de Asturias aprobaba por unanimidade unha moción pola cal se recoñecía a *asturianía incuestionable* da “Fala” (sic), cando na lexislación do Principado só se fala de “gallego-asturiano”, con guión, ou “gallego/asturiano”, con barra, ao lado do “bable/asturiano”. Son asturianas tanto a “llingua asturiana” coma a “fala”. A consideración dunha como *llingua* e da outra como *fala* xa expresa ben ás claras a consideración social das dúas que teñen os dirixentes políticos asturianos: unha é unha lingua de verdade, e amais é a asturiana por antonomasia, a outra é unha fala, un resquicio folclórico pouco serio.

porque non se ofrece, e mesmo hai algúns onde o que se imparte desde hai varios anos é asturiano puro no estándar central³.

3.3. Estandarizar a lingua para normalizar a toponimia

O galego desta zona presenta unhas características lingüísticas moi especiais que farían recomendable que, dadas as condicións sociolingüísticas, polo menos transitoriamente funcionase unha norma con concesións dialectais.

Así, en 1990 a MDGA publicou as *Normas ortográficas e morfolóxicas del galego de Asturias*, unha adaptación das normas da RAG-ILG de 1980 ás particularidades lingüísticas e sociais do galego en Asturias. Os autores eran coñecidos: Antón Meilán, Antón Santamarina, Xoán Babarro, Benigno Fernández Braña, Francisco Fernández Rei etc., é decer, consumados especialistas en dialectoloxía e standardoloxía galegas e falantes formados e conscientes de galego eonaviego.

Tres anos máis tarde da publicación das *Normas* da MDGA, en 1993, reaccionou o Principado e publicou unha *Proposta de normas ortográficas y morfolóxicas del gal(l)ego-asturiano*, editada polo Serviciu de Publicaciones del Principau. Non se sabe quen foron os expertos aos que recorreu o Principado, pero o evidente é que esta *Proposta* é un calco das *Normas* da MDGA con concesións á vulgaridade e interferencias forzadas da norma do asturiano. Vexamos as (mínimas) diferenzas no seguinte texto:

Texto na *Norma* elaborada pola MDGA:

FV,JA: *Notas etnolingüísticas del concello da Veiga*. Xeira Laboría, 2, A Caridá, 1997, 108 páxinas [con ilustracións de X.]

³ O 2 de xuño de 2006 en toda a prensa asturiana a Asociación Xeira denunciaba a precariedade da escolarización en "gallego-asturiano", manifestando que eran moi poucos os centros de ensino onde se impartía e que mesmo había centros, como o de Luña-Ibias, onde o que se viña impartindo ano tras ano era asturiano central estándar.

JAFV, catedrático de Lingua e Literatura Españolas nel Instituto Xovellanos de Xixón dende hai xa ben de cursos, doctorouse en novembre de 1996 na Universidad de Uvieu cuha tese que estudiaba a fala da Veiga e el sou concello. Era el resultado de ben de anos recollendo material oral e falando con xente vella de todos os sitios del concello. El estudio da fala e el léxico, publicados entrambos e dous aspectos nos libros El habla de Vegadeo (A Veiga y su concejo) (1997) e Vocabulario da Veiga (1998), completábase na tese cuhas notas etnolingüísticas sobre A Veiga que son as que agora amañan este volume. El ser humano, as súas edificacións, as súas ferramentas, a labranza, a pesca, el ferro, a madeira, os nomes cos que se describe el tempo, os animales, a comida, as festas... todo contado dende A Veiga e cun xeito e uha precisión que namáis dan os anos de traballo e dedicación que el autor leva ás costas.

Texto na Proposta do Principado (Consejería de Cultura):

FV, JA: Notas etnolingüísticas del conceyo da Veiga. Xeira Llaboría, 2, A Caridá, 1997, 108 páxinas [con ilustracións de X.]

JAFV, catedrático de Llingua y Lliteratura Españolas nel Instituto Xovellanos de Xixón dende hai xa dellos cursos, doctorouse en novembre de 1996 na Universidad d'Uviéu con úa tesis qu'estudiaba a fala da Veiga y el sou conceyo. Era el resultao de ben d'anos recovendo material oral y falando con xente veya de todos os sitios del conceyo. El estudio da fala y el léxico, publicaos entrambos y dous aspectos nos llibros El habla de Vegadeo (A Veiga y su concejo) (1997) y Vocabulario da Veiga (1998), completábase na tesis con úas notas etnolingüísticas sobre A Veiga que son as qu'agora amañan este volume. El ser humano, as súas edificacións, as súas ferramentas, a labranza, a pesca, el ferro, a madera, os nomes cos que se describe el tempo, os animales, a comida, as festas... todo cuntao dende A Veiga y con un arte y úa precisión que namáis dan os anos de trabayu y dedicación qu'el autor lleva ás costas.

O modelo das Normas da MDGA abéirase máis ao falado na zona B, a asturnegueirsa xa vista máis arriba (*no molín, el tou caldeiro taba chen de centén* etc), a fala eonaviega maioritaria demolingüística e xeograficamente. En troques, o modelo da Proposta do Principado achégase ao falado na zona C, a

baixonaviega, a que presenta algúns e moi sobranceiros trazos lingüísticos en común co asturiano central (*illustraciois, lliteratura, lleva, llibros*) pero amais con morfoloxía e ortografía propias do asturiano: *y, trabayo, veyá, resultao, qu'agora, d'anos, con úas* etc.

Co revestimento gráfico desta *Proposta*, non oficial pero si oficiosa, desde a Oficina de Política Llingüística e desde a Xunta de Toponimia del Principau d'Asturies estanse a fixar as formas dos topónimos da zona galegófono. De maneira que hoxe nesta zona poden coexistir formas populares inalteradas, formas deturpadas pola presión do castelán e formas deturpadas pola presión do asturiano. Vexamos, por exemplo:

	Forma oficial (aínda)	Forma correcta e popular
Deturpación polo castelán	<i>Carbajal</i> (Navia) <i>Carbajales</i> (A Veiga)	Carballal Carballales
Deturpación polo asturiano	<i>Carbayal</i> (Castropol) <i>Carbayal</i> (Eilao)	Carballal Carballal
Formas populares conservadas	<i>Carballofalso</i> (Ibias) <i>Carballo del Cuito</i> (Ibias)	Carballo Falso Carballo del Cuito

3.4. A Xunta de Toponimia del Principado de Asturias

A Xunta de Toponimia, equivalente á Comisión de Toponimia da Xunta de Galicia, foi creada en 1985 no seo da Consejería de Cultura, pero non é ata o ano 2002 en que se regulan as súas funcións e protocolos de recuperación e fixación toponímicas. Está formada pola presidenta da ALLA, dous expertos en filoloxía asturiana, un representante do RIDEA (Real Instituto de Estudios Asturianos), un representante da universidade asturiana e dous cargos políticos, dos cales un é o responsable da política lingüística asturiana. A ALLA abandonou a Xunta de Toponimia por considerar que o traballo desta xa o realizara e publicara a Academia de la Llingua Asturiana en 2000: *Nomes de conceyos, parroquies, pueblos y lugares del Principau d'Asturies*.

Non hai nesta Xunta de Toponimia ningún experto en toponimia galego-portuguesa nin en toponimia propia do Eo-Navia. Este organismo oficial asturiano xa fixou a toponimia de varios concellos, entre eles o eonaviego de El Franco (2005), e nestes momentos traballa na fixación dos topónimos doutros concellos eonaviegos (Grandas de Salime e Ibias entre outros).

Debido á adopción das *Propostas* asturianizantes nunca oficializadas, A Xunta de Toponimia está consagrando oficialmente toda unha mouteira de asturianismos (e vulgarismos) gráficos na fixación de topónimos eonaviegos. Vexamos algúns exemplos:

- a) Emprego do grafema *y* en lugar de *ll* ou *i*: *Caleya*, *El Caley*, *Cornayo*, *Payoza*, *Carbayal*, *Peruyeira*, *Busmayor*, *Conceyu d'Ayande*, *Conceyo d'Ibias*, *Caye A Fonte*, *Ayande* etc.
- b) Emprego do apóstrofo nos topónimos compostos: *A Veiga d'Eo*, *San Pedro d' Hío*, *San Pedro d'Agüeira*, *Vilanova d'Ozcos* etc.
- c) Eliminación case sistemática dos artigos -galegos-, en (Os) *Teixois*, (As) *Mestas*, (Os) *Pedrouzos*, (A) *Casía*, (O) *Castro*, (Os) *Boudois* etc., profusamente usados en textos turísticos, culturais e divulgativos.
- d) Usos de *b/v* á asturiana (ou á castelá) e non á galega: *San Cristobo*, *San Esteba*, *Viduedo*, *Berducedo* etc.
- e) Non se respecta o -d- intervocálico nas terminacións en -ado: *El Peneo*, *El Poblao*, *Montefurao* etc., e mesmo nunha ocasión conságrase unha hipercorrección: *Naraído* en lugar de *Naraío* (Santiso de Abres).
- f) Acentuación á asturiana: *Nadóu*, *Nogueiróu* etc.

- g) Incoherencias na haxiotoponimia: a Xunta e a ALLA recomiendan *San Tiso* pero *Santantolín de Ibias*; ou teiman en fundir nasais: *Samartín*, *Samamede* etc.
- h) Consagración de vulgarismos tipo *Castripol*, *Bual*, *Carbueira*, *Abilleira*, *Pilliceira*, etc.
- i) Asturianizacións forzadas como *San Lluís* ou *San Chuís* no canto do autóctono *San Llouguís*.

Toda a presión que recibe a poboación sobre a toponimia eonaviega fai que, por exemplo, no lugar chamado *As Mestas*, concello de Coaña, uns propietarios dun restaurante-sidraría denominen o seu local como *Les Mestes* (así, en asturiano central), nun intento de hiperasturianización artificial.

Esta “normalización” asturianizante é máis ficticia ca real, porque a inmensa maioría da rotulación viaria, documentación etc., segue aparecendo coa toponimia castelanizada. Sen embargo, entre os simpatizantes das linguas emerxentes estas “normalizacións asturianizantes” poden prender por simpatía. E poño exemplos: en libros editados en galego⁴ nos últimos dez anos achamos *Chao Samartín*, *Tixeu*, *Picu da Mina* (en Boussoño, San Martín de Ozcos) ou *San Lluís*, en lugar dos autóctonos *Chao de San Martín*, *Teixedo*, *Pico da Mina* ou *San Llouguís*, aínda que o máis triste é que os libros galegos (e os medios de comunicación galegos) seguen consagrando as formas deturpadas polo castelán nos territorios exteriores de lingua galega: *Villanueva de Ozcos*, *Vega de Espinareda*, *Chano*, *Pereje*, *Vegadeo* etc. É imprescindible que na Galicia administrativa se coñeza e se utilice a toponimia verdadeira das zonas galegófonas asturiana, leonesa, zamorana ou estremeña, pois é unha maneira de lles axudarmos a normalizala.

⁴ Por exemplo: Carballo Arceo, X. (2006): *Arqueoloxía de Galicia. Itinerarios polo pasado*, Vigo: Nigratea, ou Calo Lourido, F. (1997): *A cultura castrexa*, Vigo: A Nosa Terra.

Na *Enciclopedia gallega universal* da editorial Ir Indo os topónimos galegos de Asturias, Castela-León e Estremadura aparecen coa súa forma acastrapada. A casa editora non quixo poñelos na forma autóctona galega, e deste xeito contribuír ao coñecemento e difusión da forma verdadeira entre os galegos, argüíndo a razón de que as formas “normalizadas” non eran oficiais neses territorios. Por esta mesma razón moitos de nós non deberíamos ter empregado hai xa bastantes anos topónimos como Sanxenxo, Ourense, A Coruña,... e mais si o fixemos. Non teñen razón os que esgrimen “falta de oficialidade ou legalidade toponímica”, sobre todo cando esa dobre oficialidade vén recollida na *Carta Europea das Linguas Rexionais e Minoritarias*, art. 10.2.g), documento que desde 2002 é parte integrante do bloque de constitucionalidade da lexislación española actual.

Na meirande parte dos casos faise isto porque a fronteira política está instalada nos cerebros de moitos galegos. Dous exemplos máis: os mapas dos Ancares e do Courel de Edicións do Cumio (2002) aparecen coa toponimia galega do lado de alá da raia administrativa aínda en castrapo. Ou mesmo nos mapas que aparecen nas axendas anuais de editoras e institucións galegas a toponimia galega exterior segue aparecendo unha e outra vez deturpada.

Nos medios de comunicación o panorama é tamén desolador: alteracións e castrapizacións por todas partes, mesmo nos medios en galego, onde se continúa coa inercia denominativa en castrapo para os topónimos galegos extra-autonómicos: na TVG falábase non hai moito dun derrubamento en *Pereje* (Perexe) ou dun accidente en *Santa Eulalia de Oscos* (Santalla de Ozcos); n’*A Nosa Terra* liamos hai pouco o relato dun ex-guerrilleiro que andara agochado no lago de *La Campañana* (Campañá) e no concello de *Borrenes* (Borrés), cando estes mesmos medios non teñen reparo ningún en difundir galeguizacións infrecuentes de topónimos como *Baraxas*, *Xetafe*, *Guadalaxara*, *Badaxoz* (en galego antigo *Badallouze*) ou *Aranxuez*.

3.5. A castrapización que non cesa

A auténtica deturpación toponímica da Asturias de lingua galega é, como podemos supor, a derivada da castelanización histórica. Os procesos de deformación toponímica son exactamente os mesmos que operaron en territorio galego administrativo, coa diferenza de que nestes territorios exteriores, por inexistencia de lexislación protectora (ou por descoñecemento desta) e de vontade política, persisten de maneira esmagadora as castrapizacións:

- a) Redución dos ditongos decrecentes⁵ ou, *ei*: *Vega, Requejo, La Veguiña, Omente, Morentán, Uría, Ortiguera, Grandera, Figueras, Piñera*, etc.
- b) Adaptación fonética sistemática do fonema fricativo palatal xordo ao fricativo velar xordo do castelán: *Fojo, Seijas, Teijueira, Coruja, Vinjoy, Vijande, Teijois, Lijou, Villajane, Goje, Parajúa, Jarrio, Cereijeira, Gestoso, Arandojo, Gío, Freijulfe*,...
- c) Adaptación/tradución case sistemática de terminacións
 - -ana en lugar de -á: *Marcellana, Quintana, Mezana, Solana*,...
 - -ones en lugar de -ois: *Mendones, Peirones*,...
 - -ina(s) por -ía(s): *Cortinas, Serandinas, Figuerina, Vallina*,...
- d) Traducións parciais de topónimos compostos: *Chousavieja, Villarnovo, Villalaín, Puente del Porto, Valdebueyes* etc.

⁵ Non sempre, pois acaron de formas deturpadas aparecen formas populares inalteradas: *Rubieira, A Vruiteira, Perulleira, Piñeira*, etc.

- e) Traducións totais de topónimos transparentes: *La Laguna*, *Sierra*, *Cabañas*, *Peña del Cuervo*, *Puentenueva*, *La Brañuela*, *El Llano*..
- f) Falsas segmentacións cos artigos: *Fuente de Louteiro* (Fonte del Outeiro), *La Zoreira* (A Azoreira), *La Viouga* (A Aviouga), *La Mabona* (Llamabúa)⁶
- g) Alteracións debidas a defecto de audición: *Taladrid* (Taladriz), *Cadanosa* (Candaosa), *Cedemonio* (Zaramoño), etc.
- h) Cambios inexplicables: *Viavélez* (El Porto)⁷.

Toda esta casuística xa nos resulta moi familiar en Galicia. Nosoutros xa vimos e vivimos esa película e estamos no camiño de superala despois de case cinco lustros de Lei de Normalización.

As actuacións para normalizar a toponimia galega en Asturias e póla na súa forma tradicional, como di a lexislación asturiana e recolle a lexilación europea, abalan entre a actitude irrespetuosa, vulgarizante e asturianizante das propostas de fixación da ALLA e as decisións máis respectuosas coa lingua mais con grafía asturianizante da Xunta de Toponimia do Principado. En calquera caso, as autoridades académicas asturianas practican un autismo científico nada beneficioso para o galego eonaviego.

3.6. Danos colaterais: xentilicios e apelidos

Os medios de comunicación asturianos, que tan poderosamente contribuíron á deturpación toponímica desde o

⁶ O que nos fai acordantes do topónimo ancarés *Lama Ancha* e mais do limiao *Lama Má*, oficialmente durante moitos anos *La Mancha* e *La Mamá*.

⁷ Localidade coñecida por ser o lugar de nacemento da escritora Corín Tellado, quen se reclamaba publicamente natural de Viavélez e nunca de El Porto.

castelán, publicitan agora moi timidamente as “novas formas autóctonas” en versión asturianizante. E coas novas formas dos topónimos novos derivados destes: os xentilicios, algúns autóctonos, verdadeiros, pero moitos outros “inventados” sen criterio nas redaccións.

Nos últimos vinte anos os xornalistas espallaron a moda de que cada topónimo debería ter o seu xentilicio, cando nin o 10 % dos topónimos teñen xentilicio correspondente e amais moitos deles son derivados de alcumes ou nomes festivos⁸. E dado que toda lingua ten as regras de derivación de seu, os filólogos coñecen maiormente os étimos dos topónimos e poderían “crear ou inventar” xentilicios con certo rigor. Nas redaccións da Galicia administrativa e fóra dela invéntanse xentilicios que coa difusión e reiteración diaria acaban triunfando entre as persoas que os len/escoitan; referímonos a despropósitos como: *buenenses* (de Bueu), *forcaricenses* (de Forcarei), *niranienses* (de Nigrán), *lorentinos* (de Vilanova de Lourenzá), *agolenses* (de Agolada), *porteños* (de Porto⁹ das Portelas) etc.

O Eo-Navia non escapa a esta moda, pero cunha diferenza: os xentilicios populares existentes –de fasquía galega– son normalmente substituídos por outros inventados con intención cultista ou castelanista. Así, por exemplo, os de Santiso de Abres e Santalla de Ozcós xa non son máis *santiseiros* ou *santalleiros*, agora son *santirseños* e *santallese*, e nacen xentilicios como *veigueño*, *castropoleño*, *ozqueño*...por obra e graza da prensa escrita. A deturpación ou deformación dos xentilicios populares, así como a invención aberrante de xentilicios sen atender ao xenio e á tradición da lingua deberían ser tamén

⁸ Un exemplo: en todo o concello de Sanxenxo só existen dous xentilicios: *lilainos*, os de Padriñán e Sanxenxo vila, e *ghatos* os de Portonovo.

⁹ Dos catro concellos galegófonos de Zamora, só Porto ten xentilicio: *portexo*, para nada *porteño*, que corresponde ao Río da Prata (Arxentina). Para un concello portelao que ten xentilicio vai a prensa zamorana e múdallo “a la castellana”.

obxecto de estudo, denuncia e análise por parte do mundo académico en paralelo co estudo da toponimia¹⁰.

Un derradeiro tema que non quero deixar de tocar é o referente á onomástica de orixe toponímica. Se o topónimo está acastrapado é case seguro que acastrapado estará o apelido ao que deu lugar: *Vijande*, *Villanueva*, *Otero*, *Vinjoy* etc. En Galicia é posible agora "regaleguizar" *Seijas* en *Seixas*, *Vijande* en *Vixande*, *Otero* en *Outeiro*, *Villanueva* en *Vilanova* etc., pero en Asturias, aínda que en teoría nada o impide, na práctica resulta totalmente imposible. Deste xeito, por exemplo, a mesma vicepresidenta do goberno español D.^a M.^a Teresa Fernández de la Vega, oriúnda das ribeiras do Eo, aínda que quixera, teríao ben complicado para retomar o primitivo e auténtico apelido da súa familia materna: *da Veiga* (e non *de la Vega*), tal e como aparece aínda nalgúns documentos a principios do século XIX.

4. Cabo

Quería rematada esta chamada de atención sobre a situación do galego no Eo-Navia, do que a toponimia é unha mostra ben visible, solicitándolles publicamente á Real Academia Galega, ao Consello da Cultura Galega e ao novo Goberno galego que poñan os medios para que se realicen estudos serios da toponimia das áreas lingüísticamente galegas fronteirizas e se promocionen desde os medios de comunicación e as editoriais galegas as formas toponímicas (e xentílicas) auténticas e autóctonas.

Polo de agora, se alguén quere ver un exemplo de mapas do galego exterior coa toponimia totalmente normalizada,

¹⁰ Nas páxinas 95 e 96 do libro de Rodríguez Rodríguez, A. e Davila Ventura, M. (2004): *Lingua galega. Dúbdas lingüísticas*. Vigo: Universidade de Vigo, podemos achar unha relación bastante completa e acertada dos topónimos e xentilicios dos concellos estremeiros galegos. No documento que publicará proximamente Carlos Xesús Varela Aenlle sobre a toponimia eonaviega (2009) figura ao final un apéndice completísimo de xentilicios eonaviegos.

recoméndolle que visite a páxina web www.cousas.tk¹¹ de Ramón Pais Pazos, o autor do primeiro mapa dos *Territorios de Lingua Galega* ou primeiro mapa completo da Galicia lingüística e cultural. Que prenda o exemplo.

Bibliografía

- Academia de la Llingua Asturiana (2000): *Nomes de conceyos, parroquies, pueblos y llugares del Principau d'Asturies*. Uviéu: ALLA.
- Academia de la Llingua Asturiana (2006): *Informe sobre la fala o gallego-asturiano. Una perspectiva hestórica, social y llingüística*. Uviéu: ALLA.
- Asociación Abertal del Eo-Navia (2007): *Informe lingüístico sobre a nosa terra Eo-Navia*. Navia: SAEGAO.
- Asociación Abertal del Eo-Navia (2009): *Orientaciois prá escrita del galego de Asturias*. Vigo: ANL da Universidade de Vigo [edición en CD, no prelo]
- Babarro González, X. (2003): *Galego de Asturias: delimitación, caracterización e situación sociolingüística*. A Coruña: Fundación Barrié de la Maza.
- Cano González, Ana María (1980): "Los distintos bables de la región asturiana", en *Actas de la I Asamblea Regional del Bable*. Madrid: Editorial Nacional, 39-67
- Cano González, Ana María (1987): *Informe so la llingua asturiana / Rapport sur la Langue asturienne*. Uviéu: ALLA.
- Cano González, Ana María (1992): "Asturiano: evolución lingüística interna", en Holtus, G., Metzelsin, M., e Schmitt, Ch. (eds): *Lexikon der Romanistischen Linguistik*, vol. VI, 1, Tübingen: Max Niemeyer, 652-680.
- Comisión de Toponimia Asturiana (1985): *Conceyos y parroquies d'Asturies*. Uviéu: Conseyería de Educación, Cultura y Deportes.

¹¹ http://usuarios.lycos.es/aureano/galiza_irredenta.htm. e
...../[galego_caceres.htm](http://usuarios.lycos.es/aureano/galiza_irredenta.htm)

- Costas González, Xosé Henrique (2005): "Notas sobre o informe do Comité de expertos do Consello de Europa: 'A aplicación da Carta Europea das Linguas Rexionais ou Minoritarias en España'", *A Trabe de Ouro* 63, 417-428.
- Costas González, Xosé Henrique (2007): "Análise sobre a situación do galego no Eo-Navia, O Bierzo, As Portelas e o Val do Ellas", *Longalingua* 11, 14-25.
- Fernández Braña, B. et alii (1990): *Normas ortográficas e morfolóxicas del galego de Asturias*. Eilao: MDGA.
- Varela Aenlle, C. X. (2000): *Toponimia del Eo-Navia*. Eilao: MDGA.
- Varela Aenlle, C. X. (2009): "Repertorio toponímico da nosa terra Eo-Navia", documento en formato pdf incluído no CD *Lingua e literatura galegas de Asturias*. Vigo: ANL da Universidade de Vigo [no prelo].
- VVAA (1992): *Geografía de Asturias*, vol. II. Oviedo: La Nueva España.

OLLA E MIRA, DOUS MARCADORES DISCURSIVOS EN TRES LINGUAS: PORTUGUÉS, GALEGO E ESPAÑOL

SORAYA DOMÍNGUEZ PORTELA
CPI Doutor Daniel Monje, Navia de Suarna

RESUMO

Neste traballo preséntase unha descrición do funcionamento dos marcadores discursivos (MD) *olla / olha* en galego e portugués e *mira* en galego e español. Polo tanto, este é un traballo encadrado na liña de investigación da análise do discurso, combinada coa lingüística contrastiva. Nestas páxinas trátanse de maneira principal a cuestión de por que certas formas dos verbos de percepción adoitan desenvolver unha función discursiva, o funcionamento e os valores dos MD *olla / olha* e *mira* nun corpus exemplos das tres linguas, e mais os contrastes existentes entre elas. De maneira máis específica, analizaranse as diferenzas entre *olla* e *mira* en galego e as súas causas.

PALABRAS CHAVE

marcador discursivo, verbos de percepción, contacto de linguas, *olla/olha*, *mira*

"Look and stare, two discursive scoreboards in three languages: Portuguese, Galician and Spanish.

It is presented a description of the discursive scoreboards operational process (MD) *olla/olha* in Galician and Portuguese, and *mira* in Galician and Spanish. Therefore, this work is placed in the research line of the speech analysis, combined with contrastive linguistics. In these pages is mainly treated the question why certain perception verb forms use to develop a discursive function, the operational process and the values of MD *olla/olha* and *mira*, all that situated in an exemplified corpus

of three languages, and even more the existing contrasts among them. In a more specific way, the differences between Galician *olla* and *mira* and its causes are going to be analyzed.

KEY WORDS

discursive scoreboard, perception verbs, contact languages, *olla/olha, mira*.

1. Introducción

A lingüística contrastiva configúrase como unha ferramenta rendible para o coñecemento dun feito de lingua concreto ou un código particular, xa que a través dunha análise plural se pode acceder a un coñecemento máis profundo dun feito singular. Esa é a razón pola que á revisión dos marcadores discursivos (dende agora MD) galegos *olla* e *mira* lle adxuntamos o tratamento destas mesmas partículas noutros dous idiomas da mesma familia románica: o portugués e o español.

A razón da escolla destas dúas linguas encóntrase, fundamentalmente, en dous factores: o espazo e a historia. No que respecta ó primeiro, mentres o galego linda co portugués polo sur, co español comparte territorio administrativo, ó seren as dúas linguas cooficiais de Galicia. Xunto das dimensións locativas, as cales tamén xustificarían a escolla de, por exemplo, o asturiano, as razóns históricas argumentan a selección destes idiomas. En primeiro lugar, o galego e o portugués parten dunha mesma variante dialectal do latín, a cal, á súa vez, se resolve en dous códigos diferentes a partir da anterior escisión política. No que se refire ó español, a convivencia, de natureza conflitiva en diferente grao segundo o momento histórico, destes dous idiomas nun mesmo territorio político produciu unha serie de interferencias lingüísticas que interrelacionan estreitamente estes dous idiomas na nosa comunidade. Polo tanto, o portugués e o español non só son idiomas veciños para o galego, senón que manteñen con este unha relación suma-

mente estreita dende, practicamente, o comezo da historia dos tres.

Presentado o contexto idiomático deste traballo, réstanos delimitar a segunda parte do título do traballo, é dicir, definir o concepto MD. Dende os primeiros tempos da investigación lingüística recoñeceuse a existencia dun tipo de partículas que contaban cun alto uso, pero que non desenvolvían un rol facilmente delimitable na oración. A ambigüidade do seu funcionamento fixo que en moitas ocasións fosen consideradas dunha forma allea á gramática e metidas en caixóns estanco con diferente nomenclatura: *apoios léxicos*, *complementos externos*, etc. Dende os anos 60 do século pasado, co desenvolvemento da Análise do Discurso, estes elementos convertéronse, por primeira vez, en obxecto nuclear de estudo. O feito de que adquiriran este status non significa que se avanzase moito no seu coñecemento. A súa pluralidade semántica e funcional fai que, na actualidade, dar unha definición pechada do que é un MD non sexa posible. O conseguido ata o de agora é a delimitación dunha serie de propiedades que, en maior ou menor medida, todos os MD cumpren.

- Son palabras que non funcionan a nivel da oración, senón no discurso, xa que o seu labor é organizalo ou modalizalo ou relacionar un enunciado co conxunto do acto comunicativo.
- Perden as características que lle son inherentes á categoría gramatical da que derivan. O proceso de gramaticalización non é idéntico para todos os MD, de modo que, por exemplo, algúns que proveñen de verbos poden manter a rexencia de suxeito, aínda que sexa dunha forma máis ou menos fosilizada.
- Aparecen illados da liña entoativa.
- Non poden negarse.
- Teñen a capacidade de combinarse con outros MD.
- O funcionamento e valor dos MD son escollas propias de cada lingua, establecéndose estas partículas como un dos elementos máis definitorios dun idioma e o seu

dominio como un dos niveis finais para contar cunha competencia alta na lingua que se estea a aprender.

A delimitación dunha listaxe pechada de MD para unha lingua, sexa cal for, é unha tarefa aínda en curso. Deste modo enténdese que o habitual sexa o traballo sobre un número limitado de formas. Ese é o noso caso, ó circunscribirmos a dous MD: *olla* e *mira*.

2. *Dos verbos de percepción sensorial ós MD conversacionais*

Os MD non se definen como unha nova categoría gramatical. Son palabras ou xiros frásticos de moi diversa orixe que en determinados contextos pasan a funcionar como partículas discursivas e perden as propiedades que lle son propias á categoría da que derivan. Estes elementos teñen a súa orixe en adverbios, conxuncións, verbos, etc.

Ti vés e pidesme un favor qu'eu ch'otorgo; **pois ben**, é raro atopar quen esté disposto a lembrarse e a corresponder a el (*A Fuliada*, 1888, 2).

Oe, Pedro, nas cousas de amores non che valen razóns (Pla Xubiri, 1917, 64).

Para moitos MD o punto de partida é un lexema verbal, singularmente se describe unha percepción sensorial de natureza visual ou auditiva. O salto funcional (verbo > MD) xustifícase polo significado de partida destas unidades, cun posterior proceso progresivo de desgaste semántico. Os verbos sensoriais do tipo *oír* ou *ollar*, en primeiro lugar, describen a acción de percibir algo ou alguén co oído ou coa vista outro algo ou alguén; nun segundo nivel, o/a emisor/a procuran que o/a receptor/a oia ou vexa algo ou alguén; e, por último, xa non é preciso que se faga referencia a esas accións, senón que o único que se busca é captar a atención do que está a interrelacionar co que emite a mensaxe. Isto é, a finalidade do emisor é chamar a

atención do receptor sobre un asunto ou feito, sen que haxa unha visión ou audición dun feito concreto. En definitiva, alteran o seu punto de referencia do campo físico (visual ou auditivo) para o da acción mental, ó contextualizar situacións subxectivas, derivando deste un sentido pragmático centrado no texto do falante.

Polo tanto, enténdese que este sexa un dos recursos máis rendibles para a creación de MD dirixidos cara ó receptor. Ademais, ó compartiren orixe gramatical e semántica (verbos de percepción sensorial), enténdese que, fóra de manter as características propias para o conxunto de MD, xurdan outros puntos comúns singulares a este subtipo. Estas pódense reducir a dous (Gallardo Paúls e Marín Jordà, 2005, 60):

- Derivan de formas imperativas ou relacionadas co imperativo.
- Teñen un carácter connotativo básico que os vencella co receptor.

O feito de que estes MD sexan elementos cos que se interactúa co oínte do acto comunicativo xustifica que se clasifiquen como MD conversacionais.

Ó igual que non existe unha definición pechada de MD, tampouco hai unha clasificación única dos seus subtipos. As funcións destes elementos son de moi variada natureza: poden enfatizar a información, xerarquizala, organizala... Os aspectos que se poden considerar para a súa clasificación son moi numerosos, pero aínda así algunhas etiquetas van triunfando ó seren recorrentes nos diferentes estudos sobre estas partículas. Unha delas é o de MD conversacional para o que funciona a nivel da conversación. Son elementos que organizan a fala en quendas e pasan a caracterizar o acto comunicativo como un modelo simétrico -cando ambos interlocutores teñen o mesmo dereito non só de tomar a palabra, senón tamén de escoller o tópico discursivo, direccionándoo de acordo co tempo de participación- ou asimétrico. Isto é, o uso dos MD conversacionais restrínxense a este tipo de discurso, sexa

convencional ou un simple recurso do código escrito para facer partícipe o receptor do proceso comunicativo, aínda que non teña posibilidades reais de actuar no suposto diálogo.

3. *Mira e olla, en galego, portugués e español*

Mira e olla son dous MD derivados de verbos de percepción visual cunha distribución diferente nas tres linguas peninsulares consideradas neste traballo. Mentres no idioma portugués só *olha* se desenvolve como unha figura do discursivo, para o español esta non é unha opción, senón que con este papel aparece exclusivamente *mira*. Pola contra, o galego sitúase como ponte entre os dous idiomas veciños: as dúas partículas son MD galegos, aínda que non teñen a mesma importancia de uso. *Mira* é máis recorrente có seu compañeiro.

Para observar o comportamento e os valores que lle engaden estes elementos ó texto, temos en conta dous tipos de fontes: por un lado, un corpus de rexistros reais elaborado para esta investigación e, polo outro, facemos uso de traballos doutros autores sobre estas partículas. O peso de cada unha destas fontes será diferente segundo o idioma en cuestión.

Na lingua española o funcionamento de *mira* está amplamente estudado, polo que nos faremos eco dos resultados de diferentes publicacións sobre o tema, destacando os artigos de Martín Zorraquino e Portolés (1999) ou Gallardo Paúls e Marín Jordà (2005).

Para o portugués entran en xogo os dous tipos de vías: a bibliográfica, con traballos como o de Rost (2005), e tamén a de corpus. Neste caso elaboramos unha base de datos extraída do *Corpus Lingüístico da Universidade de Lisboa*; esta base de datos está formada, á súa vez, por diferentes corpora. Para o tratamento de *olha*, un MD conversacional, tomamos en consideración exclusivamente o *Corpus de Português Oral*, xa que nos interesaban os discursos conversacionais onde os dous interlocutores estivesen presentes e, polo tanto, onde este tipo de partículas de interacción entre eles son máis comúns. O

Corpus de Português Oral está configurado a partir de conversas recollidas na obra *Português Fundamental* de María Fernanda Bacelar do Nascimento (1987) elaborada para o proxecto *Português Fundamental*. De aquí extraemos un total de 67 rexistros onde *olla* perde o seu rol verbal e funciona como un MD.

No caso do galego, a investigación circunscríbese exclusivamente a un traballo de corpus. Este tipo de partículas na lingua galega comezaron a ser estudadas na última década e fíxose ese tratamento ou dende unha perspectiva moi xeralista (Freixeiro Mato 2005) ou circunscribíndose a un número restrinxido de marcadores entre os que non se inclúe *mira* (Domínguez Portela 2006). A fonte para a elaboración do noso corpus é o *Tesouro Informático da Lingua Galega* (TILG) formado por máis de 11 millóns de rexistros de palabras que se engloban nun total de 90 mil lemas extraídos de 1500 fontes, creado polo Instituto da Lingua Galega. De aquí, atendendo só obras datadas a partir de 1970 -para procurar o maior equilibrio posible coas épocas de estudo contemporáneas das fontes portuguesas e españolas-, obtivemos un total de 143 rexistros en que *olla* é usado como un MD e 2011 de *mira*.

Na busca dunha definición do funcionamento destes dous MD, delimitamos unha serie de aspectos que xurdiron como fundamentais para esta finalidade:

- o estilo de lingua: observar se son partículas propias dun estilo concreto ou se non hai restricións a este nivel
- as súas posibilidades combinatorias con outros MD, xa que esta non é unha capacidade xeral para o conxunto de MMDD
- a súa posición dentro da oración: se é fixa ou permite mobilidade
- a súa colocación dentro da liña entoativa do discurso, e en particular se esixe unha linde prosódica.
- a súa complementación, é dicir, se ten capacidade de rexeer complementos herdados da súa orixe verbal ou se perdeu esta propiedade

- o seu valor dentro da oración ou do discurso, a súa finalidade discursiva
- a súa incapacidade de ser negado, en liña co común para o conxunto dos MD

Moitos destes aspectos axudarannos a determinar se estamos ante partículas cun funcionamento totalmente discursivo, ou se aínda son elementos en proceso de cambio gramatical ó preservar parte do funcionamento propio ó seu pasado verbal.

4. *Olha e olla*

4.1. *Estilo de lingua*

Tanto en portugués coma en galego, defínese como un MD conversacional por axudar a entender como ocorre o diálogo. Mais non é común en todo tipo de conversas, senón que recollemos un condicionante constante nos dous idiomas: o estilo de lingua coloquial. En portugués, dos 67 rexistros totais do corpus de estudo, só 1 se podería adscribir a un estilo de lingua formal;

Como é um povo e como é necessariamente uma política, não é. A. Pois, e olha, por exemplo em relação ao papel fundamental que a televisão tem nos meios sociais (Bacelar do Nascimento, 1987, 306).

e no galego, só 5 rexistros dos 143 totais.

Luar frío cobra noite son en todo augas femias; olla que a metalurxia é sustento do mundo (Méndez Ferrín, 1994, 33).

O léxico, o tema tratado e o posicionamento do emisor a respecto do receptor sitúannos nun ámbito formal, pero é xustamente o MD *olla* o que suaviza a formalidade da

intervención. Polo tanto, *olla*, nas dúas linguas, adscíbese ó grupo dos MD conversacionais propios da coloquialidade.

lingua	estilo de lingua	
	coloquial	formal
portugués	66	1
galego	138	5

4.2. Valores do MD

No plano do significado, a escolla deste MD responde á necesidade de captar a atención, basicamente en dous niveis: sobre o receptor ou sobre a información. Á súa vez, en cada un deles *olla* / *olha* adquire diversos subvalores.

Cando a finalidade de *olla* / *olha* é conseguir a atención do receptor, xorde porque se quere asegurar a comprensión exitosa da mensaxe. Nestes casos o emisor vai:

- comezar unha nova intervención orixinal para informar sobre algo que considera relevante:

Olla, cretino Amado: rematou a Asamblea (Méndez Ferrín, 1980, 28).

Mana, olha, sabes, que afinal a moreia que eu comprei não prestava para nada (Bacelar do Nascimento, 1987, 131).

É bastante habitual que o introducido sexa unha pregunta,

E olla unha cousa, ¿ese Don Erno teu castigáte tamén cando bafordas mal? (Cabana, 1996, 73).

Olha: rapazes, sabe por onde se nota? Pelo peito! (Bacelar do Nascimento, 1987, 92).

unha advertencia, consello ou aviso.

Olla , Garela, millor será que cales e que non me fagas falar (Blanco Amor, 1974, 27).

Olha, já me vou, já ando desde sábado a fazer viagem (Bacelar do Nascimento, 1987, 244).

- responder unha pregunta dun outro:

¿Que por qué falo deste xeito... cando pertence á gramática non a un libro de «pasá-lo tempo»? Pois olla, fun a Madrí onde fan os presidentes que as sobriñas e os parentes, sen poñer nada de si, poidan amosa-los dentes (Barxa Iglesias, 1984, 11).

O que é que tu fizeste? X: Olha, fomos 'a... sabes quem é o NP? A: Sim. X: Fomos à... temos ido agora à praia. (Bacelar do Nascimento, 1987, 102).

- ou buscar unha actuación consecutiva no receptor, xa que domina a función apelativa no discurso:

¡Non, co fol non que espallas a borralla... olla como espallas a cinza que voa emporcallando todo! (Díaz Fernández, 1985, 87).

Olha, espera lá, olha agora vai começar a feitiçeira (Bacelar do Nascimento, 1987, 138).

Polo tanto, a súa función neste nivel achégase moito á dos organizadores do discurso, ou sexa, non só modaliza o que se di, senón que tamén o xerarquiza dentro da conversa.

Por outra banda, cando o focado é a información, a chamada de atención é sobre o dato en si mesmo. Pódese querer destacar por diferentes razóns, para:

- reafirmar a intensidade, grao ou importancia da mensaxe que se acaba de dar:

Olla o bonito que es (Ledo Andión, 1985, 95).

O meu menino, olha, desde que almoçei ainda não o vi mais (Bacelar do Nascimento, 1987, 244).

- transmitir a estupefacción ou sorpresa -negativa ou positiva, segundo o contexto- que produce unha información no receptor:

Olla o que di este télex (Cid Cabido, 1988, 125).

Mesmo os sonsinhos, sempre vão fazendo das suas, olha... B: Até andam com divorciadas... (Bacelar do Nascimento, 1987, 103).

- descargar de dramatismo ou suavizar a incomodidade que pode producirle o que se vai dicir ó emisor e/ou ó receptor, por a nova información se opor á opinión do outro ou por non ser esperable:

Olla , non sei. O Xacinto di que non fixeron ningún mal (Rodríguez Troncoso, 1996, 57).

C: Vamos, vamos embora! X: Pois, vê lá, tu é que sabes. Olha, eu não, não forço porque sei, eu sei que é (Bacelar do Nascimento, 1987, 305).

- focalizar a información por ser o dato final, a conclusión dunha exposición

Olla: tes razón cando dis que eu soio nascín pra traballar (Catoira, 1977, 51).

O amigo NP arregalou-se-lhe logo o olho, e estas duas coitaditas de perna fina, olha, foram logo desprazadas (Bacelar do Nascimento, 1987, 103).

- ou destacar a explicación ou argumentación para guiar a comprensión do receptor, xa que o emisor considera que o exposto non queda claro

¿Non se louvaban de portarme e seren meus parentes? E olla que a min pouco me estraña que non quixesen escoitar o que ti lles pedías (Cabana, 1996, 163).

Como é um povo e como é necessariamente uma política, não é. A: Pois, e olha, por exemplo em relação ao papel fundamental que a televisão tem nos meios pequenos como Mangualde, praticamente digamos como meio de televisão é o único (Bacelar do Nascimento, 1987, 306).

Con estas dúas últimas finalidades -introducir unha conclusión ou unha explicación- retoma novamente a súa función de organizador do discurso, xa que xerarquiza a importancia ou regula a orde dos elementos discursivos.

No referente á predisposición desta partícula en adquirir cada un destes valores e subvalores segundo a lingua, observamos un funcionamento, a grandes trazos, común. En primeiro lugar, tanto no galego coma no portugués, *olla/olha* ten as mesmas posibilidades de ser usado para facer unha chamada de atención sobre o receptor coma sobre a información.

Valor	subvalor	portugués	galego
chamada de atención sobre o receptor	total	45%	49,5%
	informar	10,5%	18%
	resposta	1,5%	3%
	preguntar	4,5%	0,5%
	advertir	21%	18%
	apelativo	7,5%	10%
chamada de atención sobre a información	total	55%	50,5%
	intensidade ou énfase	25%	24,5%
	estupefacción	3%	16%
	suavizar	6%	1,5%
	conclusión	10,5%	7%
	explicación	10,5%	1,5%

Ademais, dentro de cada grupo de sentidos, destacan os mesmos no galego ca na lingua lusa: se se procura captar a atención do receptor será para informalo ou advertilo e se o focado é a información propiamente dita será para reforzar a súa intensidade ou importancia ou porque se trata dunha conclusión.

Pola contra, encóntranse diferenzas na predisposición que cada idioma rexe a respecto dos restantes subvalores que pode adquirir este MD. Son todos eles secundarios, pero sitúanse

dunha forma diferente dentro da escala de normalidade de uso en cada un dos códigos lingüísticos. Deste modo, *olla* como marcador de sorpresa é máis habitual na lingua galega ca na veciña, mentres que na lusa recórrese en máis casos a *olhar* para focalizar a explicación que se vai introducir. Son, sempre, diverxencias porcentuais en valores non nucleares desta partícula. Polo tanto, podemos concluír que o funcionamento discursivo deste MD é moi similar entre o galego e o portugués.

4.3. Posición no discurso

Outro punto compartido polas dúas linguas encóntrase na posición que ocupa *olla* na oración. A situación maioritaria é a inicial, superando o 75% dos rexistros nos dous idiomas.

posición	tipos	portugués	galego
inicial		76%	80%
posposto a un elemento externo	nexo coordinante	9%	8%
	MD	6%	7%
	vocativo	3%	1%
posposto a un elemento oracional	tras suxeito	3%	
	tras verbo		1,5%
	tras complemento temporal	3%	0,5%
	tras complemento locativo		0,5%
final			1,5%

Cando non se coloca neste lugar, encontramos tres situacións diferentes:

- *precédese por un elemento externo:*

A maioría dos casos en que *olha*, en portugués, e *olla*, en galego, non ocupa a posición inicial na oración débese a que este lugar é cuberto por outro elemento alleo á sintaxe da estrutura, é dicir, anteponse un segundo elemento externo que concorre con este MD por situarse no mesmo lugar na construción. Antecedén a este MD ou un nexo coordinante:

Sei que non o es, se o queres así. Pois non faltaría máis. Pero olla qué ocasión tan boa para buscar un cambio de situación (Martínez Oca, 1983, 189).

É interesante a participación de bons orquestradores na canção portuguesa. A: Mas olha, não achas por exemplo que a canção portuguesa é ainda, quer dizer, ainda é muito incipiente (Bacelar do Nascimento, 1987, 276).

outro MD:

O sabor do cornecelo ten as roxas nos seus labios: pois olla que o cornecelo precúra-no os boticarios (Lorenzo Fernández, 1973, 124).

Olha, espera lá, olha agora vai começar a feiticeira (Bacelar do Nascimento, 1987, 138).

ou un vocativo:

Palomiña encantadora, olla que son cazador (Lorenzo Fernández, 1973, 128).

Mana, olha, sabes, que afinal a moreia que eu comprei não prestava para nada (Bacelar do Nascimento, 1987, 131).

- *precédese por un complemento oracional*

En menor medida pode posporse a complementos da oración que o MD está a modalizar. Son casos moi minoritarios e responden sempre a unha mesma finalidade comunicativa: realizar unha focalización do dato, sexa o xa dito (cando se pospón ó suxeito ou ó verbo):

E morreram, morreram, morreram, e os ciganos, coitados, pronto, olha, morreram (Bacelar do Nascimento, 1987, 279).

Pastorciños e magos de gran sabencia achéganse olla á túa beira! (Cao Martínez, 1983, 158).

ou sexa o que se vai dicir (ó se pospor á indicación temporal ou locativa):

No día seguinte, olha, tudo o que la pôs, foi tudo! (Bacelar do Nascimento, 1987, 270).

Diante túa, olla, un grande cadro de mandos (Méndez Ferrín, 1986, 116).

- *conclúe a oración*

Só na lingua galega se rexistrou unha última opción: a posición final na oración é ocupado por *olla*. Pero nestes casos, este elemento non funciona só, senón que se inclúe nun MD composto, que toma a forma de unidade frástica, aínda que fosilizada, independente. Á súa vez, o primeiro elemento desta unidade discursiva é *olla*.

Xa quedaba tranquilo, olla ti que cousa (Eiré, 2000, 69).

Polo tanto, o que semellaba unha diverxencia de funcionamento non é tal, xa que o que se sitúa no último lugar da estrutura é un MD oracional, no que *olla* ocupa a posición inicial, antecedendo incluso o suxeito. De modo que, tamén nestes casos, *olla* se caracteriza por esixir introducir a unidade discursiva.

Esta é a primeira discordancia de consideración que encontramos entre o galego e o portugués no que se refire ó funcionamento deste MD. Pero será a ponte que nos leve ó aspecto do seu comportamento máis afastado entre linguas: as capacidades combinatorias de *olha* e *olla*.

4.4. *Capacidades combinatorias*

A posibilidade de aparecer unha acumulación de MMDD é habitual; nalgúns casos só son acumulacións de partículas externas que modalizan ou organizan o discurso cada un dunha forma, mentres que noutros funcionan como elementos compostos que producen un único efecto sobre el.

Con *olla* e *olha*, as posibilidades de que este MD apareza xunto doutro marcador son diferentes segundo o idioma. Hai unha situación inversa: mentres en galego a tendencia é que *olla* se combine con outra partícula, en portugués *olha* adoita aparecer como un MD simple e sen acompañamento doutras unidades periféricas.

lingua	funcionamento do MD	
	illado	combinado
portugués	60%	40%
galego	36%	64%

Os elementos que poden acompañar o noso redúncese a tres tipos: outros MD ou complementos externos, o vocativo e complementación verbal herdada do pasado gramatical do marcador. Das dúas primeiras xurdiron exemplos en galego e en portugués, mentres que da última só apareceu na lingua galaica.

4.4.1. *Olha / olla xunto doutros MD ou complementos externos*

Nestes casos, hai dúas posibles situacións: o funcionamento destes elementos como unidades independentes acumuladas, ou, a partir da súa xuntanza, a configuración dun novo MD composto.

Son acumulacións de MMDD ou MD e complemento externo os encontros:

combinatoria	nº de rexistros	
	portugués	galego
<i>olla, olla</i>	-	5
<i>ala, olla (ti)</i>	-	1
<i>olha, (Vº,) sabes</i>	2	-
<i>olha, espera lá</i>	2	-
<i>pronto, olha</i>	1	-

Como se pode observar, o galego e o portugués non comparten tipos de xuntanzas discursivas. No comezo deste traballo referimos que os elementos que funcionan como MD e os seus valores é un trazo moi singularizador de cada idioma. A explicación para isto pode encontrarse en que dentro destas partículas podemos establecer dous grandes grupos: un primeiro onde se inclúen elementos que, por evolución semántica, acaban desenvolvendo este rol discursivo na maioría dos idiomas (por exemplo, os derivados dos verbos sensoriais), e un segundo moito máis dependente de modas ou estilos lingüísticos, que responden a procesos de creación singulares de cada comunidade de falantes. Boa parte dos elementos cos que se combina *olha / olla*, sen configurar unha unidade composta, pódense incluír no segundo subgrupo, o que xustifica o afastamento entre códigos.

a. Repetición de MD

É unha fórmula só recollida no corpus galego. O uso reiterado desta unidade funciona sempre dunha mesma maneira: procúrase captar a atención do interlocutor, habitualmente porque hai unha mudanza de tema. Entoativamente configuran unidades diferentes, o que mostra que non se desenvolven como un MD composto, senón como a acumulación reiterativa dun mesmo:

Claro que ti de raparigo non eras tan feo coma agora... Olla, olla ... Dá gusto a un pai ver durmir o seu neno... (Cabana, 1996, 111).

b. ala + olla

Tamén exclusivo da lingua galega, neste caso *olla* é precedido por outro elemento externo da oración. Non comparten en ningún caso liña entoativa e a súa acumulación é unha estratexia coa que o emisor busca reforzar a focalización da conclusión que se vai introducir, marcándoa como unha realidade sorprendente a través do uso de *ala*.

Alá, olla ti, era a auga a que faltaba (Martínez Oca, 1993, 14).

A partir de aquí, consideramos as combinatorias recollidas na lingua portuguesa.

c. olha (+ V) + sabes

Xunto de *olha sabes* aínda pode aparecer un terceiro elemento: un vocativo. *Sabes* sempre se coloca despois dos outros constituíntes e nunca comparte liña entoativa con ningún deles. Finalmente, no que respecta ó valor de *olha*, nestes casos funciona como unha voz de chamada cara ó receptor: búscase captar a súa atención porque vai ser informado ou advertido sobre unha realidade. Os outros dous elementos funcionan cara a mesma finalidade, a de captar a atención do destinatario.

Mana, olha, sabes, que afinal a moreia que eu comprei não prestava para nada (Bacelar do Nascimento, 1987, 131).

d. olha + espera lá

Espera lá é un MD de estrutura frástica que en dous rexistros segue a *olha*. Coma no caso anterior, tampouco comparten liña entoativa e a combinatoria de *olha, espera lá* tamén funciona como unha chamada de atención sobre o receptor; con estas estruturas búscase unha reacción no comportamento do interlocutor, especialmente a través do segundo marcador.

Olha, espera lá, olha agora vai comezar a feitiçeira (Bacelar do Nascimento, 1987, 138).

e. pronto + olha

Estes dous elementos tampouco comparten liña entoativa e, mentres *pronto* interrompe a enunciación, *olha* focaliza o dato que introduce marcándoo como a conclusión ou resultado do proceso comunicativo que se está a realizar.

E os ciganos, coitados, pronto, olha, morreram (Bacelar do Nascimento, 1987, 279).

Noutros casos, a xuntanza de varios elementos externos resólvese na configuración dun MD composto.

MD composto	nº de rexistros	
	portugués	galego
<i>olla / olha que</i>	10	40
<i>pois olla</i>	3	7
<i>olla como</i>	-	11
<i>olla (ti) por onde</i>	-	1
<i>olha lá</i>	2	-
<i>olha agora</i>	1	-

Fóra dos dous primeiros, que aparecen nas dúas linguas, tampouco coinciden o galego e o portugués nos MD compostos con *olha / olla*.

a. olha / olla que

Este é o encontro máis recorrente nos dous idiomas.

Bati à porta do quarto: "Olha que já são horas." (Bacelar do Nascimento, 1987, 141).

Si calaras, si calaras... Olla que eres comenenciado (Blanco Amor, 1974, 34).

Forman unha única unidade discursiva que, semanticamente, pode aparecer con dous valores.

- focaliza o dato, reforza a intensidade da calidade ou grao especificado

Oh! Olha que chumbada, do futebol! (Bacelar do Nascimento, 1987, 113).

Olla que fermosas arbres -exclamaba ela, de súpeto (Cortezón Álvarez, 1981, 53).

- chama a atención do receptor, porque é introducida unha advertencia.

Mas olha que em dereito em Lisboa com alguns profesores eu sei que era assim! (Bacelar do Nascimento, 1987, 181).

Disme que vou tan velliño que non podo cos calzós: olla que a troita i o unto canto máis vellos, millor (Lorenzo Fernández, 1973, 71).

O funcionamento conxunto dos dous elementos explica que nun número moi reducido de construcións -1 en portugués e 4 en galego- haxa un corte na liña entoativa entre estes elementos, xa que nestes casos funcionan como unidades diferentes:

Ora um bocadinho de terra pequenino, um bocadinho de cenoura... e olha, que nós conseguíamos menos, ali! (Bacelar do Nascimento, 1987, 243).

¡Olla, qué sorpresa, o fillo do Aser (Blanco Amor, 1972, 169).

b. pois olla / olha

Pois en todos os casos se antepón a *olla*, e os dous sempre comparten liña entoativa.

Pois olha, fica sabendo que fazes mal em não ler (Bacelar do Nascimento, 1987, 149).

¿Que por qué falo deste xeito... cando pertence á gramática non a un libro de «pasá-lo tempo»? Pois olla, fun a Madrí onde fan os presidentes que as sobriñas e os parentes, sen poñer nada de si, poidan amosa-los dentes (Barxa Iglesias, 1984, 11).

Só nun rexistro en portugués aparece entre eles o nexo coordinante *e*,

Pois, e olha, por exemplo em relação ao papel fundamental que a televisão tem nos meios de comunicação (Bacelar do Nascimento, 1987, 306).

pero entendemos que é unha variante dun mesmo MD, máis do que unha ruptura na liña entoativa entre elementos.

Este MD composto céntrase no receptor: pasa a ser unha chamada para que o interlocutor poña atención sobre o que se

vai dicir, sexa unha opinión, advertencia, argumentación ou resposta.

c. olla como

Totalmente paralelo a *olla que* encontramos o conxunto *olla como* recollido na lingua galega. As dúas palabras comparten en todo momento liña entoativa e funcionan como voz de chamada sobre a información para reforzar a intensidade, grao ou calidade do dito. Este é un caso, coma *olla que*, en que se pode observar o paso intermedio na evolución de verbo de percepción a MD, xa que a partícula conserva, en certo modo, a idea de captación a través dos ollos a pesar de que se utiliza basicamente como un elemento focalizador.

-Ven aquí, muller. Acouga. Olla como non hai ninguén (Martínez Oca, 1993, 141).

d. olla (ti) por onde

Configura unha unidade entoativa independente do resto da oración e funciona como un focalizar da conclusión por presentarse como un feito sorprendente. Este é un claro exemplo de marcador composto onde os complementos rexidos non funcionan a nivel oracional, senón como expresión fixa discursiva, xa que *por onde* non remite en ningún caso unha referencia locativa pola que dirixir os ollos.

Mágoa de seren casados, tanto un coma o outro, se non olla ti por onde podían facer unha parella de intelectuais (Martínez Oca, 1995, 76).

e. olha lá

Este e os seguintes son MMDD compostos exclusivos da lingua galega.

Olha lá forma unha unidade entoativa illada do resto da construción coa que o emisor procura a captación da atención do receptor porque vai comezar a súa intervención, normalmente pisando a do outro. Polo tanto, é unha marca na procura de recuperar ou pasar a posuír a quenda de fala.

Olha lá, há alguma diferença de técnica para saltar um cav(...) para saltar um obstáculo comprido e, ou um obstáculo alto? (Bacelar do Nascimento, 1987, 151).

f. olha agora

Na única construción do noso corpus portugués onde se fai uso deste MD, os dous elementos forman parten dunha mesma unidade entoativa, independente do resto do discurso. Utilízanse para dirixir a atención sobre o propio discurso marcándoo como un feito que provoca estupefacción no receptor.

Aquilo às vezes até doía, e tinham que ficar sem o cabel(...)
C: Olha agora! (Bacelar do Nascimento, 1987, 236).

En definitiva, xunto doutros MD ou complementos externos *olha* amosa dous tipos de comportamentos: os dous MD funcionan como unha única unidade ó compartir liña entoativa e finalidade na estrutura (é dicir, xunto de *que, como, pois, lá, por onde* ou *agora*) ou, pola contra, son diferentes MD acumulados que presentan valores non exactamente sinónimos ou sinxelamente diferentes (por exemplo, con *pronto* ou *ala*).

4.4.2. Olha / olla xunto dun vocativo

A última concorrencia deste marcador cun complemento externo á oración é xunto dun vocativo. A voz de chamada habitualmente posponse ó MD e estes dous elementos case nunca comparten liña entoativa.

Olha, pá, não é para vos estar a desiludir mas essa coisa de vendedores e de inquéritos dura um mês, vocês ganham seis contos num mês (Bacelar do Nascimento, 1987, 140).

Olla, Garela, millor será que cales e que non me fagas falar (Blanco Amor, 1974, 27).

A través do vocativo aparece explícita unha referencia ó receptor, o que reforza a idea de *olha / olla* captar a súa atención, podendo ter esta procura distintas finalidades, como xa vimos nos valores do MD.

4.4.3. Olla con complementación verbal propia

Finalmente, colocamos como unha última opción de combinación a posibilidade de *olla* aparecer complementado por outras unidades. En portugués, tal e como a negación deste marcador, é unha imposibilidade sintáctica. Pola contra, mentres o galego comparte a incapacidade de negar *olla*, diverxe da lingua lusa ó permitir acompañarse por un suposto suxeito (*ti*), obxecto directo (*olla unha cousa*) ou circunstancial de modo (*olla ben*).

combinatoria	nº de rexistros
	galego
olla ti	8
olla unha cousa	1
olla ben	6

A pesar de xurdir como unha novidade a respecto do recollido para o portugués, tampouco podemos extrapolar deste tipo de estruturas que *olla* no galego conserva, como MD, o seu funcionamento verbal ou parte deste. Estas construcións son restos fosilizados que se configuran como MD frásticos. Unha mostra do dito encontrámolo en dúas características do comportamento destas unidades compostas. Unha delas refírese en concreto ó MD *olla ti*.

Ei, Roxelio!, o Santiago, o do tostadeiro de café, olla ti quen vai aparecer por aquí agora (Martínez Oca, 1995, 77).

Sintacticamente, semella que estamos ante un verbo (*olla*) xunto do seu suxeito (*ti*), pero para esta consideración xorden dificultades dende a mesma disposición dos elementos. O combinado *olla ti* sempre segue esta orde, é dicir, o verbo antepónse sempre ó pronome persoal tónico; isto choca coa regra básica da colocación dos constituíntes oracionais na lingua

galega, onde a distribución primeira é sempre suxeito + verbo + complementos; en galego só se encontra esta disposición dos elementos cando o verbo está en modo imperativo e focalízase o suxeito. Neste caso o suxeito non é o foco de atención, senón que o encontro *olla ti* respecta a disposición dos elementos habitual para *olla* como unidade discursiva, pois nestes contextos sempre esixe a posición inicial no discurso.

En segundo lugar, outro trazo característico destas tres construcións (*olla ti*, *olla ben* ou *olla unha cousa*) é que en todos os casos que rexistramos os membros da unidade frástica fosilizada os elementos comparten liña entoativa. Un dos trazos definitorios dos MD compostos de *olha / olla* vimos que era este comportamento melódico que aquí se respecta tamén.

Boeno, dáme un bico, e olla ben que xa estás ceibe destas falcatuadas (Catoira, 1977, 8).

E olla unha cousa, ¿ese Don Erno teu castigaste tamén cando bafordas mal? (Cabana, 1996, 73).

Polo tanto, estamos ante MD compostos que derivan de estruturas frástica, pero xa fosilizadas.

No que se refire á finalidade comunicativa destas estruturas no discurso, esta é de dous tipos:

- funcionan como captadores da atención do receptor porque se vai introducir unha pregunta ou advertencia; este é o uso tanto de *olla ben* coma de *olla unha cousa*.

Olla ben o que fago con il (Blanco Amor, 1972, 37).

E olla unha cousa, ¿ese Don Erno teu castigaste tamén cando bafordas mal? (Cabana, 1996, 73).

- chaman a atención sobre a información en si, ó configurala como un feito que produce estupefacción, por ser unha conclusión ou por querer intensificar a súa importancia; con este uso escóllese a fórmula *olla ti*.

Pero dicir non dixo nada, olla ti, se cadra gustaballe (Eiré, 2000, 110).

En conclusión, non estamos ante un verbo cos seus complementos, senón ante MD frásticos, ó seren unidades oracionais fosilizadas que funcionan no nivel discursivo.

A través das opcións combinatorias deste MD fomos definindo a súa situación na liña entoativa. Vimos que, cando funciona como un MD composto, os formantes do construto deben compartir liña melódica; mentres, se temos unha acumulación de MD ou un MD xunto de diferentes elementos externos da oración, o habitual é que haxa algún corte entoativo entre as partículas. Partindo deste postulado, ó seren maiores na lingua galega as capacidades combinatorias (co resultado dunha acumulación de elementos ou o funcionamento único destes) de *olla* ca na lusa, fai supor que o comportamento deste MD na liña entoativa será máis plural no referido idioma.

posición	construción	portugués	galego
illado	MD simple	63%	39%
	MD composto	15%	10%
non illado	MD simple	10%	17%
	MD composto (<i>olla que / como</i>)	12%	34%

Así, encontrámonos con que o funcionamento xenérico nas dúas linguas é paralelo: sempre que *olla* ou *olha* non formen parte dun MD composto cos elementos *que* ou *como*, o esperable será que tras el haxa unha ruptura na liña entoativa. Pero esta regra esixe unha matización: a consecuencia das maiores capacidades combinatorias deste elemento con outras partículas que non esixen o referido corte entoativo, as posibilidades de ausencia de ruptura aumentan na lingua galega.

En resumo, tanto en portugués coma en galego, *olla / olha* defínese como un MD:

- conversacional
- propio da coloquialidade
- situado, comunmente, na posición inicial do discurso
- usado para captar a atención, sexa sobre o receptor, sexa sobre a información
- combinable con outros elementos discursivos
- con autonomía prosódica

Pero tras este funcionamento xenérico e compartido xorden diverxencias entre linguas. As máis importantes encóntranse na capacidade de *olha / olla* aparecer xunto doutros elementos en cada un dos idiomas; en galego, as combinatorias son máis variables e comúns, o que tamén produce que o comportamento deste elemento na liña entoativa non sexa tan facilmente predicible coma no idioma luso. Isto é, en galego *olla* ten un funcionamento máis diversificado. Finalmente, outro punto onde diverxen entre linguas é o corpo funcional desta partícula. Os valores base de *olha / olla* son os mesmos nos dous idiomas, pero non hai coincidencia nos secundarios ou periféricos. Neste caso non hai unha diverxencia na amplitude de opcións, senón que presentan unha xerarquía de predisposición diferente para o conxunto de sentidos.

5. *Mira*

Como indicamos no inicio do traballo, *mira* utilízase como MD nas linguas galega e española, mentres que para o portugués só existe como forma verbal:

Agora, aquietada a imaginação e o ressentimento, mira e remira a alcova solitária (Assis, 1994, cap. 42).

ou como substantivo:

Tem olhos de canção e a mira boa (Carvalho, 1971).

A diferente distribución idiomática non significará necesariamente unha discordancia no funcionamento entre *olla* / *olla* e *mira*, senón que, polo contrario, os puntos comúns serán os máis.

A respecto do estilo de lingua, *mira* aparece no corpus galego como un MD propio da coloquialidade. De 2011 rexistros só 7 se poden adscribir á formalidade por cuestións léxicas e temáticas, aínda que, ó igual có visto para *olla/olha*, neste caso é o MD *mira* o que introduce un matiz de coloquialidade na conversa. A razón disto está en que o uso de *mira* (ou de *olla/olha*) significa a escolla dunha P2 para referirse ó receptor/a, indicativo de proximidade entre os participantes no acto comunicativo.

Todos eses que estaban en organizacións distintas ó Partido acabaron por ingresar nel e ata a ocupar postos de responsabilidade. Mira, o rector da Universidade daquela época era un home que coa Metrópole non pasara de administrativo de segunda e dicían que fora dos antigos independentistas radicais. (Alonso, 2001, 12).

En español, na bibliografía de consulta sobre *mira*, este é definido como unha partícula conversacional, pero non é habitual que se concretice o estilo de lingua que lle é propio (por exemplo, en Martín Zorraquino e Portolés 1999 non hai ningún tipo de especificación sobre o estilo de lingua). A única referencia que atopamos foi no traballo de Pons Bordería (1999). O autor analiza este elemento en ocorrencias coloquiais, polo que podemos concluír que o define como unha partícula prototípica deste estilo.

Y le dije oye mira que ha venido un chico a dejar tu bolso que te lo has dejado en el coche (Pons, 1998, 222).

Deste modo tanto *olla* coma *mira*, indistintamente da lingua, son partículas propias da coloquialidade.

Un segundo punto compartido no funcionamento entre os dous MD é a súa incapacidade para aparecer como unha construción negativa. Nos tres idiomas non é posible negar estes MD.

¿Cómo che farei entender o que pasa?... Mira: os pais de Manoliño deben moitos cartos (Moure Mariño, 1971, 35).

*¿Cómo che farei entender o que pasa?... **Non** mira: os pais de Manoliño deben moitos cartos.

Tamén coinciden na súa posición fundamental inicial dentro da estrutura oracional que modalizan.

Na lingua galega, *olla* e *mira* teñen unha distribución moi parella no que se refire ás súas posibilidades de colocación dentro do discurso.

posición	tipos	galego
inicial		72%
posposto a un elemento externo	nexo coordinante	12%
	MD ou complemento externo	10,5%
	vocativo	1%
posposto a un elemento oracional	tras suxeito	1%
	tras verbo	0,5%
	tras complementos satélites	1,5%
final		1,5%

Mira ou é a primeira palabra

¿Cómo che farei entender o que pasa? ...Mira: os pais de Manoliño deben moitos cartos... (Moure Mariño, 1971, 35).

ou este lugar é cuberto por outro elemento externo co que concorre nesa posición.

O amo dixolle: Pois mira: primeiro, nunca vaias polos atallos (Alumnos Seminario de santa Catalina de Mondoñedo, 1999, 74).

¿De onde é? De Cantador de Enriba, Vaia, mira ti, non se lle nota nada o acento da rexión (Torres, 1971, 173).

Manolo, mira, hoxe pola mañá saíume unha rabañada de doce xabaríns (Alumnos Seminario de santa Catalina de Mondoñedo, 1999, 66).

Xuntando esas dúas posibilidades sumamos o 95,5% do corpus. O restante 4,5% repártese en dous contextos que tamén eran minoritarios para *olla* / *olha*: a posposición a elementos da oración modalizada polo MD ou a posición final no discurso (posibilidade allea para a lingua lusa).

A tipoloxía de complementos oracionais que pode anteporse a *mira* diversifícase a respecto do recollido para *olla* / *olha*, o que se pode xustificar polo tamaño dos corpora (moito maior para *mira*). Pero, de todos os xeitos, os elementos máis recorrentes que presentan esta capacidade son os mesmos que adiantaban a *olla*: por un lado, complementos satélites (con indicación temporal ou de condición amosándose como as máis repetidas), que se illan da información que se vai focalizar a través de *mira*;

E hoxe, mira, do máis inda vou indo, pero das pernas nada (Regueira Fernández, 1989, 94).

e por outro, elementos nucleares (suxeito e verbo) que, ó seren separados da unidade oracional e romper coa liña entoativa habitual, son eles os elementos destacados.

Pois seica tiña dereito a ela, mira ti, por se ter chamuscado un pouco alá atrás nun accidente de tráfico (Ribeiro Loureiro, 1989, 49).

Resta, finalmente, unha última opción no repertorio de posicións de *mira* na lingua galega. Tal e como *olla* neste idioma, cando *mira* segue a información que se quere focalizar e ocupa o último lugar da construción, aparece normalmente combinado con outros elementos ou configurando unha estrutura frástica illada, na que el é o primeiro elemento.

Tanta gracia que, iso: non habería de parar de rir ata mañá. Ti mira (Rei Ballesteros, 1992, 286).

Pero, a diferenza de *olla*, cubrindo esta posición pode aparecer el só como MD.

¡É que che levaba máis de dez mil pesos, mira! (Regueira Fernández, 1989, 39).

É unha posibilidade moi reducida para o MD, o que explica que o seu valor focalizador se intensifique. Polo tanto, *mira* na lingua galega conta, de forma moi extraordinaria, coa opción de pechar o discurso, sempre que se queira aumentar a carga comunicativa da partícula.

Neste punto, aínda xurdiu unha singularidade no funcionamento de *mira* na lingua galega. Cando se rexistra o encontro *mira* + *ti*, resto fosilizado do pasado verbal do MD, a única opción construtiva non é a anteposición do verbo, como acontecía con *olla*, senón que este se pode pospor ó pronome (*ti mira*).

Non habería de parar de rir ata mañá. Ti mira (Rei Ballesteros, 1992, 286).

Esta orde dos elementos recóllese en 25 rexistros fronte a 168 en que se comparte a mesma disposición cá propia para *olla*. Ou sexa, a orde propia ós constituíntes da oración para a lingua galega é a menos común para este MD composto. Deste modo enténdese que, máis que estarmos ante unha situación lingüística diferente á vista para *olla ti* (a fixación dunha estrutura frástica), a redución de uso do encontro *ti mira* carga o MD composto dunha maior calidade intensificadora sobre a información que se quere focalizar, de modo que se mostra como unha estratexia comunicativa singularmente marcada.

En español, Gallardo Paúls et al. (2005: 62) indican que “el carácter parentético (...) hace que los marcadores discursivos [entre os que está *mira*] se sitúen en la periferia de la oración, generalmente al comienzo, si bien también hemos encontrado algunos casos de posición final”. Por conseguinte, marcan como esperable a posición inicial para este MD:

Mira, non lo sé es que, es todo y no es nada (Pons, 1998, 224).

E recoñecen a posibilidade de que peche a estrutura oracional como moito menos habitual.

Yo, ni lo he cogío, mira (Pons, 1998, 220).

Ó referírmonos ó valor que adquire *mira* unha vez se pospón a certas unidades sintácticas do discurso que o MD vai modalizar, fixemos unha referencia á utilidade comunicativa que adquire a partícula na conversa: facer unha chamada de atención sobre o receptor ou sobre a información.

Mira na lingua galega, en trazos xerais, conta coa mesma utilidade comunicativa ca *olla*, tanto no portugués coma no galego: presenta dúas funcións fundamentais (captar a atención, sexa do receptor, sexa sobre a información), as cales, á súa vez, inclúen diferentes subfuncións. Como acontecía con *olha/olla*, a importancia dos dous grandes macrosentidos de *mira* é moi similar.

Valor	subvalor	galego
chamada de atención sobre o receptor	total	48%
	informar	14,5%
	resposta	11%
	apelativo	5,5%
	preguntar	1%
	advertir	16%
chamada de atención sobre a información	total	52%
	intensidade ou énfase	26,9%
	estupefacción	16%
	suavizar	0,1%
	conclusión	4%
	explicación	5%

No que se refire ás subfuncións, dominan as mesmas ca con *olha / olla*. Cando se procura captar a atención do receptor é

porque se vai introducir unha advertencia ou sinxelamente informal sobre un aspecto.

—¿Vedes alguén? —Non, seguíde —Mira que avisedes, poñen multa de quinientas pesetas (Torres, 1971, 106).

E, por outro lado, a focalización da información procura na gran maioría das estruturas intensificar a calidade, importancia ou interese do dato introducido.

Mira que hai xente pesada, machuca que machuca co tema (Torres, 1971, 23).

En galego, tanto con *olla* coma con *mira*, o seguinte valor máis empregado é marcar o dato como un feito sorprendente ou que causa estupefacción.

Pero mira, eu cando lle (ó meu irmán) vin o paraguas, no baile, co paraguas, no baile, co paraguas colgado por atrás, mira, aquilo, aquilo, aquel... (Regueira Fernández, 1989, 31).

Unha vez que nos afastamos dos sentidos dominantes, diverxen lixeiramente as porcentaxes de uso dos outros sentidos que poden adquirir os referidos MD na lingua galega, pero non a distribución na escala de predisposición de cada un deles (neste plano, pola contra, *olha* en portugués si ten un funcionamento diferencial). Polo tanto, no que respecta á funcionalidade comunicativa, *olla* e *mira* son partículas equivalentes no idioma galaico.

Por outro lado, no conxunto dos traballos consultados encontrámonos con que a terminoloxía utilizada para describir os valores de *mira* en español é moi diversa. Aínda sendo así, a través dos exemplos e das definicións destas etiquetas podemos establecer certos paralelismos coas que utilizamos para a nosa análise. Seguindo as propostas de Pons (1998), Gallardo e Marín (2005) e Galué (2002), defínese o funcionamento focalizador desta partícula en dúas direccións:

- captar a atención do receptor, un valor especialmente focalizado por Pons (1998, 217)

Pues mira que tomar el baño en tu piscina, eso también tiene, eso es boniquet (Pons, 1998, 217).

- repercutir na intensidade da información, ó indicar “un cambio en la evolución del discurso” (Gallardo Paúls, 2005, 66); no traballo de Gallardo Paúls (2005, 78) estes valores en *mira* ocupan ó redor do 66% do seu corpus.

Miara, ehh, yo, no tengo tardes libres, eh? (Gallardo Paúls, 2005, 63)

Pero a diferenza do visto para *olla*, tanto en galego coma en portugués, e para *mira* en galego, neste MD en español o máis recorrente non é o seu uso como un elemento focalizador, senón como unha estratexia organizadora do discurso. Emprégase para introducir as quendas de fala, para comezar, mudar ou pechar un tema ou para pasar a tratar un subtema. Por exemplo, son un elemento fundamental para a estrutura global da conversa (Gallardo Paúls, 2005, 66):

Bueno/ yo/ es que/ ¿sabes lo que pasa?// hay cosas/ que de repente?// parece que no/ ¿no?// pero/ inglés y francés// por ejemplo/ Raúl/ al final cuesta quee/ de repente/ quiero expresar unha cosa enn// enn nee/ castellano// ¿eh?/ y de repente/ puess/ no sé por qué/ me salen palabras en inglés y francés/ ¿no? Y a esto// valenciano// ¿eh? Yy// no sé/ este/ me gustaría – tengo/ necesidad dee expresar/ paraa/ que todo el mundo / ¿eh?/ en vez de – dee afasia/ me gustaría/ no afásico/ pero mira/ es más? / se ve qu que/ mucho tiempo/ seguramente/ pero me es igual (Gallardo Paúls, 2005, 66).

Deste modo, mentres nas outras linguas predomina o uso de *olla* ou *mira* como estratexia intensificadora, en español este sentido de *mira* semella de menor peso fronte á función de organizador do discurso.

Alén de diferenciarse *mira* en galego e español no que se refire á súa funcionalidade no discurso, tamén presentan un comportamento distinto a respecto das capacidades combinatorias do MD.

Na lingua portuguesa, *olha* é un MD que se usa habitualmente sen acompañamento, e de habelo serán outros

MD (cos que se pode acumular ou combinar) ou un ou máis vocativos.

En español, *mira*, como todos os MD, ten a capacidade de combinarse con outras partículas máis ou menos fixas, pero tampouco aparece como a situación maioritaria. Os encontros máis frecuentes de *mira* rexístranse tamén xunto doutros complementos externos (*en cambio, entonces, es decir, finalmente, no obstante, sin embargo, pues, por lo tanto, interxeccións...*) ou con outros MD.

¿Qué es lo que pensaban al inicio y que es lo que piensan ahora? (...) Pues mira, de entrada, hay una cosa particular en la carrera, por lo menos en la escuela que nos tocó o que elegimos estar que fue la escuela de arte teatral de Bellas Artes (Muñoz, 2009).

Nos diferentes traballos só aparecen especificadas determinadas combinatorias entre MD. Pons Bordería (1998, 222) rexistra a acumulación de *oye mira* e destaca a 'énfasis que añade la repetición';

Y le dijee oye mira que ha venido un chico a dejar tu bolso que te lo has dejado en el coche (Pons, 1998, 222).

ou Galué (2002, 37) "el marcador [usado para recuperar a quenda de fala] aparece reforzado por otra forma apelativa como *ven acá* con la que se procura, de manera enfática, captar la atención del otro".

H3. No era un engaño.

Y3. Porque qué era lo que trataba [de decirle al pueblo].

H4. pero, mira, ven acá

Y4. yo no voté porque Chávez me estaba ofreciendo a mí una constituyente (...) (Galué, 2002, 37).

De modo que só refiren casuísticas de acumulación de MD e non rexistran a posibilidade de que *mira* sexa un formante dun MD composto.

Na lingua galega as posibilidades diversifícanse. *Olla* neste idioma adoita aparecer xunto doutras partículas: MD, outros complementos externos ou restos fosilizados de com-

plementación de *olla*. Para *mira* nesta lingua rexistrouse a mesma situación. Este MD adoita acompañarse dunha destas partículas, funcionando como unha acumulación de elementos externos á oración ou como un MD composto.

lingua	funcionamento do MD		
	illado	combinado	
		MD composto	acumulación de MD
galego	34,5%	48,5%	17%

Aínda que os tipos de partículas que se xuntan son os mesmos en *olla* ca en *mira*, con este último recollemos un abano moito maior de elementos. De todos os xeitos, os combinados máis recorrentes coinciden cos rexistrados para *olla* na lingua galega.

Así, cando encontramos unha acumulación de MD o máis recorrente é que apareza unha reiteración de *mira* (48 rexistros)

E pois, ¿sabías qué fixo? Mira, mira... fixo masa con el, e pois comeu (Costas González, 1986, 247).

ou a xuntanza deste MD e un vocativo (301 rexistros);

¿Cómo foi que o trincaron..., despois? —Mira, rapaz, naquela fonda do Bidueiro barullaban demasiado (Gómez Vilavella, 1992, 139).

mentres que se o resultado final é un MD composto, dominan, ó igual ca para *olla*, *mira que* (440 rexistros),

Mira que gastou tempo varada nesta cama (Torres, 1971, 67).

mira como (79 rexistros)

Mira como se dá a tempra ¿ves? (Varela Buxán, 1975, 12).

ou *pois mira* (137 rexistros).

Eso, i así, sentados, e pagannos por estar sentados. Pois mira, sentados vamos andando (Vázquez Somoza, 1986, 140).

Os valores dos combinados e dos encontros de MD son os mesmos cós que presentan os correspondentes formados con *olla*. A única diferenza entre estes dous MD está en que xunto de *mira* as posibilidades combinatorias pluralízanse, singularmente no que se refire á acumulación de MD ou MD e complemento externo:

bueno mira:

Dixen eu: —Bueno, mira, eu vouche comer (Couceiro Pérez, 1976, 166).

mira sabes:

¡En fin, ho...! O carallo vintenove, mira, ¿sabes? Fora a partir daquel momento cando empezara el a se encontrar verdadeiramente cómodo con ela igual ca se un tabú acabase de ser tombado entre eles dous (Rei Ballesteros, 1992, 309).

mira claro:

—¡Quedouselle morta! {GL}— Morta; mira, claro, era co mal o animal subía por as paredes (Sánchez Rodríguez, 1983, 219).

mira vaia:

Teño eu, prestoumo Ramiro antes. Mira, vaia montón (Borrazás, 1991, 35).

Polo tanto, *mira* amosa unha maior capacidade ca *olla* de compartir espazo con outras partículas, froito do seu alto uso como MD na lingua galega.

En resumo, é na lingua galega onde os MD *olla* e *mira* presentan como habitual unha situación que nos outros dous idiomas só é unha opción de construción. Nas tres linguas as formas coas que se combinan estes elementos son outros MD ou diferentes complementos externos, pero *mira* destaca na lingua galega sobre o resto na diversidade de elementos cos que poder coaparecer.

Imos resaltar destas outras opcións dúas estruturas en que *mira* se segue por un nexo prepositivo (*para* e *de*). Son formantes dun MD composto, xa que comparten unidade entoativa e o significado sempre é o mesmo: focalizan a

información para marcala ou como un feito que produce estupefacción ou como exemplo que explique o antes dito.

Mira pr'el, como se lle conoce o que...o que le (Regueira Fernández, 1989, 28).

¡Cuantos borrachos hai por eiquí neste sitio! Mira deste ladrón a qué anda, a prender este cabalo (Roux, 1982, 37).

Estes adxuntos sérvennos á vez para tratar aqueles MD compostos formados por unha estrutura frástica fosilizada de *mira*. Observando especialmente o conxunto *mira para* (25 rexistros) semella que estamos ante un verbo de percepción visual xunto do complemento de dirección cara a onde se debe levar a vista. Pero non é tal. *Mira* perdeu o seu significado sensorial e o complemento xa non indica un destino locativo, senón que introduce o elemento que se quere focalizar. Dalgún modo herda a interrelación orixinal, pero perde a natureza de estrutura con acción verbal. Pasan a funcionar como un MD.

Hai un resultado idéntico xunto doutros elementos que son restos de complementación verbal. Tal e como con *olla*, resólvense nun MD composto os encontros *mira ti* (ou *ti mira*), no que o segundo elemento remite ó suxeito fosilizado do verbo e, comunmente, pecha a unidade (168 rexistros),

¡Ai!... eu aínda me quero mellor aquí de noite ca na roza de día, mira ti (Graña Muíño, 1995, 63).

aínda que tamén pode iniciala (25 rexistros);

O pobre home chegou sangrando á aldea e alí o curaron. Ti mira que historia (Toro, 1993, 142).

mira ben (15 rexistros), MD que introduce unha advertencia e ten un funcionamento paralelo a outros do tipo *a ver*;

¡Mira ben o que fas, filliña, mira! (Varela Buxán, 1975, 107).

ou o encontro de *mira* cun complemento directo fosilizado. Neste último tipo, amais de *unha cousa* (9 rexistros), tal e como o rexistramos con *olla*,

-Mira unha cousa. ¿Por que non te agachas no noso soto?
(Calo Lourido, 2001, 115).

tamén presentan un mesmo funcionamento os encontros *mira o que che digo* (1 rexistro)

Apañámonos sen ti, aquí non nos fas falta. E mira o que che digo, que serán tódo-los meus desexos que non veñas, querido Ramón (Angueira Viturro, 1999, 183).

ou *mira o que son as cousas* (1 rexistros), este último marcando a información como un resultado sorprendente e abandonando xa a idea de advertencia.

Entón un parente mercou aquilo e marcharon para Madrid, e mira o que son as cousas, ó pouco tempo morreulle-lo tío (Araúxo, 1996, 99).

Nos tres casos a fixación das construcións tamén vén favorecida polo uso de lexemas semanticamente xenéricos como *é cousa*, que permite o uso discursivo do conxunto.

Por outra banda, temos que lembrar que para *olha* en portugués este tipo de combinatorias non foron recollidas no corpus, amosando un maior desligamento entre o seu funcionamento verbal e o discursivo. Pola contra, *mira* en español pode combinarse cun elemento tamén xenérico (*una cosa*), derivado dun complemento directo do verbo, e pasar a funcionar como un MD composto. Gallardo Paúls nunha nota de rodapé matiza a súa afirmación de que *mira* non ten complementos ó indicar que “en la conversación espontánea o en el debate (...) es frecuente identificar ejemplos como *mira una cosa*” (2005, 61). Neste traballo non consideramos que nestas situacións o MD *mira* (ou *olla*) conserven a capacidade de rexer complementación verbal, senón que son construcións frásticas fixadas que adquiren o rol de MD, neste caso composto.

Mira una cosa si tu fueras a otro país te gustaría ke te trataran bien o como una mierda (Ayoze, 2008).

Polo tanto, recollemos en español un dos exemplos de MD frásticos con *olla* ou *mira* na lingua galega. Pero é o único caso de todos as posibilidades rexistradas para estes dous MD

no idioma galaico; por exemplo, o combinado máis recorrente neste código, *olla / mira ti*, non ten paralelo en español. Nesta lingua, estruturas como *mira, tú* non funcionan como un MD único, senón como un MD e un vocativo, situados en dous cortes entoativos diferentes.

O comportamento na liña entoativa é útil para opor o funcionamento de *mira ti* a *mira, tú*, porque en grande medida o seu desenvolvemento neste nivel é un dos trazos que caracterizan os MD *olla / olha* e *mira* como partículas discursivas.

En portugués *olha* trae consigo un corte na liña entoativa como principio, singularmente cando non forma parte dun MD composto, xa que nestes casos as opcións de que a liña entoativa sexa continua aumentan. En galego, con este mesmo MD, a situación diversifícase porque, ó crecerea as súas capacidades combinatorias, redúcese a súa esixencia de separación do resto do discurso. Deste modo vimos que o funcionamento a este respecto é diferente se *olla* funciona como un MD simple (o habitual é que se ille do resto), a se forma parte dun MD composto (domina a ausencia de ruptura entoativa), fundamentalmente xunto de *que* ou *como*. Con *mira*, nesta mesma lingua, a situación é parella á descrita para *olla*: o comportamento na liña entoativa destes MD diversifícase, froito das súas capacidades combinatorias, pero pódese establecer unha regra de predisposición segundo funcione como un MD simple (tende cara ó illamento entoativo) ou un MD composto (prefire a ausencia de interrupción entoativa, xa que é un único elemento discursivo)

posición	construción	galego
illado	MD simple	36%
	MD composto	15,5%
non illado	MD simple	20%
	MD composto (olla que/como)	28,5%

En español, Pons Bordería apunta que “desde el punto de vista entonativo, estas formas no suelen estar integradas en la línea melódica del enunciado en el que se insertan” (1998, 218); polo tanto, o habitual é que apareza illado da estrutura.

Es que, mira, ¿qué grupo de música es? (Pons, 1998, 224).

Deste modo, o funcionamento de *mira* en español e de *olla* en portugués son máis próximos, porque para estes dous idiomas a situación prototípica é a illada, mentres que no galego -para os dous MD- o illamento da partícula vai depender en grande medida das súas combinatorias.

Concluindo, tanto en español coma en galego, *mira* defínese como un MD conversacional propio da coloquialidade, incompatible coa negación e con predisposición para cubrir a posición inicial da oración. A partir de aquí comezan as diverxencias que afastan este MD entre os dous idiomas, pero que aproximan *olla* e *mira* na lingua galega. Estamos a nos referir ó valor comunicativo das partículas (focalizador en galego, pero organizador do discurso en español), ás súas capacidades combinatorias (moito maiores en lingua galega) e, vencellado con isto último, ó seu comportamento na liña entoativa (máis variable no idioma galaico). De modo que, aqueles puntos que afastan *mira* en galego e español, son os que sitúan *olla* e *mira* (en galego) nun mesmo paralelo.

6. Conclusión

A finalidade deste traballo era facer unha análise de dous MD cunha mesma orixe (derivan de verbos de percepción visual) en tres linguas contiguas e membros dunha mesma familia (galego, portugués e español). A análise ten como obxecto a busca de puntos comúns e discordancias entre as catro formas, fundamentalmente para tentar xustificar a diferenza de uso entre *olla* e *mira* no mesmo código, é dicir, en galego.

Estes MD, indistintamente da lingua en consideración, comparten trazos definitorios do seu funcionamento:

- o seu nivel de actuación é a conversación, é dicir, son ferramentas do discurso dialogado con interacción entre os participantes;
- o estilo de lingua condiciona o uso destas partículas, xa que a súa presenza trae consigo coloquialidade no tratamento da información e na relación entre emisor e receptor;
- e particularidades definitorias da categoría MD son cumpridas por todos eles:
 - non é posible a súa negación
 - teñen a facultade de combinarse con outros MD ou complementos externos á unidade sintáctica

Por outro lado, os aspectos que afastan o funcionamento de *olla* ou *mira* entre o galego e a outra lingua en que o respectivo MD é utilizado son xustamente características compartidas polas dúas partículas na lingua galaica:

- a disposición do MD na orde de elementos

Tanto *olha/olla* coma *mira* (en galego e en español) ocupan comunmente a posición inicial na estrutura; se non o fan, será porque este lugar xa foi cuberto por outra partícula discursiva que concorre cos referidos MD por esta posición.

Pero neste punto xorde unha diferenza entre *olha* e *olla*: na lingua galega, este MD pode pechar a estrutura. O que é unha diferenza entre linguas pasa a ser unha coincidencia entre *olla* e *mira*, xa que este último, tanto en galego coma en español, pode tamén cubrir esta posición.

- os efectos do MD na liña entoativa do discurso

Funcionar como unha partícula discursiva significa un rol periférico na estrutura; á súa vez, este papel tradúcese normalmente nun illamento entoativo. Deste modo, enténdese que tanto con *olha* en portugués coma con *mira* en español a súa presenza signifique un corte na liña entoativa. Pola contra, estes dous MD en galego non carretan necesariamente unha

interrupción desta natureza; a capacidade de *olla* e *mira* neste idioma de situarse a parte da liña entoativa do discurso vai depender das partículas coas que se combine: se estamos ante unha acumulación de complementos externos, o común será que o MD se ille do discurso, pero se forma parte dun MD composto, entón non haberá unha única unidade entoativa.

- a predisposición de adxuntarse a outros elementos externos

Para todos os MD a combinación con outros MD ou outros elementos externos é unha propiedade común. A diferenza está no grao de desenvolvemento desta capacidade. En galego estes dous elementos acostuman acompañarse doutras partículas periféricas á estrutura sintáctica, resultando nunha acumulación de elementos discursivos ou configurando MD compostos; pola contra, tanto en portugués coma en español, aínda sendo a combinación unha opción de *mira* ou *olha*, non representa unha porcentaxe de peso similar ó recollido para as formas galegas.

- a conservación de pegadas sintácticas do seu pasado verbal

Estes MD derivan de verbos de percepción visual; este funcionamento verbal xustifica que nalgúns casos houbera unha fixación frástica que se desenvolveu nun único MD. En galego, tanto *olla* coma *mira* aparecen funcionando como un MD composto xunto de complementos verbais fosilizados de diversa natureza (o suxeito, diferentes complementos directos ou un circunstancial de modo), mentres que en portugués non se recolleu esta posibilidade e en español só se rexistrou xunto dunha das posibilidades recollidas na lingua galaica.

- o valor comunicativo das partículas

En español *mira* ten como función primeira a de organizador do discurso, mentres que a de focalizador é secundaria. Para os outros dous idiomas, estes MD son

focalizadores que, nun plano secundario, poden aproximarse ó rol de organizadores.

Á súa vez, como focalizadores, dirixen a atención ou sobre o receptor ou a propia información. O peso destes dous sentidos é similar para os dous MD e tanto en galego e portugués; as diverxencias entre idiomas encóntranse na predisposición cara ós subvalores secundarios que estas partículas poden adquirir. Teñen unha xerarquía diferente que opoñen, por un lado, *olla* e *mira* en galego a, polo outro, *olha* en portugués.

En conclusión, o comportamento destes catro MD ten moitos puntos comúns, pero o de *mira* e *olla* en galego é practicamente idéntico. Por conseguinte, entramos agora no segundo obxectivo deste traballo: a busca dunha xustificación que explica a diferenza de uso que se rexistra entre estes dous MD no idioma galego.

Se o funcionamento de *mira* e *olla* é moi similar, a escolla preferente na lingua galega do primeiro fronte ó segundo non se pode deber a razóns puramente lingüísticas. Haberá que buscar a explicación noutro nivel, e nós encontrámolo no plano sociolingüístico.

Son numerosos os exemplos de pares máis ou menos sinónimos galegos, dos que un dos membros se viu promocionado, entre outras razóns, por coincidir cunha forma española: desde (22908 rexistros no corpus Corga) fronte a dende (10952 rexistros no corpus Corga), primo (723 rexistros no corpus TILG) fronte a curmán (587 rexistros no corpus TILG), etc. Entre estes pares, sempre se rebela como o máis usado a forma idéntica á da lingua española. É dicir, a presión da lingua (elite) colonizadora explica a selección feita pola comunidade, xa que esa escolla está máis ligada á idea de poder e de prestixio social que os falantes teñen (ou tiveron).

A linguaxe é o lugar onde os individuos se representan e constitúen o mundo, pero é necesario ver a linguaxe como lugar de interacción de constitución de identidades e de coparticipación. Así se explica que nunha situación de conflito lingüístico actual e dunha anterior diglosia alongada no marco

temporal a comunidade de falantes tentase, se non asimilar todos os parámetros que estaban vencellados cun status superior, polo menos achegarse o máis posible a eles; ou sexa, a nivel lingüístico, se non se pasa a falar castelán, cando menos promover as formas coincidentes coas deste idioma en detrimento das diferenzas. Este é o criterio básico que xustifica o éxito de *mira* fronte a *olla* no galego.

Fontes utilizadas:

Lingua Portuguesa

Corpus do Português Oral: <http://www.clul.ul.pt>

Assis, Machado de (1994): *Quincas Borba en Obra completa*. Rio de Janeiro: Editora Nova Aguilar.

Bacelar do Nascimento, Maria Fernanda, et al. (1987): *Português fundamental: métodos e documentos*. Tomo I. Lisboa: INIC, CLUL.

Carvalho, O. G. Rêgo de (1971): *Somos todos inocentes*. <<http://www.corpusdoportugues.org>>.

Lingua española

Ayoze (2008): "Que pasada" en *Argentina Forums*. <<http://www.justlanded.com>>. [07/05/08].

Gallardo Paúls, Beatriz; María Josep Marín Jordà (2005): "Marcadores discursivos procedentes de verbos perceptivos en el discurso afásico", *Revista de Investigación Lingüística* 8, 53-94.

Galué, Dexy (2002): "Marcadores conversacionales: un análisis pragmático", en *Boletín de Lingüística*: 18. Caracas: Universidad Central de Venezuela, 27-48.

Muñoz, Sandra (2009): "Entrevista a Dana Aguilar y Hector Peña", en *Tampico Cultural*. <<http://www.tampicocultural.com>>.

Pons Bordería, Salvador (1998): "Oye y mira o los límites de la conexión", en *Los marcadores del discurso: Teoría y análisis*. Madrid: Arco, 213-228.

Lingua galega

Tesouro Informático da Lingua Galega (TILG):

<http://www4.usc.es/TILG/>

A Fuliada 11, 12/04/1888.

Alonso, Dámaso (2001): *Narraciones orales gallego-asturianas*.

Tomo I. Madrid: Gredos.

Alumnos do Seminario de santa Catalina de Mondoñedo (1999):

A carón do lume. Lugo: Citania.

Angueira Viturro, Anxo (1999): *Pensa nao*. Vigo: Xerais.

Araújo, Emilio (1996): *Tempos serodios*. Vigo: Cumio.

Barxa Iglesias, Francisco (1984): *Os cabalos da troula*. Vigo: Nova Galicia.

Blanco Amor, Eduardo (1972): *Xente ao lonxe*. Vigo: Galaxia.

Blanco Amor, Eduardo (1974): *Teatro para a xente*. Vigo: Galaxia.

Borrazás, Xurxo (1991): *Cabeza de chorlito*. Santiago de Compostela: Sotelo Blanco.

Cabana, Darío Xohán (1996): *Morte de Rei*. Vigo: Xerais.

Calo Lourido, Francisco (2001): *Salseiros*. A Coruña: Espiral Maior.

Cao Martínez, Camilo (1983): *Idilios do Feal de Ansemil*. Vigo: Ed. do autor.

Catoira, Manuel (1977): *A cadea e A outra banda do silencio*. Lugo: Celta, Lugo.

Cid Cabido, Xosé (1988): *O camiño de Middeharnis*. Vigo: Xerais.

Cortezón Álvarez, Daniel (1981): *A vila sulagada*. Sada-A Coruña: O Castro.

Costas González, Xosé Henrique (1986): *Aproximación lingüística á fala de Vincios*. Memoria de Licenciatura. Santiago de Compostela: USC.

Couceiro Pérez, Xosé Luís (1976): "El habla de Feás", en *Verba*, anexo 5. Santiago de Compostela: USC.

Díaz Fernández, Xerardo (1985): *Contos*. Montevideo: Patronato da Cultura Galega.

Eiré, Afonso (2000): *Eu tamén fun coas vacas*. A Coruña: Espiral Maior.

- Gómez Vilavella, Xosé María (1992): *Historias intra da bisbarra: Castroverde, Baleira, Pol*. Lugo: Deputación de Lugo.
- Graña Muíño, J.L. (1995): *Cerdido: lingua e outras realidades*. Vigo: Ed. do autor.
- Ledo Andión, Margarita (1985): *Trasalba ou Violeta e o militar morto*. Vigo: Xerais.
- Lorenzo Fernández, Xoaquín (1973): *Cantigueiro popular da Limia Baixa*. Vigo: Galaxia.
- Martínez Oca, Xosé Manuel (1983): *Beiramar*. Vigo: Xerais.
- Martínez Oca, Xosé Manuel (1993): *Os culpables*. A Coruña: Vía Láctez.
- Martínez Oca, Xosé Manuel (1995): *Náufragos en terra*. Vigo: Xerais.
- Martínez Oca, Xosé María (1981): *A escura chamada dos cavorcos*. Vigo: Xerais.
- Méndez Ferrín, Xosé Luís (1980): *Crónicas de Nós*. Vigo: Xerais.
- Méndez Ferrín, Xosé Luís (1986): *Antón e os inocentes*. Santiago: Sotelo Blanco.
- Méndez Ferrín, Xosé Luís (1994): *Estirpe*. Vigo: Xerais.
- Miranda, Xosé (1998): *Morning Star*. Vigo: Xerais.
- Moure Mariño, Luís (1971): *Sempre matinando*. Vigo: Galaxia.
- Pla Xubiri, Juan (1917): *A tola de covas*. Madrid: La Itálica.
- Regueira Fernández, Xosé Luís (1989): *Textos dialectais do norte da Terra Chá*. Tese de doutoramento, Santiago de Compostela: USC.
- Rei Ballesteros, Anxo A. (1992): *Loaira*. Vigo: Galaxia.
- Ribeiro Loureiro, Manuel (1989): *O coronel ten a quen lle escribir*. Vigo: Xerais.
- Rodríguez Troncoso, Manuel (1996): *Raíces rurais*. Sada: O Castro.
- Roux, Martine (1982): *Fala o fistor e faise día: algúns aspectos da tradición oral galega*. Sada: O Castro.
- Sánchez Rodríguez, Helena (1983): *A fala da parroquia de Seteventos*. Memoria de Licenciatura, Santiago de Compostela: USC.
- Toro, Suso de (1993): *Tic-tac*. Barcelona. Ediciones B.

- Torres, Xohana (1971): *Adiós María*. Buenos Aires: Ediciones Galicia.
- Varela Buxán, Manuel D. (1975): *O ferreiro de Satán. Taberna sin dono*. Lugo. Celta.
- Vázquez Somoza, Xosé Luís (1986): *O galego en Zolle*. Tese de licenciatura, Santiago de Compostela: USC.

Referencias bibliográficas

- Domínguez Portela, Soraya (2006): "Pois: dende a conxunción ó marcador discursivo", *Cadernos de Lingua* 28, 145-173.
- Freixeiro Mato, Xosé Ramón (2005): *Os marcadores discursivos: conectores cotraargumentativos no galego escrito*. A Coruña: Universidade da Coruña.
- Gallardo Paúls, Beatriz; María Josep Marín Jordà (2005): "Marcadores discursivos procedentes de verbos perceptivos en el discurso afásico", *Revista de Investigación Lingüística* 8, 53-94.
- Galué, Dexy (2002): "Marcadores conversacionales: un análisis pragmático", en *Boletín de Lingüística*: 18. Caracas: Universidad Central de Venezuela, 27-48.
- Martín Zorraquino, María Antonia; José Portolés Lázaro (1999): "Los marcadores del discurso", en *Gramática descriptiva de la lengua española*. Vol. 3. Madrid: Real Academia Española / Espasa Calpe, 4051-4213.
- Pons Bordería, Salvador (1998): "Oye y mira o los límites de la conexión", en *Los marcadores del discurso: Teoría y análisis*. Madrid: Arco, 213-228.
- Rost, Cláudia Andrea (2005): "Mapeamento dos contextos de actuação de olha e veja", en *Abordagens funcionais da língua: temas de pesquisa*. Trindade: Universidade Federal de Santa Catarina.

A PROSODIA DE CAMELLE: DESCRICIÓN ACÚSTICA DUNHA FALA¹

ANA BELÉN ESCOURIDO PERNAS
Universitat de Barcelona

RESUMO

Preséntase o Córpora e a metodoloxía que o grupo galego do proxecto *Atlas Multimedia Prosodique de l'Espace Roman* (AMPER) emprega nas súas investigacións. E seguindo estas pautas descríbese o comportamento dos principais índices acústicos da prosodia (F0, duración e enerxía), a partir das gravacións feitas a unha informante feminina natural do concello de Camelle, na Costa da Morte.

ABSTRACT

"The prosody of Camelle: an speech acoustic description"

It is presented the Corpus and methodology that the Galician group of the project *Atlas Multimedia Prosodique de l'Espace Roman* (Amper) uses in its researches. Following these guidelines, it is described the behaviour of the main prosody acoustic indexes (FO, length and energy), from the recordings done to a feminine informant from Camelle Council, in Costa da Morte.

PALABRAS CHAVE

prosodia, dialectoloxía, focalización, clash acentual

KEY WORDS

Prosody, dialectology, focalization,...

¹ Este traballo realizouse co financiamento do Ministerio de Educación y Ciencia, ao abeiro do proxecto "Descripción y estudio de los modelos entonativos en el gallego actual" (HUM2005-08282-C02-01).

1. Introducción

Esta investigación segue a metodoloxía do proxecto internacional *Atlas Multimedia Prosodique de l'Espace Roman* (AMPER). Este proxecto nace coa intención de estudar a prosodia nas linguas románicas, seguindo uns obxectivos e unha metodoloxía común, de tal xeito que se poidan cotexar estruturas fónicas idénticas de calquera variedade lingüística e achegar conclusións fiables en torno aos diferentes parámetros acústicos que interveñen na prosodia. O proxecto ten unha orientación dialectolóxica, polo que se pretende afondar nas variedades de cada dominio lingüístico coa finalidade de confeccionar un atlas multimedia da Romania, no que poidamos consultar gráficos de F0, duración e enerxía, así como escoitar os arquivos de son orixinal e sintetizado.

A coordinación xeral do proxecto é responsabilidade dos profesores Michel Contini e Antonio Romano do *Centre de Dialectologie* da Université Stendhal-Grenoble III (Francia). Porén, para abranguer mellor a totalidade do espazo da Romania, fóronse incorporando distintos equipos de investigación multidisciplinares en varios dominios lingüísticos².

Para o dominio galego, formouse un grupo ao amparo do *Instituto da Lingua Galega* (ILG) da Universidade de Santiago de Compostela e coordinado pola profesora Elisa Fernández Rei. O comité galego do AMPER dedícase ao estudo da prosodia nas variedades dialectais do galego, que abranguen a comunidade autónoma de Galicia, parte de Asturias e de Castela e León.

Seguindo as indicacións específicas do proxecto realizáronse gravacións en varios puntos da xeografía galega. Dentro da área fisterrá (F. Fernández Rei 1991), seleccionouse como fala representativa da zona a vila de Camelle, no extremo norte do occidente da provincia da Coruña.

² Dentro da península ibérica formouse AMPER-PORTUGAL para o dominio do portugués, AMPER-CAT para o do catalán, AMPER-ASTUR para o asturiano, AMPER-Galicia para o galego e AMPER-ESP para o español.

Neste estudo faremos unha descrición do comportamento dos principais índices acústicos da prosodia (F0, duración e enerxía), a partir das gravacións feitas a unha informante feminina de 46 anos, natural da vila, e sen estudos superiores.

2. *Córpora*

Córpora deseñados para o proxecto AMPER

Un dos principais inconvenientes á hora de abordar un estudo da entoación de calquera lingua encóntrase na escolla dun método apropiado aos nosos intereses que nos permita conseguir un repertorio de unidades fiable: por un lado, un corpus ríxido en canto á súa estrutura (sintaxe, léxico, etc.) non permite recoller a espontaneidade e naturalidade da fala cotiá e, por tanto, a entoación non será fiel; e á inversa, outro máis libre, que permita en maior ou menor medida unha fala distendida do informante, dificulta a comparación de enunciados semellantes entre todas as linguas románicas e a consecución de conclusións fiables. Para poder cubrir con maior amplitude a prosodia dos informantes, o proxecto AMPER utiliza catro tipos de córpore diferentes, desenvolvidos especificamente para cada unha das linguas estudadas: o *corpus fixo*, o *corpus dirixido* ou *inducido*, o *Map Task* e o *libre*, cos que se pretenden paliar os problemas expostos.

a) O obxectivo do **corpus fixo** consiste en gravar unha serie de frases cunha estrutura determinada que permita comparar a entoación propia das diferentes linguas románicas nos mesmos esquemas métricos e sintácticos. Para o seu desenvolvemento, pídesse ao informante que entoe unha serie de enunciados a partir duns deseños elaborados e que se lle presentaron previamente. A utilización de debuxos en fichas de papel, en lugar da lectura de frases escritas débese a un intento de aproximación ao proceso mental xerado polo locutor no acto creador da fala. O proceso de escritura require o emprego dunha zona cerebral

distinta que o da oralidade, polo que a utilización dun corpus formado por oracións escritas pode inducir ao informante ao emprego dunha entoación de lectura.

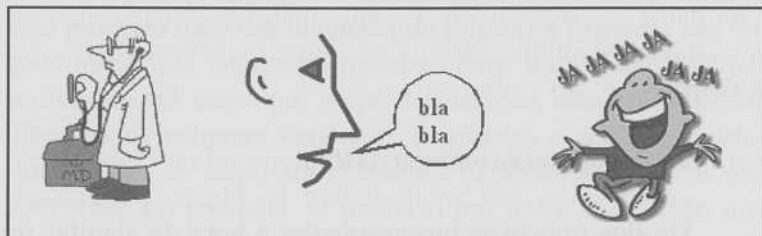


Figura 1: Exemplo das fichas utilizadas na gravación. Corresponde a *O médico falaba co cómico.*

O *corpus fixo*, que empregamos neste estudo, está formado por unha serie de frases coa estrutura:

Suxeito + Verbo + Complemento

O suxeito e o complemento están formados por unha frase nominal composta por un artigo, un substantivo e ademais, naquelas frases que denominamos con extensión, un adxectivo. Segundo o tipo de acentuación: oxítóna, paroxítóna e proparoxítóna empréganse tres suxeitos diferentes (*ferrolán, soldado, médico*), tres complementos (*capitán, cabalo, cómico*) e 3 modificadores (*folgazán, pequeno, rápido*) que se combinan entre si formando 126 frases diferentes (63 enunciativas e 63 interrogativas):

<i>O ferrolán</i>	<i>folgazán</i>		<i>co capitán</i>	<i>folgazán</i>
<i>O soldado</i>	+ <i>pequeno</i>	+ <i>falaba</i>	+ <i>co cabalo</i>	+ <i>pequeno</i>
<i>O médico</i>	<i>rápido</i>		<i>co cómico</i>	<i>rápido</i>

No entanto, antes da realización deste corpus fixo definitivo, cuxos resultados integrarán o Atlas Multimedia, o grupo galego do AMPER elaborou previamente un *corpus fixo experimental* co que se pretendían observar os problemas que inevitablemente aparecen na práctica. Con este corpus previo fixéronse

gravacións a unha informante de Santiago de Compostela, analizadas e descritas case na súa totalidade en Fernández Rei et al. (2005a e 2005b) e coas que se fixeron test de percepción en Martínez Celdrán et al. (2006 e 2007). Gravouse tamén unha informante do concello de Goián, analizada parcialmente en Fernández e Moutinho (2006). Este corpus experimental tivo que ser modificado por varias razóns:

- Algunhas das formas léxicas utilizadas presentaban variación dialectal no timbre das vogais (p.ex. *déspota* e *cadelo*).
- Decidiuse evitar a vogal palatal <i>, especialmente na sílaba tónica. A duración intrínseca desta vogal é menor con respecto ás outras.
- Evitar variacións de xénero do tipo: *tatexo/tatexa*.

b) No **corpus inducido**, o entrevistador contextualiza unha determinada situación cotiá na que o informante se debe situar para facer unha pregunta, respondela ou expresar un estado de ánimo determinado tal e como faría nunha situación natural. Con este cuestionario trátase de conseguir a entoación máis espontánea e cotiá do informante, posto que as respostas non seguen un padrón sintáctico e dependen da propia elección (anexo 1).

c) No chamado **Map Task**³, preséntanse dous mapas dunha mesma zona, inventada ou non. Nun deles hai unha ruta marcada, co seu inicio e a súa meta, que no outro non aparece (este último é o que se lle dá ao informante). Esta ruta pasa por unha serie de lugares que poden ou non coincidir nos dous mapas. Un mesmo lugar incluso pode ter diferentes nomes nos dous mapas. O informante debe debuxar o percorrido da ruta que se lle describe oralmente, facendo as preguntas necesarias. O obxectivo deste exercicio é a recollida das expresións xurdidas do diá-

³ Basado no HCRC Map Task Corpus. Para máis información consultar a web: <http://www.hcrc.ed.ac.uk/maptask/>

logo, no que abundan os equívocos, rectificacións, aclaracións, preguntas, etc. (anexo 2).

d) O **corpus libre** é un monólogo libre do informante sobre calquera tema, sen apenas intervención do entrevistador.

Para o Atlas Multimedia, nunha fase inicial, unicamente serán incluídos os datos relativos ás oracións do corpus fixo, posto que son as únicas que permiten facer unha comparativa prosódica entre as distintas linguas románicas. O resto das gravacións utilízanse para cualificar o grao de naturalidade das gravacións realizadas co corpus fixo e quedan nunha base de datos do proxecto para futuras investigacións.

Cuestionario utilizado neste estudo

A nosa investigación realizouse a partir das frases gravadas co corpus fixo, xa na súa versión definitiva (anexo 1).

Analizamos, por tanto, 126 contextos diferentes, dos cales 18 enunciados (9 enunciativas e 9 interrogativas) non presentan extensión ou modificador nin no suxeito nin no complemento. Un total de 54 enunciados (27 enunciativas e 27 interrogativas) posúen un adxectivo como expansión do suxeito. E finalmente, 54 enunciados (27 enunciativas e 27 interrogativas) presentan modificador no complemento.

3. Gravacións

As gravacións fixéronse a unha muller de 46 anos natural da parroquia de Camelle no concello de Camariñas (A Coruña), lugar no que residiu sempre⁴. Para unha mellor calidade do son, empregamos unha gravadora dixital portátil -DAT-, *Sony TCD-D10 PRO II* e un micrófono *Shure* de cabeza, externo e unidireccional, propiedade do laboratorio de fonética do Instituto da Lingua Galega (ILG) da Universidade de Santiago de Compos-

⁴ As gravacións fixéronse no ano 2005.

tela. As gravacións realizáronse nun cuarto da vivenda da informante, co que se pretende favorecer a espontaneidade e naturalidade na entoación, procurando evitar calquera tipo de ruído ambiental.

Para o corpus fixo realizáronse 3 repeticións de cada un dos enunciados coa finalidade de calcular unha media, máis representativa do que unha única mostra. As fichas de debuxos presentáronse por separado e en orde aleatoria.

4. Metodoloxía

Unha vez feitas as gravacións procedeuse á dixitalización do sinal a través do programa *GoldWave* 4.25. Para cada enunciado créase un arquivo *wav* monofónico cun rango de mostra de 16000 Hz. Co mesmo programa, límpase o sinal de ruídos, normalízase a intensidade e prepárase para a posterior análise.

Tendo en conta a estrutura sintáctica da oración e as posicións acentuais, asígnaselle a cada arquivo de son un código, que é común para todas as linguas integradas no AMPER. Así, por exemplo, o arquivo que contén *kwti*⁵ na súa denominación, permítennos saber que a oración é unha interrogativa coa estrutura: *suxeito* (*sen extensión*) + *verbo* + *complemento* (*sen extensión*). A este código engádeselle ao inicio catro díxitos que permiten saber o lugar da gravación e as características dos informantes (sexo, nivel cultural, etc); así o código *oK11* indica que os arquivos foron producidos por unha muller con estudos primarios natural da vila de Camelle.

A continuación, segméntanse as vogais de cada oración a partir dun oscilograma e dun espectrograma xerado polo pro-

⁵ Este código refírese, por exemplo, á frase "El saxofón se toca con paciencia" no corpus do español, "El capità no porta la caputxa" no catalán, "El rapacín xugaba n'equipu" no asturiano, "O capataz toca no Toneca" no portugués, "La buesunà tosta la tsabreta" no occitano ou "O ferrolán falaba co cabalo" no galego.

grama *MatLab* 6.5.⁶, coas que extraemos os datos numéricos relativos á frecuencia fundamental, duración e intensidade. Esta información é recollida en ficheiros de texto (con extensión *.txt*) cos que se poden calcular medias e xerar gráficos. Este mesmo programa pode xerar unha síntese dos ficheiros de son, eliminando o contido léxico e mantendo unicamente as características prosódicas das oracións. Estes arquivos permiten comprobar auditivamente a fiabilidade dos datos do programa, ademais da realización de tests de percepción.

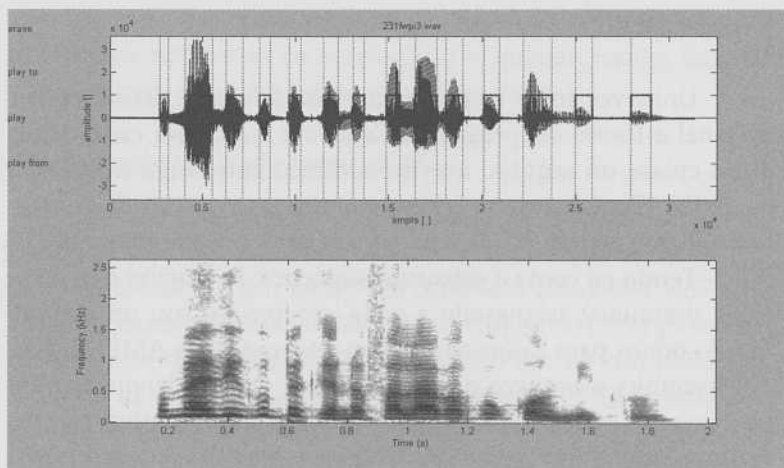


Figura 2: Segmentación de vogais en O médico pequeno falaba co cómico (MatLab)

Con *MatLab* calculamos a media aritmética das tres repetición realizadas pola informante para cada unha das 126 oracións analizadas neste traballo (ver anexo1). Para esta operación tense en conta a frecuencia fundamental media da informante (205 Hz), que determinamos co propio programa, e que nos permite estudar a curva melódica dentro do rexistro individual da falante.

⁶ Nesta versión desenvolvéronse aplicacións específicas para o proxecto feitas polo profesor Antonio Romano da Università di Torino e membro do proxecto.

Finalmente, nunha folla de cálculo *Excel* fixemos (diferenciando o tipo de acentuación –oxítona, paroxítona e proparoquítona-) a media de todas as realizacións do suxeito (*ferrolán, soldado e médico*), do complemento (*capitán, cabalo e cómico*), das extensións do suxeito (*folgazán, pequeno e rápido*) e da forma verbal *falaba* (aínda que os datos do verbo non serán considerados neste traballo). Os gráficos que mostramos nas seguintes epígrafes foron realizados en *Excel*, nos que se dá unha medida por vogal.

5. Frecuencia Fundamental

Pasando xa aos resultados logrados co método de análise referido anteriormente, describiremos o comportamento da F0 nas dúas modalidades oracionais tratadas no proxecto: enunciativas neutras ou non marcadas e interrogativas de tipo absoluto, que trataremos en epígrafes separadas. Foron desbotadas do cálculo das medias xerais todas aquelas pronunciacións que non se encadraban nestas características e presentaban matices expresivos ou certa intencionalidade.

Enunciativas

Suxeito e complemento simples

En canto á relación que se establece entre a posición do acento e o contorno de entoación, o comportamento do suxeito oxítono é semellante ao do suxeito proparoquítono. Por un lado, a frecuencia fundamental destes suxeitos realiza un único pico tonal asociado á sílaba que leva o acento léxico da palabra, posto que é nesta sílaba tónica onde xeralmente comeza un ascenso de frecuencia. Por outro lado, o máximo tonal nos dous tipos de suxeitos alcázase na primeira sílaba átona que segue á tónica. No suxeito oxítono (sen postónicas), é a sílaba pretónica da forma verbal a que continúa o ascenso iniciado pola tónica.

No entanto, no suxeito paroxítono imos encontrar unha diferenza importante con respecto a esta conduta descrita: na

sílaba tónica prodúcese unha inflexión tonal descendente (é dicir, un val tonal e non un pico tonal coma nos outros suxeitos), que non inflúe na colocación do máximo tonal do suxeito sobre a súa postónica.

No gráfico 1 non comparamos exemplos concretos da informante, senón os contornos resultantes do cálculo dunha serie de medias segundo cada aparición dos tres suxeitos simples estudados (diferentes pola posición do acento): *ferrolán*, *médico*, *soldado* (véxase a epígrafe 4, sobre a metodoloxía seguida).

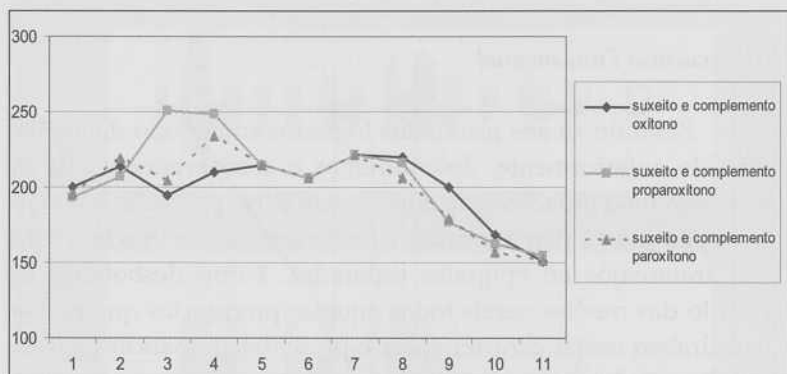


Gráfico 1: Media dos suxeitos e dos complementos simples

O contorno melódico do suxeito oxítono e do paroxítono é moi semellante. As principais diferenzas parecen atoparse na posición da primeira sílaba con respecto á terceira (máis elevada no suxeito oxítono) e na elevación máis ou menos pronunciada da cuarta sílaba, que se corresponde coa tónica no suxeito oxítono e coa postónica no suxeito paroxítono. Observámolo nas seguintes figuras, nas que mostramos a media final do suxeito oxítono (gráfico 2) e do suxeito paroxítono (gráfico 3)⁷ e, para que se aprecie mellor o ascenso total do suxeito, trazamos unha liña dende a primeira ata a última sílaba do suxeito.

⁷ Nos gráficos 2 e 3, os valores das formas verbais non se corresponden a unha media xeral de todas as realizacións de *falaba* coma no resto dos gráficos deste estudo, senón que se fixo a media aritmética en relación as súas aparicións cos suxeitos (*O ferrolán falaba* ou *O soldado falaba*).

Os maiores incrementos de frecuencia parecen ir asociados sempre ás postónicas nos suxeitos simples: na única postónica do suxeito paroxítono e na postónica medial no proparoxítono. O ascenso final é maior no suxeito paroxítono que no oxítono, pola simple razón de que este acaba en sílaba acentuada e non hai postónica. A única que podería continuar o ascenso de F0, en coherencia cos outros suxeitos, é a primeira sílaba átona da forma verbal. Os datos mostran que isto ocorre efectivamente, aínda que a porcentaxe de ascenso non é tan acusada como nos exemplos paroxítonos e proparoxítonos.

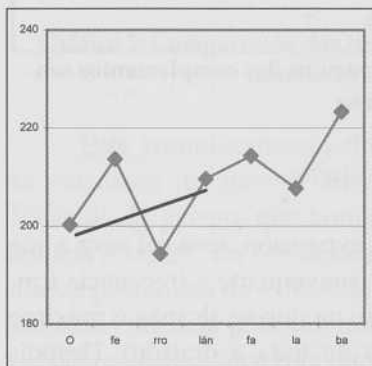


Gráfico 2: F0 correspondente a O *ferrolán falaba...*

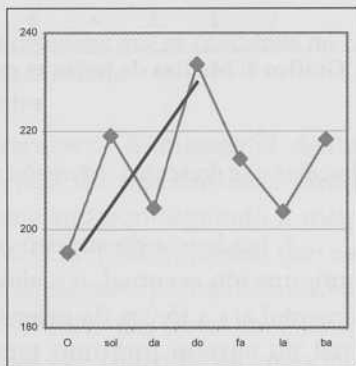


Gráfico 3: F0 correspondente a O *soldado falaba...*

Pasando á cadencia final do complemento sen extensión, non podemos afirmar que a configuración acentual do seu núcleo inflúa no maior ou menor descenso da frecuencia. A tendencia declinatoria final prodúcese dun xeito progresivo e suave ata alcanzar a base tonal da informante, próxima aos 150 Hz. Podémoslo observar no gráfico 4, onde mostramos a media dos tres finais (*capitán*, *cabalo*, *cómico*) en todas as oracións enunciati-vas analizadas, con ou sen extensión no suxeito.

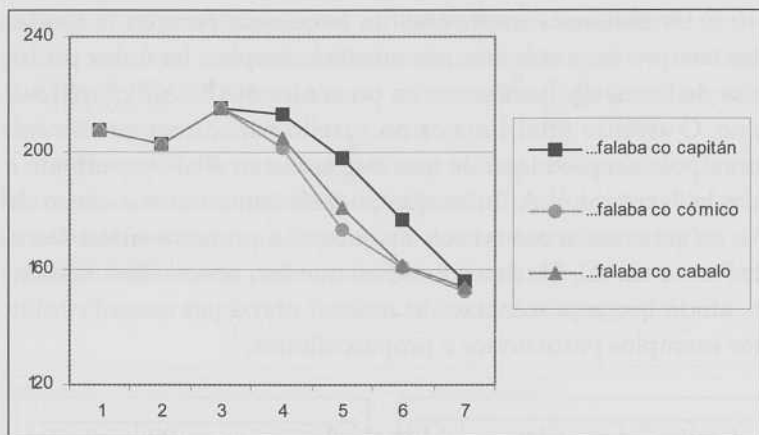


Gráfico 4: Medias de todas as realización dos complementos sen extensión

Oracións que presentan extensión no suxeito

A tendencia do suxeito con expansión, sexa cal sexa a súa configuración acentual, é a elevar suavemente a frecuencia fundamental ata a tónica da extensión, na que se alcanza o máximo tonal do suxeito (máximo tamén de toda a oración). Despois desta sílaba acentuada do modificador, prodúcese unha caída importante da frecuencia, alcanzando un ton medio no que se mantén ata o complemento final. En extensións proparoxítonas, a diminución de F0 prodúcese na primeira postónica e en extensións oxítonas na pretónica da forma verbal.

Como exemplo mostramos o gráfico 5 no que incluímos as oracións que presentan no suxeito un mesmo núcleo oxítono (ferrolán) combinado coas tres extensións que difiren na colocación do acento.

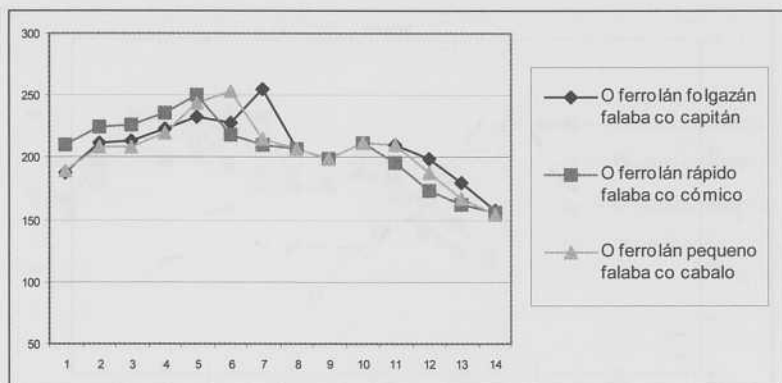


Gráfico 5: Comparación das tres extensións que se combinan no suxeito co núcleo oxítono

Este comportamento da frecuencia fundamental dentro da extensión do suxeito diferénciase do descrito para outros dialectos do galego, que tamén se estudaron seguindo a metodoloxía AMPER. En Fernández Rei et al. (2005b) dicíase que nas sílabas postónicas da extensión alcanzábase o punto máis alto da ascensión iniciada pola súa tónica. Dentro das nosas análises rexistramos realizacións concretas⁸ nas que o suxeito da enunciati- va segue este comportamento das postónicas descrito en Fer- nández Rei et al. (2005b), en contraposición ao descrito nesta epígrafe. É o caso dos exemplos mostrados nos seguintes gráfi- cos feitos co programa MatLab⁹.

⁸ Estes exemplos foron descartados no cálculo das medias, a pesar de que se entendían como realizacións neutras, porque constituían realizacións mi- noritarias dentro da totalidade das mostras. A dificultade na gravación de ora- cións afirmativas radica na ampla variedade que temos de enunciar un suceso ou acción sen utilizar a énfase como recurso, polo que a escolla das enunciati- vas sempre se fai a partir da realización máis frecuente dentro da zona xeográfica.

⁹ Este programa achega tres valores de frecuencia por cada vogal. A curva melódica dos exemplos que extraemos do programa MatLab foron some- tidos a un certo grao de estilización, para solventar problemas de estimación da frecuencia especialmente na primeira e na última sílabas.

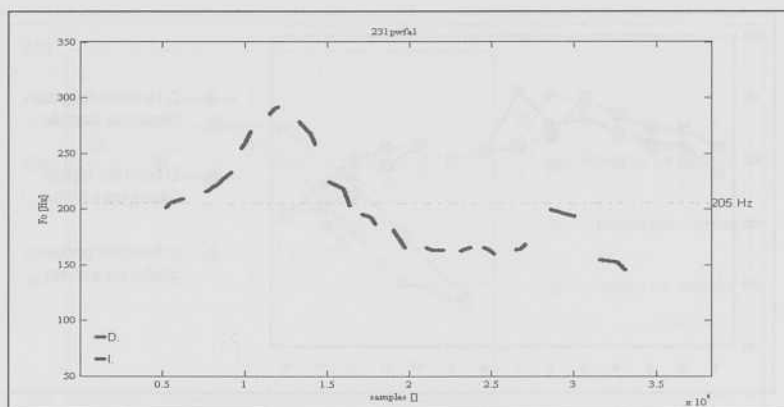


Gráfico 6: Configuración acentual nunha das repeticións de *O médico falaba co cómico pequeno*.

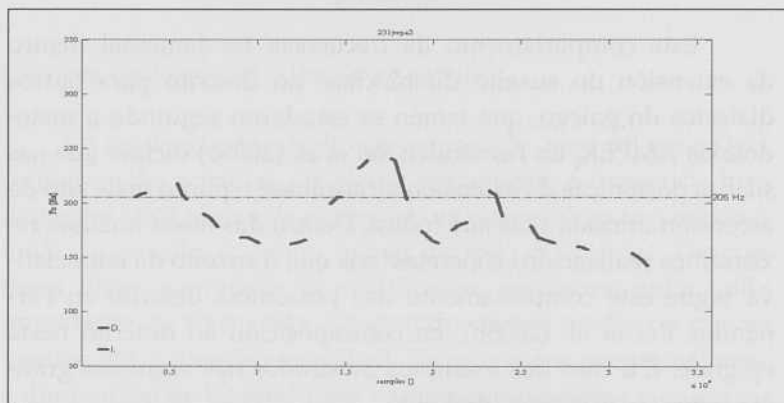


Gráfico 7: Configuración acentual nunha das repeticións de *O ferrolán rápido falaba co cómico*.

Por outra banda, se comparamos as oracións con e sen extensión (gráfico 8), decatámonos de que o máximo tonal nas oracións simples sitúase sobre as sílabas postónicas do suxeito; pola contra, nas oracións con modificador o pico tonal máis elevado realízase na extensión e sempre asociado á súa sílaba tónica.

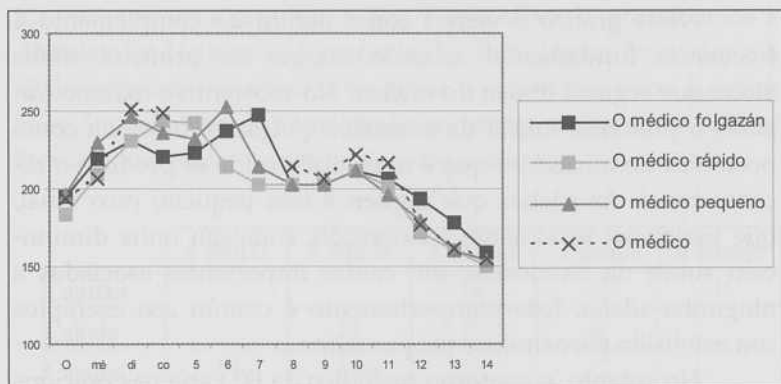


Gráfico 8: Núcleo proparoxítono con e sen extensión

Oracións que presentan extensión no complemento

As oracións que presentan extensión no complemento teñen un comportamento final peculiar, que depende da posición do acento dentro do substantivo que funciona como núcleo de frase. Observámolo con claridade no gráfico 9, no que mostramos as medias das oracións que combinan a extensión oxítónica cos tres núcleos estudados (*capitán*, *cómico*, *cabalo*) no complemento. O suxeito dos tres rexistros é oxítono (*O ferrolán...*).

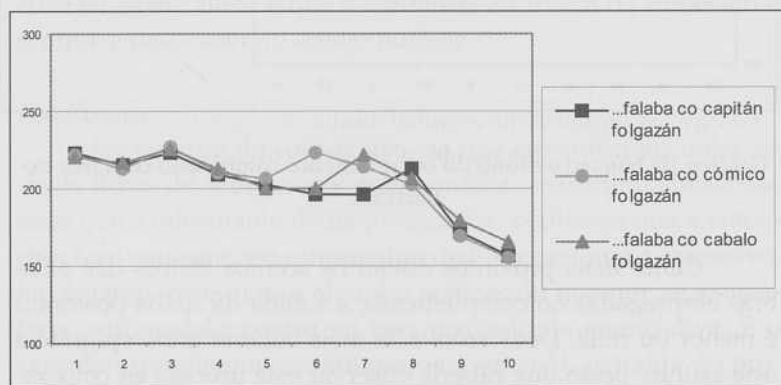


Gráfico 9: Os tres núcleos do complemento combinados coa extensión oxítónica.

Neste gráfico 9 vemos como dentro do complemento a frecuencia fundamental ascende sempre na primeira sílaba átona que segue á tónica do núcleo. No substantivo oxítono (*capitán*) a pretónica inicial da extensión (*folgazán*) funciona como postónica do núcleo, xa que é nesa sílaba onde se produce o ascenso tonal. As sílabas que seguen a este pequeno pico tonal, que pertencen xeralmente á extensión, comezan unha diminución suave da frecuencia, sen caídas importantes asociadas a ningunha sílaba. Este comportamento é común aos exemplos con extensión paroxítona e proparoxítona.

No entanto, o contorno melódico da F0 varía nas oracións coa combinación *núcleo oxítono + extensión proparoxítona*, é dicir, nos exemplos en que aparecen dous acentos léxicos contiguos (gráfico 10).

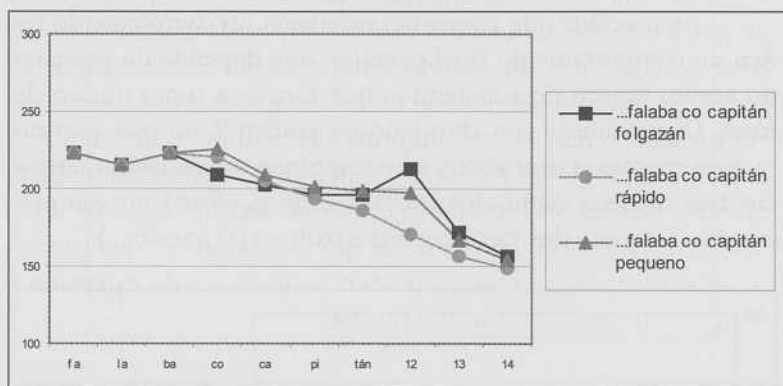


Gráfico 10: Núcleo oxítono do complemento combinado coas tres extensións

Canto máis próximos estean os acentos léxicos das palabras empregadas no complemento, a subida da sílaba postónica é menor ou nula. Dedicáremoslle máis adiante unha epígrafe a este asunto, posto que cabería observar este proceso en conxunción coa duración.

Para comprobar en que medida o número de sílabas átonas entre os dous acentos do complemento inflúe na maior ou menor elevación da sílaba postónica, como acabamos de descri-

bir, elaboramos a seguinte táboa. Nesta táboa 1, presentamos a distancia en Hz entre a frecuencia da sílaba tónica do núcleo e a sílaba seguinte, sexa tónica ou átona e pertenza ou non á extensión. Dispoñemos os datos tendo en conta o número de sílabas átonas entre os dous acentos do complemento.

	4 átonas	3 átonas	2 átonas	1 átona	0 átonas
Capitán			14	-2	-16
Cabalo		22	23	6	
Cómico	17	16	13		

Táboa 1: Distancia en Hz dende a tónica do núcleo ata a sílaba seguinte -tónica ou átona-.

Á vista dos datos anteriores podemos dicir que a subida da frecuencia fundamental é maior canto máis afastados se encontran os acentos do complemento e menor canto máis achegados. No único exemplo que temos en que os dous acentos están pegados e non existe ningunha sílaba átona no medio, non se produce ningunha ascensión, senón unha diminución de frecuencia. Este mesmo exemplo cos dous acentos seguidos (...*co capitán rápido*) permítenos afirmar que o ascenso ten que producirse en sílaba átona e que a tendencia da tónica da extensión é sempre a descender no acento nuclear.

Focalización

Na enquisa do corpus fixo, na que mostrabamos unha serie de fichas de debuxos que representaban estruturas sintácticas fixas que a informante debía pronunciar, pediamos que a entoación fose natural e sen ningún tipo de marca ou matiz expresivo. No entanto, rexistramos algunha realización que non se axustaba a este modelo neutro ou non marcado da enunciativa. É o caso do exemplo que mostramos na gráfico 11, extraída do programa MatLab, que representa unha das repeticións feitas pola informante da oración *O ferrolán pequeno falaba co cómico*. A audición destes exemplos indícanos que se está a producir unha focalización no complemento da oración.

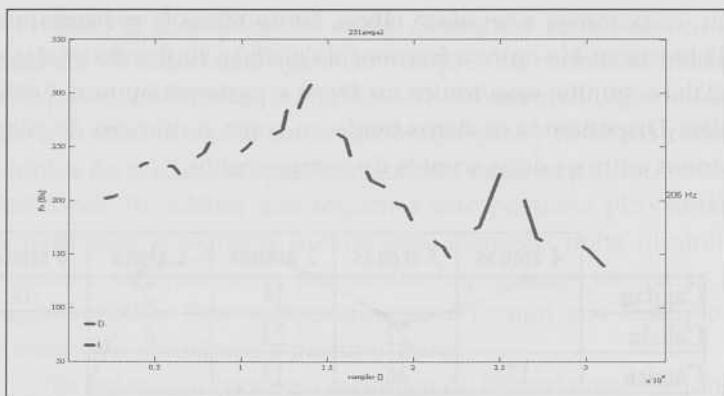


Gráfico 11: Configuración acentual da 1ª repetición de *O ferrolán pequeno falaba co cómico*¹⁰.

Se nos fixamos no final do exemplo da gráfico 11, podemos observar que, en contraposición co descrito para as oracións neutras (véxase gráfico 4 da epígrafe 6.1.1), realízase un último pico tonal definido sobre a tónica do complemento. Esta realización tonal do foco en posición final de oración non é a mesma que describiamos en Escourido (2007), na que achegabamos unha serie de trazos comúns sobre un corpus deseñado para o estudo deste fenómeno. Neste traballo dicíamos que as oracións con foco, tanto e posición inicial como en posición final, presentaban un pico tonal asociado ás primeiras sílabas pretónicas do constituínte focalizado:

Este tipo de realizacións comeza nunha frecuencia elevada que se mantén sen movementos tonais relevantes ata o final. No complemento prodúcese un pequeno ascenso nas sílabas pretónicas e posteriormente sobre a tónica ten lugar unha caída abrupta alcanzando o baixo ton final da oración” (Escourido 2007: 265).

¹⁰ A realización do suxeito, que non segue o modelo maioritario descrito ata agora para o suxeito con modificador, dado que a postónica da extensión continúa o ascenso das sílabas anteriores, alcanzando o máximo tonal da oración. Este comportamento máis achegado ao descrito en Fernández Rei et al. (2005b), sobre o estudo das oracións con extensión no suxeito dunha fala feminina de Santiago de Compostela.

A partir dos datos desta comunicación e para que se poida comparar coa realización da gráfico 11, mostramos o gráfico 12, no que aparece un exemplo de focalización no substantivo paroxítono final da oración.

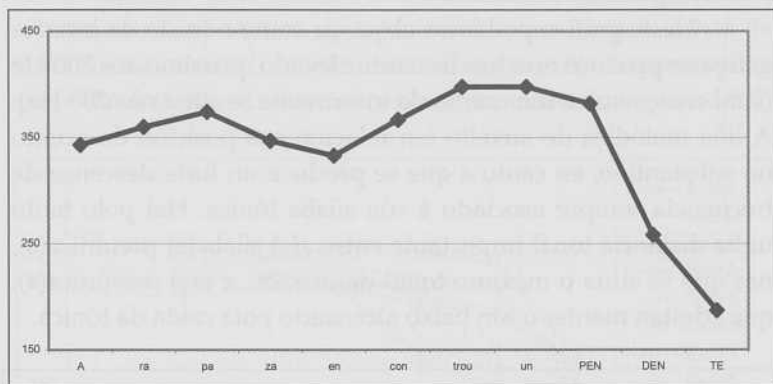


Gráfico 12: Oración con foco no constituínte final da oración.
Exemplo tomado de Escourido (2007)

Nunha realización como a descrita neste artigo sobre o foco, o pico tonal debería aparecer sobre a sílaba átona anterior á tónica (sobre a contracción "co" de ...*falaba co cómico*), en cambio, neste exemplo da gráfico 11, a elevación de frecuencia relaciónase claramente coa tónica. Debemos pensar talvez que son diferentes formas de interpretación dos enunciados e do obxecto de focalización, posto que a metodoloxía non nos permite controlar este tipo de variables contextuais: nun dos modelos interpretaríase unicamente o substantivo final da oración como elemento focalizado, e no outro modelo o foco recae no conxunto da frase nominal que conforma o complemento.

Interrogativas

Suxeito e complemento simples

Para o estudo das interrogativas sen ningún tipo de modificador nin no suxeito nin no complemento, calculamos a partir

das realizacións reais das oracións (véxase epígrafe 4) unha serie de medias de todas as aparicións de cada suxeito e de cada complemento. Finalmente co resultado elaboramos o gráfico 13, no que mostramos as tres medias das interrogativas segundo a súa configuración acentual.

Neste gráfico podemos observar como o inicio da interrogativa se produce nun ton bastante elevado, próximo aos 300 Hz (lembramos que o ton medio da informante se sitúa nos 205 Hz). A liña melódica do suxeito ten relación coa posición do acento no substantivo, en canto a que se produce un forte descenso de frecuencia sempre asociado á súa sílaba tónica. Hai polo tanto unha distancia tonal importante entre a(s) sílaba(s) pretónica(s), nas que se sitúa o máximo tonal da oración, e a(s) postónica(s), que adoitan manter o ton baixo alcanzado pola caída da tónica.

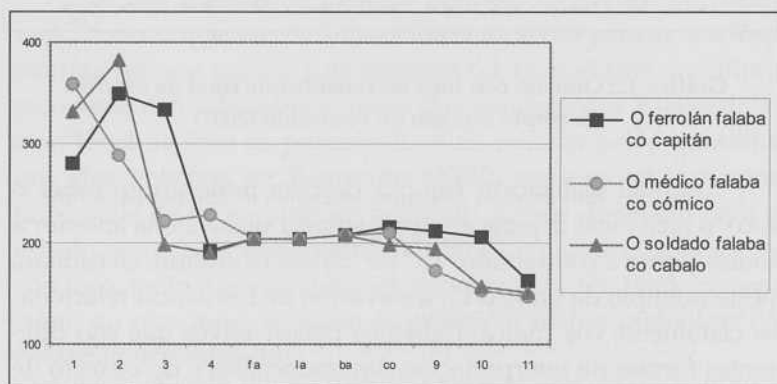


Gráfico 13: Comparación dos tres suxeitos na modalidade interrogativa.

Dentro do complemento rexistramos tamén un pequeno descenso de ton que contrasta coa inmovilidade do predicado. Esta diminución de F0 é menos significativa que a que se produce dentro do suxeito, pero parece ter relación de igual xeito coa configuración acentual do substantivo empregado. Elaboramos un gráfico, que leva o número 11, no que mostramos todas as medias dos enunciados con complemento simple. Para que fora máis claro empregamos tres tipos de tramas segundo a posición

acentual no substantivo: as oracións con complemento proparoxítono levan o gris claro, as que teñen complemento paroxítono a trama con raías descontinuas e as oracións con complemento oxítono o gris escuro. Neste gráfico, efectivamente, observamos como hai unha certa relación entre a caída tonal e o acento do substantivo. Con todo, a agrupación de cores (e polo tanto de oracións cun mesmo complemento) é moito máis dispersa que a que viamos dentro do suxeito (véxase por exemplo o gráfico 12).

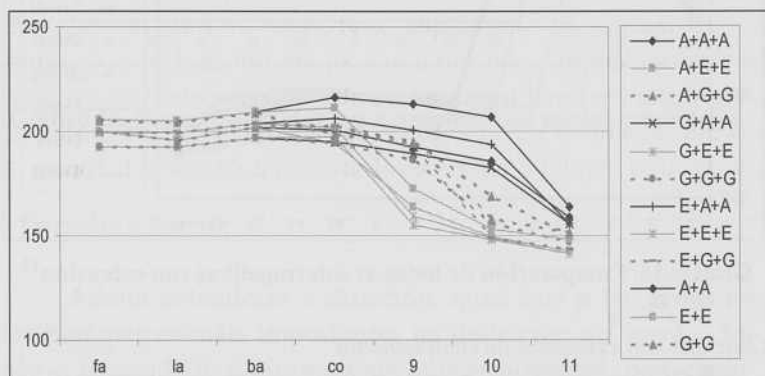


Gráfico 14: Comparación dos complementos de todas as oracións interrogativas estudadas¹¹.

Oracións con extensión no suxeito

Esta caída tonal abrupta descrita para as oracións simples, prodúcese no segundo acento do suxeito nas oracións con extensión, é dicir, na sílaba tónica do adxectivo modificador. No gráfico 15, onde comparamos todas as interrogativas con extensión, empregamos unicamente tres cores para sinalar a posición do acento no modificador: gris claro, gris medio e gris escuro, que representan respectivamente as oracións con extensión proparoxítona, con extensión paroxítona e con extensión oxítona no suxeito. Así, podemos comprobar nun golpe de vista como se agrupa esta descenso en picado do suxeito ao acento do adxectivo modificador. En primeiro lugar, aparece a caída das interro-

¹¹ Na lenda dos gráficos 14 e 15: E significa esdrúxulo, G grave e A agudo.

gativas con extensión proparoxítona na 5ª sílaba, o resto das oracións continúan no mesmo ton ata a 6ª sílaba na que ten lugar a baixada da extensión paroxítona e en último lugar, na 7ª sílaba, aparece a caída da extensión oxítona.

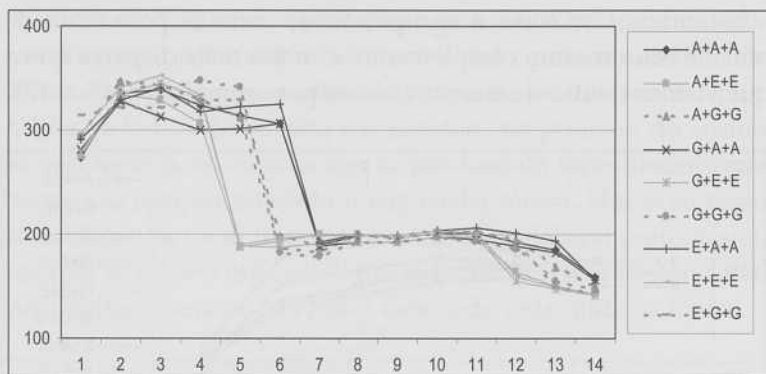


Gráfico 15: Comparación de todas as interrogativas con extensión¹²

Oracións con extensión no complemento

Con respecto ao complemento simple, as interrogativas con extensión final manteñen durante máis tempo o ton baixo do predicado (en torno aos 200 Hz). O núcleo tende a manter en suspensión a frecuencia na altura marcada polo suxeito e a partir da extensión iníciase un descenso progresivo cara ao final, sen ter relevancia aparentemente a posición acentual do modificador (gráfico 16).

¹² Na lenda: E significa esdrúxula, G grave e A aguda.

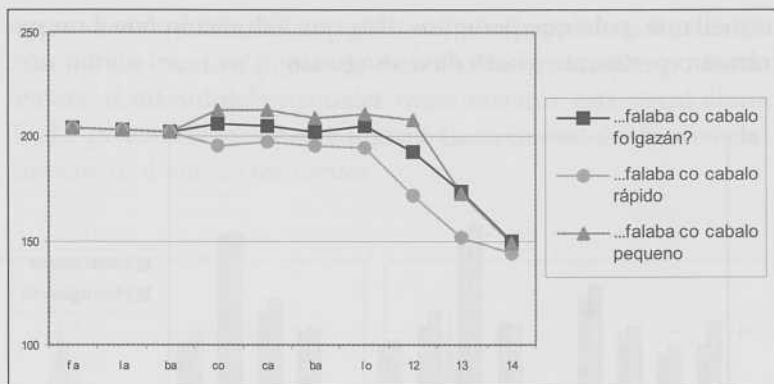


Gráfico 16: Comparación das tres extensións con núcleo paroxítono

7. Duración e Enerxía

Adoita entenderse a duración, igual que a F0, como un dos parámetros máis importantes na definición do acento. Tamén a intensidade é un dos seus índices acústicos, pero xeralmente no galego non parece seguir un comportamento unitario. Si podemos intuír tanto na duración como na enerxía a súa asociación aos acentos da oración.

Trataremos de describir a continuación cal é o comportamento xeral destes dous índices acústicos e que relación gardan coas variacións melódicas sinaladas.

Duración

Na totalidade do noso corpus, tanto nas oracións con extensión como nas simples, observamos como as sílabas de maior duración son sempre as que levan o acento léxico da palabra. Nos seguintes gráficos mostramos a duración dos suxeitos (gráfico 17) e dos complementos (gráfico 18), nas oracións sen extensión, tendo en conta a súa configuración acentual. Diferenciamos tamén con distintos grises as enunciativas e as interrogativas. O comportamento de ambas modalidades oracionais é moi

semellante, polo que podemos dicir que a duración non é un parámetro pertinente á hora de distinguilas.

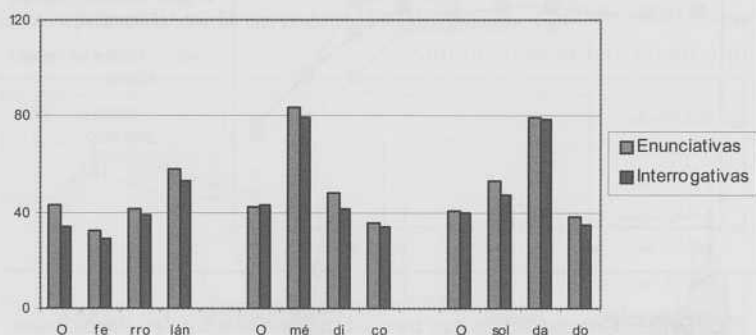


Gráfico 17: Duración dos suxeitos nas oracións sen extensión

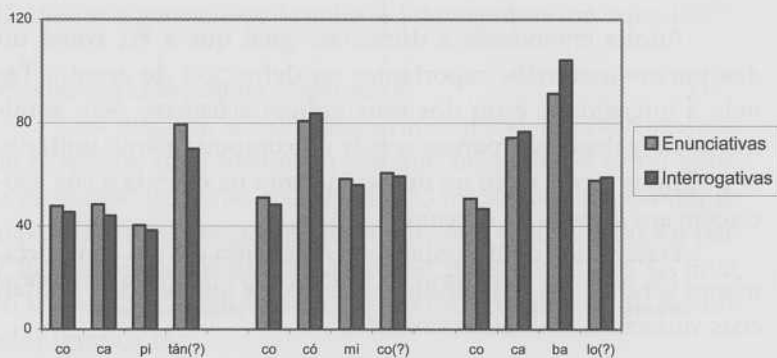


Gráfico 18: Duración dos complementos nas oracións sen extensión

O comportamento das átonas depende da súa función silábica e oracional. En xeral, as sílabas pretónicas adoitan ser máis longas conforme se van achegando á tónica, mentres que as postónicas diminúen a duración en relación á súa proximidade coa tónica.

Por outro lado, a duración da vogal final ante pausa debe ser interpretada con cautela, posto que con frecuencia non é doada de segmentar. Habitualmente os falantes galegos tenden

a pronunciala como unha vogal xorda, máis próxima á articulación dunha fricativa (figura 3) e é difícil saber en que momento remata. É máis doado naqueles casos nos que esta vogal átona final é pronunciada con voz (figura 4), na que se distinguen claramente os distintos formantes.

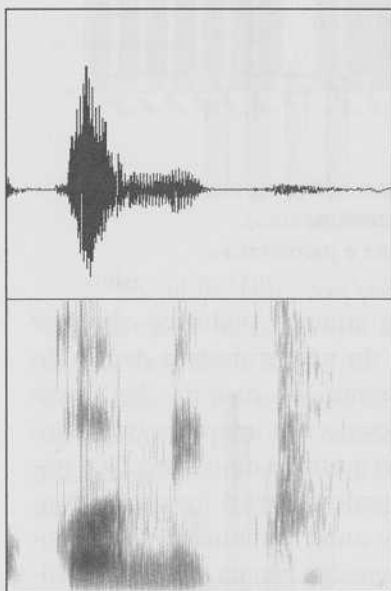


Figura 3: Oscilograma e espectrograma de cómico

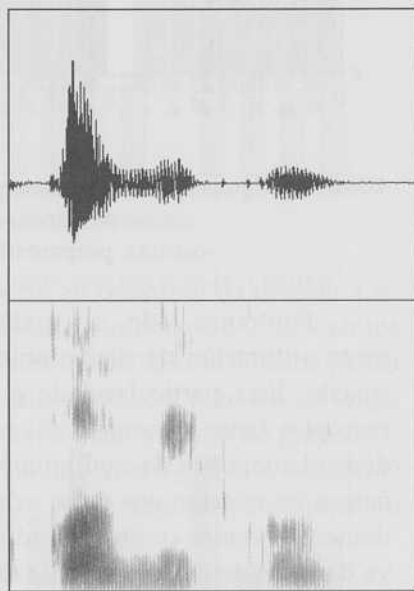


Figura 4: Oscilograma e espectrograma de cómico

Vexamos agora o comportamento das extensións dentro do suxeito e dentro do complemento. Como xa indicamos, as sílabas tónicas do suxeito con extensión na maioría dos casos teñen maior duración ca as átonas. Poñemos como exemplo disto o gráfico 19, no que mostramos as tres extensións (con distinta colocación do acento) combinadas con suxeito paroxítono, na modalidade afirmativa e interrogativa.

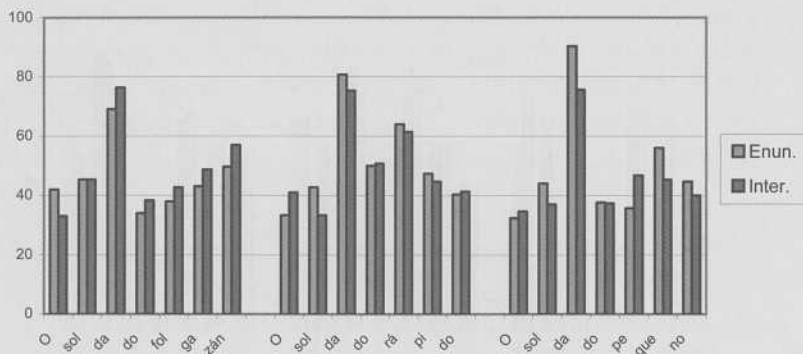


Gráfico 19: Duración dos suxeitos con núcleo paroxítono combinados coas tres extensións -oxítona, proparoxítona e paroxítona-.

Por outro lado, no gráfico anterior podemos observar como a duración da sílaba tónica do núcleo destaca dentro do suxeito. Esta particularidade é común na maioría dos nosos exemplos tanto nas enunciativas como nas interrogativas e independentemente da configuración acentual do núcleo. Se o poñemos en relación aos datos achegados sobre a frecuencia fundamental, vemos como se produce unha diminución significativa de F0 asociada ao acento da extensión, ben na súa sílaba tónica, ben na súa primeira postónica (ver epígrafe 6.1.2 para as enunciativas e epígrafe 6.2.2 para as interrogativas); polo que a duración parece ser a encargada de sinalar o acento da sílaba tónica do núcleo do suxeito, onde o pico da F0 non é o máis destacado do suxeito.

Canto ao complemento con extensión, tamén destaca a duración das súas dúas sílabas con acento léxico. A sílaba tónica do modificador xeralmente é a máis longa do complemento, aínda que a diferenza de duración entre as tónicas non é tan grande coma no suxeito. Exemplificamos isto no gráfico 20, onde mostramos os complementos con núcleo paroxítono combinado coas tres extensións con variación na configuración acentual.

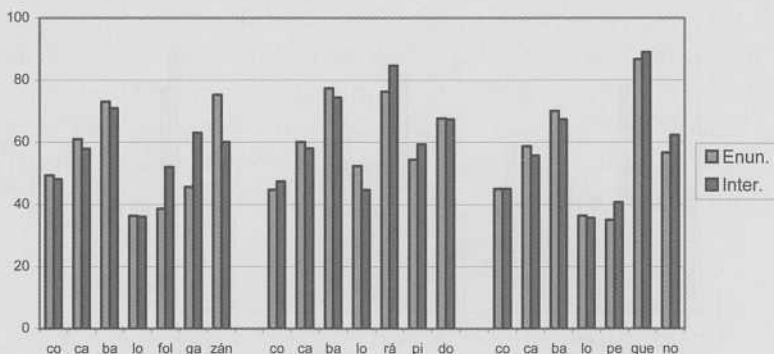


Gráfico 20: Duración dos complementos con núcleo paroxítono combinados coas tres extensións

Por outro lado, con respecto ao conxunto da oración a sílaba de maior duración está asociada xeralmente á última tónica da oración. Nas oracións con extensión no suxeito a sílaba tónica do seu núcleo é a segunda de maior duración, pero a máis longa da oración atópase no complemento final. Vémolo no gráfico 21, no que mostramos como afirmativa e interrogativa a oración *O médico pequeno falaba co cabalo(?)*, con extensión paroxítona no suxeito.

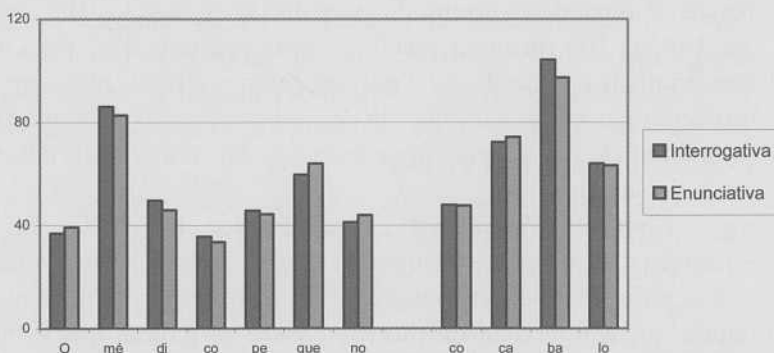


Gráfico 21: Duración de *O médico pequeno falaba co cabalo(?)*¹³

¹³ Para que os gráficos 20 e 21 foran máis ilustrativos, eliminamos os datos da duración da forma verbal e dividimos en dous grupos o suxeito e o predicado.

Nas oracións con extensión no complemento, vemos como a importancia do suxeito simple é clara e a miúdo a súa sílaba tónica compite en duración coa tónica da extensión do complemento. Observámolo no gráfico 22, no que representamos a oración *O soldado falaba co cabalo pequeno(?)* na modalidade afirmativa e interrogativa.

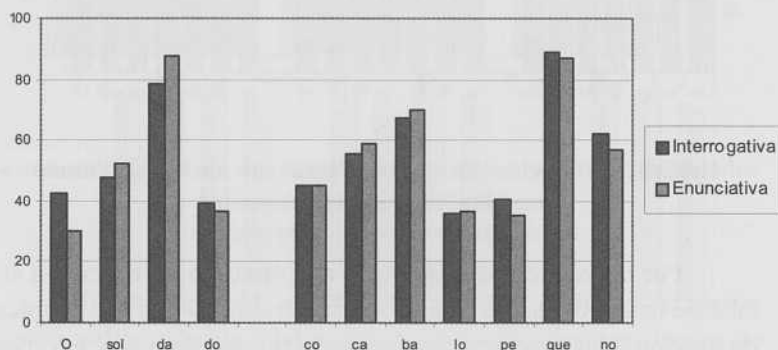


Gráfico 22: Duración de *O soldado falaba co cabalo pequeno(?)*

Enerxía

Á vista dos datos extraídos, a enerxía non segue un padrón regular como o descrito para a F0 ou a duración, por iso só podemos falar de tendencias xerais ou trazos comúns. No entanto, cabería comprobar mediante un test de percepción se estas pequenas variacións de intensidade descritas a continuación son realmente captadas polo oído humano e por tanto significativas no proceso comunicativo.

A primeira tendencia que podemos indicar, común ás dúas modalidades oracionais estudadas, é que no suxeito a sílaba de maior intensidade é xeralmente a tónica, tanto se hai extensión coma se non. Atopamos varias excepcións pero só en núcleos oxítonos, nos que a sílaba tónica perde intensidade en favor da súa segunda pretónica. É significativo que o único caso irregular se encontre no oxítono, posto que é o único suxeito onde a tónica aparece trabada por unha consoante. O que nos

leva a facer a seguinte pregunta: ¿cabería pensar que a nasal trabante inflúe dalgún xeito nos datos da enerxía?

No gráfico 23 mostramos os tres suxeitos, segundo a súa configuración acentual, nas dúas modalidades estudadas: afirmativa e interrogativa, que seguen un comportamento semellante.

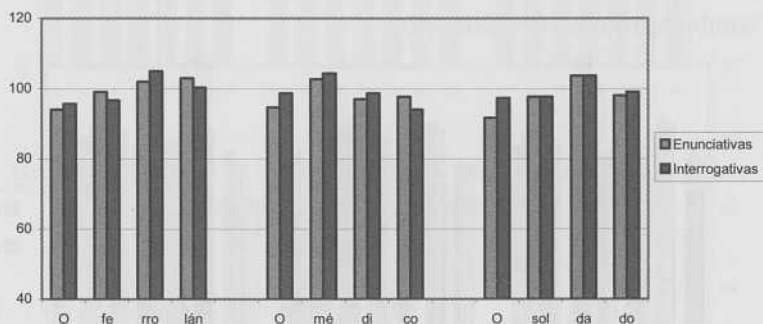


Gráfico 23: Enerxía en dB dos tres suxeitos simples

Unha segunda tendencia rexistrámola no complemento (gráfico 24), onde maioritariamente o comportamento é a perder enerxía nas dúas últimas sílabas da oración. O acento léxico da palabra parece influír na maior ou menor diminución final: a intensidade da penúltima sílaba é maior en *cómico* (sendo átona), que en *cabalo* (sendo tónica). No final oxítono, obsérvase como a sílaba tónica perde intensidade con respecto á antepenúltima, aínda que a súa enerxía sexa maior que a penúltima.

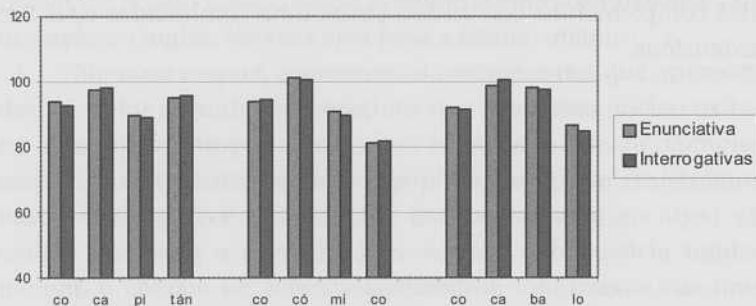


Gráfico 24: Enerxía en dB dos tres complementos simples

Por último, con respecto ás extensións, non hai grandes variacións de enerxía que nos permitan facer afirmacións categóricas ao respecto. Si observamos, no entanto, que tenden a mostrar o mesmo comportamento que as oracións simples. Dentro do suxeito, as sílabas tónicas teñen lixeiramente maior intensidade. Vémolo, por exemplo, no gráfico 25, no que mostramos a enerxía en *dB* dos suxeitos con núcleo proparoxítono combinados coas tres extensións.

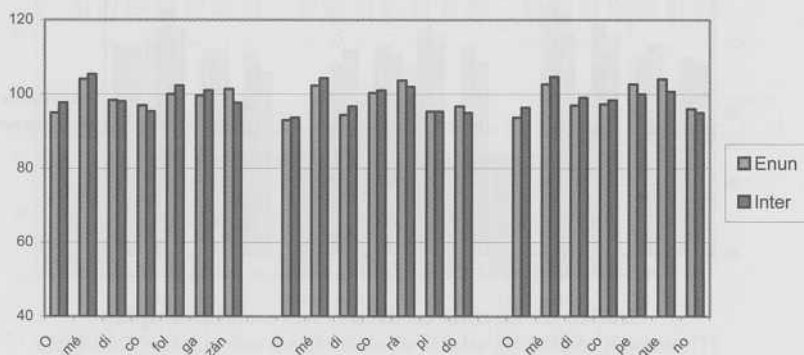


Gráfico 25: Enerxía en *dB* dos suxeitos con núcleo proparoxítono combinados coas tres extensións

E no complemento, as últimas sílabas adoitan diminuír a intensidade independentemente da posición do acento, do mesmo xeito que describiamos para as oracións sen extensión. Obsérvese isto no gráfico 26, no que mostramos a enerxía en *dB* dos complementos con núcleo paroxítono combinados coas tres extensións.

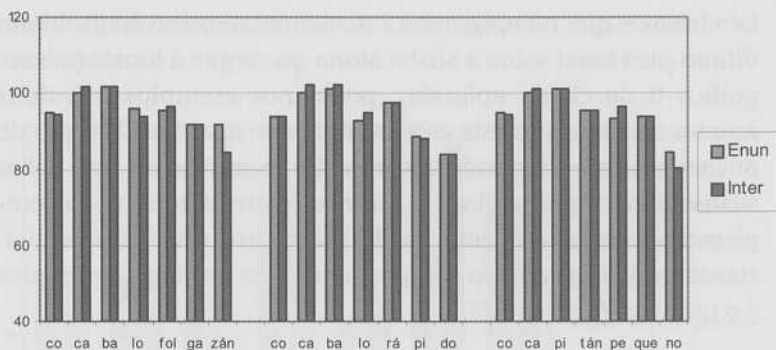


Gráfico 26: Enerxía en dB dos complementos con núcleo paroxítono combinados coas tres extensións

8. Oracións con dous acentos seguidos

Segundo se postula dende a escola da Fonoloxía Métrica (Selkirk 1984), as linguas de todo o mundo, na súa organización rítmica, prefiren contextos con alternancia métrica entre sílabas en posición forte (F) e sílabas en posición débil (D). Así, as secuencias de sílabas con dous ou tres acentos consecutivos séntense como pouco naturais e tenden a ser evitadas. Son varios os procedementos prosódicos rexistrados para solucionar un *clash* acentual ou choque de acentos, o máis documentado talvez sexa o “desprazamento acentual” (*stress shift*), que consiste na re colocación do acento léxico nunha posición anterior na palabra, por exemplo no inglés: *thirteen men* pasa a *thirteen men*.

No noso corpus, encontramos contextos nos que aparecen dous acentos seguidos nas oracións que presentan núcleo oxítono e extensión proparoxítona, ben no suxeito ben no complemento. Sorprendentemente, o comportamento que rexistramos nestes casos parece reforzar este choque acentual en lugar de evitalo que sería o esperable. En relación á frecuencia fundamental, o choque de acentos maniféstase unicamente nas oracións con extensión no complemento. Non observamos ningunha incidencia da F0 nas oracións con extensión no suxeito.

Lembremos que na epígrafe 6.1.3. viamos como se realizaba un último pico tonal sobre a sílaba átona que segue á tónica (véxase gráfico 9 da citada epígrafe), porén nos exemplos con dous acentos consecutivos este ascenso final non aparece. O acento do núcleo asóciase ás postónicas e nestes exemplos ao non haber sílabas átonas que realicen o ascenso (entre as tónicas do complemento) necesariamente ha de producirse unha desacentuación dun dos acentos do complemento, que nos nosos exemplos é o primeiro (gráfico 10¹⁴).

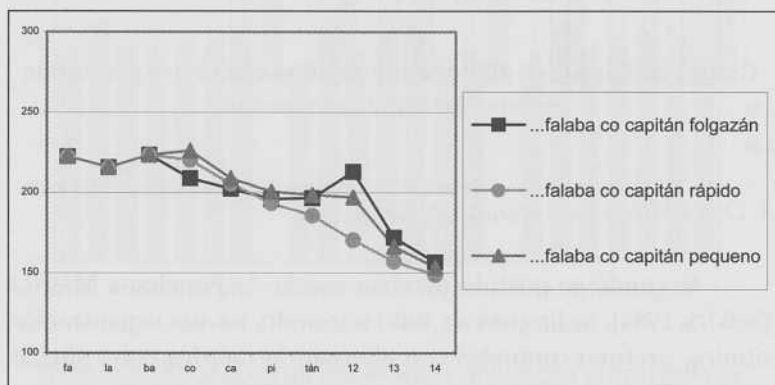


Gráfico 10: Núcleo oxítono do complemento combinado coas tres extensión.

Vexamos agora cal é o comportamento da duración nestes casos con dous acentos seguidos. Para iso, elaboramos os seguintes gráficos con enunciativas, nos que mostramos por un lado os complementos con núcleo oxítono comparados cos modificadores nas tres posicións acentuais (gráfico 27), e por outro lado os suxeitos co núcleo oxítono combinado cos tres modificadores (gráfico 28). Nas oracións interrogativas, rexistramos este mesmo comportamento das enunciativas, pero prodúcese dunha forma menos acusada. No gráfico 27 observamos como a duración da sílaba 11, que se corresponde coa tónica do núcleo, au-

¹⁴ Reproducimos o gráfico 10 da epígrafe 6.1.4 para contrastalo con maior comodidade cos datos da duración.

menta significativamente cando a continuación vén a tónica do modificador.

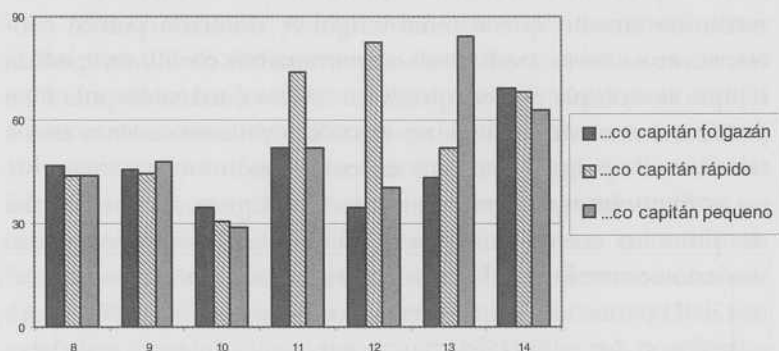


Gráfico 27: Duración das sílabas do complemento nas enunciativas co seu núcleo oxítono

No suxeito tamén se produce un incremento considerable da duración da tónica do núcleo (sílabas 4), nos exemplos en que hai dous acentos contiguos. No gráfico 28, a duración destaca cando os dous acentos léxicos van inmediatamente seguidos. Non sucede isto, cando hai algunha sílaba átona no medio.

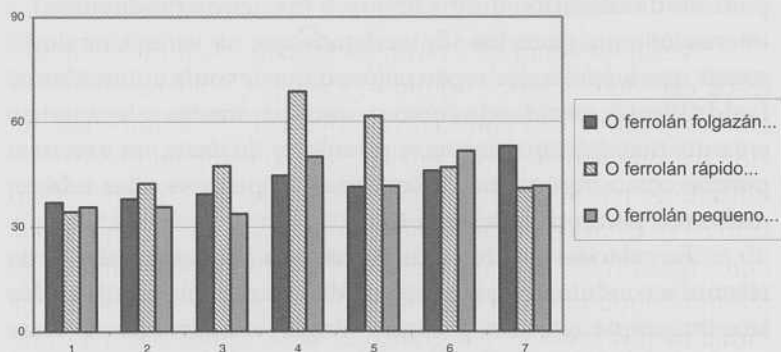


Gráfico 28: Duración das sílabas do suxeito nas enunciativas co seu núcleo oxítono

Mirando os dous gráficos observamos que no complemento o primeiro acento ten menor duración có segundo, a diferenza do suxeito que é máis longo. A duración parece contrarrestar en certa medida o comportamento da F0, na medida en que no complemento o primeiro acento é marcado pola F0 e posúe a menor duración, e no suxeito o primeiro acento non é tan marcado pola F0 e posúe a maior duración.

Este comportamento de acentos contiguos, no que a sílaba do primeiro acento aumenta a súa duración sen producirse traslado acentual, xa é descrito por outros autores como Prieto et al. (2001) para o catalán, Silvermann e Pierrehumbert (1990) para o inglés ou Arvaniti (1994) para o grego. No entanto, non deixa de ser un comportamento interesante posto que é contrario á suposta tendencia das linguas a evitar contextos acentuais difíciles ou pouco frecuentes, xa que o choque de acentos vese reforzado.

9. Conclusións

Neste estudo, analizamos os tres parámetros que afectan, segundo Cruttenden (1986) e outros, ao deseño dos distintos padróns da entoación dunha lingua: a frecuencia fundamental, a intensidade e a duración. Entendemos que as variacións de F0 son as que teñen maior repercusión, o que levou a autores como Ladd (1996) a considerala como o único parámetro relevante; no entanto, non debemos esquecer os valores da duración e enerxía porque como observamos anteriormente, parecen estar relacionados coa percepción dos acentos.

En relación aos datos da **frecuencia fundamental**, é interesante a conduta dos picos tonais dentro das enunciativas. No suxeito simple que non presenta ningún modificador na frase nominal, prodúcese unha elevación de ton habitualmente asociada á(s) sílaba(s) postónica(s). No suxeito oxítono, a sílaba pretónica da forma verbal funciona en realidade como postónica do suxeito, en tanto que culmina o seu ascenso. A tónica nestes

exemplos ben comeza suavemente esta elevación tonal, ben descende para logo ascender rapidamente na postónica. Este é o caso do suxeito paroxítono.

Cando o suxeito presenta unha extensión ou modificador, o comportamento é diferente. A elevación tonal do suxeito remata na sílaba tónica do modificador e non na(s) postónica(s) como acabamos de describir no parágrafo anterior. As sílabas átonas posteriores descenden en frecuencia significativamente. Estes datos non coinciden tampouco cos achegados noutros estudos do proxecto AMPER realizados coa mesma metodoloxía. Así en Fernández Rei et al. (2005a e 2005b), analizando unha fala de Santiago de Compostela, dicíamos que o suxeito con e sen extensión alcanzaba o seu máximo tonal sempre nas postónicas. Nos exemplos analizados de Camelle só encontramos dous casos en que as postónicas continúan e culminan o ascenso da tónica, polo que entendemos que esta é unha das posibilidades dentro da zona xeográfica, pero non representa a conduta máis frecuente das enunciativas neutras nesta falante.

Dentro do predicado tamén encontramos diferenzas entre o complemento simple e o complemento con extensión. A posición do acento no substantivo do complemento simple non parece condicionar no descenso paulatino da frecuencia fundamental. No complemento con extensión tampouco é significativa a configuración acentual do adxectivo modificador final. Si é relevante dentro do núcleo, posto que no complemento con extensión se produce un último pico tonal na sílaba átona inmediatamente posterior á tónica: na sílaba postónica do núcleo paroxítono, na primeira postónica do núcleo proparoxítono e na pretónica da extensión co núcleo oxítono. Comprobamos tamén que nesta elevación tonal sobre a postónica do núcleo do complemento ten moita influencia a posición dos acentos léxicos do substantivo e do adxectivo dentro do complemento: canto máis afastados maior é a elevación e menor canto máis achegados. O único caso no que non se produce ningún pico tonal é cando os dous acentos son consecutivos (...co *ferrolán rápido*).

Con respecto ás interrogativas, o característico desta modalidade está no lugar en que se produce o descenso tonal

brusco. No modelo de interrogativas de Camelle a caída aparece sobre a última sílaba acentuada do suxeito.

O comportamento desta modalidade camellá é moi peculiar con respecto ás interrogativas (do tipo absoluto ou totais) doutras zonas xeográficas do territorio galego, que se describi-ron noutros traballos centrados na entoación do galego. Así, en Fernández Rei (1995:186), describe os padróns máis comúns das interrogativas do galego a partir de varios informantes de distintas zonas xeográficas:

“as interrogativas totais caracterízanse por comezar nun nivel alto, que se mantén en suspensión ata a vocal que precede ó núcleo tonal, sexa tónica ou átona, para rematar nun nivel baixo, por debaixo da frecuencia fundamental media, no mesmo núcleo tonal”.

Este padrón das interrogativas absolutas é común ao descrito en traballos posteriores por outros investigadores con diferentes informantes, como por exemplo no traballo de investigación tutelado de Rodríguez (2003), que segue a mesma metodoloxía que Fernández (1995) ou nos artigos do proxecto AMPER: Fernández et al. (2005a e 2005b) sobre a entoación dunha fala de Santiago de Compostela:

“As interrogativas caracterízanse por presentaren o primeiro pico tonal no suxeito, tamén asociado ao acento do substantivo, que alcanza o seu máximo tonal na sílaba posterior á tónica, para a partir dese momento manterse nun ton notablemente alto ata a última sílaba tónica do enunciado, onde se produce un descenso brusco (moitas veces, superior a unha octava) ata a base tonal da falante, que se estende na(s) postónica(s) en complementos paroxítonos e proparoxítonos” (Fernández et al. 2005b:151).

Existen outros traballos nos que se describen outros contornos melódicos diferentes para as interrogativas absolutas, seguindo un enfoque dialectal. Sobrino (1997) unha serie de padróns das interrogativas totais en oito informantes do Baixo

Miño. Estes padróns caracterízanse por comezar a interrogativa nunha tonalidade media na que se desenvolve ata a última sílaba tónica do complemento que ascende con maior ou menor pendente. Nas oracións con sílabas postónicas o ton volve descender cara ao ton final da oración¹⁵. E por último, en Fernández et al. (2007), achegabamos unha serie de datos con respecto á realización das interrogativas (sen extensión) de Bueu. Diciamos que o seu comportamento era semellante ao descrito para as interrogativas de Santiago de Compostela a excepción do final, a partir do núcleo tonal: o tonema final que era descendente en Santiago, tiña unha realización ascendente en Bueu¹⁶.

Por outro lado, volvendo aos datos do noso traballo, podemos dicir que a **duración** e a **enerxía** non son parámetros pertinentes na distinción das modalidades oracionais estudadas (afirmativas neutras e interrogativas totais). Pero si parecen sinalar a posición dos distintos acentos da oración, posto que as sílabas que se producen con maior duración e intensidade son as que levan a tonicidade da palabra.

Neste estudo, tamén falabamos daqueles casos nas oracións con extensión en que aparecían dous acentos léxicos continuos (*ferrolán rápido* e *capitán rápido*). No noso estudo, observamos que nestes casos a duración da sílaba do primeiro acento (*ferro-lán* e *capi-tán*) aumentaba considerablemente con respecto á duración da mesma sílaba nun contexto sen *clash* acentual, seguida de sílaba átona. Con respecto á frecuencia fundamental unicamente encontramos variacións nos exemplos con extensión no complemento. Nestes casos, producíase no complemento un incremento tonal na sílaba átona inmediatamente posterior á tónica do núcleo. Pero, cando aparecían dous acentos contiguos a frecuencia descendía suavemente, sen realizar ningún pico tonal.

¹⁵ Fai referencia tamén a un padrón que só rexistra unha vez nunha soa falante e que é completamente diferente aos anteriores: "A secuencia empeza nun nivel alto e descende de xeito continuado. A tónica final xa chega por debaixo da frecuencia fundamental media" (Sobriño 1997: 115).

¹⁶ Só en oracións con postónicas no substantivo do complemento.

Por último, pensamos que a partir das análises e conclusións deste traballo, achegamos unha serie de datos que contribuirán ao deseño dun mapa dialectal da prosodia do galego, segundo os obxectivos e metodoloxía do proxecto internacional AMPER. Porén, tamén se abordan cuestións apenas investigadas para o galego, como a focalización e o choque de acentos, que foron tratadas abundantemente en traballos centrados na prosodia de diferentes linguas (inglés, portugués, español, catalán, etc). Nestes casos, o noso obxectivo, esperamos que cumprido, é proporcionar material e descricións que permitan un posterior afondamento nestes temas.

10. Referencias bibliográficas

- Arvaniti, A. (1994): "Acoustic features of Greek rhythmic structure", *Journal of Phonetics* 22, 239-268.
- Cruttenden, Alan (1986): *Intonation*. Cambridge: Cambridge University Press.
- Escourido, Ana (2007): "Aproximación ao estudo da focalización no galego", en González González, Manuel, Elisa Fernández Rei e Begoña González Rei (eds.), pp. 261-277.
- Fernández, Elisa (1995): *Aproximación á entoación dos enunciados interrogativos en galego. Estudio acústico*. Memoria de licenciatura. Universidade de Santiago de Compostela (inédita).
- Fernández Rei, Elisa e Lurdes de Castro Moutinho (2006): "A fronteira xeográfica do Minho: ¿tamén fronteira prosódica?", in Cristina Mourón e Teresa Iciar Moralejo (eds.): *Studies in contrastive linguistics. Proceedings of the 4th International Contrastive Linguistics Conference*. Santiago de Compostela: Universidade, 265-276.
- Fernández Rei, Elisa; Manuel González González, Luís Xuncal Pereira e Maruxa Caamaño Varela, (2005a): "Achega á entoación dunha fala do centro de Galicia. Contribución para o AMPER", *Géolinguistique*, Hors de Série 3, 87-102.

- Fernández Rei, Elisa; Ana Escourido; Maruxa Caamaño e Luís Xuncal (2005b): "A entoación dunha fala de Santiago: fronteira prosódica entre suxeito e predicado", *Estudios de Fonética Experimental*, XIV, 141-165.
- Fernández Rei, Elisa; Ana Escourido Pernas e Maruxa Caamaño Varela (2007): "Caracterización prosódica das interrogativas da Costa da Morte e do Morrazo", en González et al. (eds.), 305-315.
- González González, Manuel; Elisa Fernández Rei e Begoña González Rei (eds.): *III Congreso Internacional de Fonética Experimental*. Santiago de Compostela: Xunta de Galicia.
- Ladd, D. Robert (1996): *Intonational phonology*. Cambridge: Cambridge University Press.
- Martínez Celdrán, Eugenio; Ana María Fernández Planas; Josefa Dorta Luís e Elisa Fernández Rei (2006): "La intercomprensión d'interrogatives absolutes amb tonemes descendents del català, de l'espanyol de canàries i del gallec", *Estudis Romànics* 28, 7-28.
- Martínez Celdrán, Eugenio; Ana María Fernández Planas; Josefa Dorta Luís e Elisa Fernández Rei (2007): "Reconocimiento de variedades lingüísticas a partir de la entonación: el caso de algunas interrogativas de Tenerife, Santiago de Compostela y Barcelona", en González González, Manuel; Elisa Fernández Rei e Begoña González Rei (eds.): *III Congreso da Sociedade Española de Acústica Forense*. Santiago de Compostela: Xunta de Galicia, 225-247.
- Prieto, Pilar; S. Oliva; B. Palmada; P. Serra; B. Blecua; S. Llach e V. Oliva (2001): "Anàlisi acústica de la resolució de xocs accentuals en català", *Estudios de Fonética Experimental* XI, 11-40.
- Rodríguez Álvarez, Xosé Antonio (2003): *Patróns de entoación das interrogativas*. traballo de investigación tutelado polo profesor Xosé Luís Regueira (inédito).
- Selkirk, E. (1984): *Phonology and syntax: the relation between sound and structure*. Cambridge, Ma.: MIT.
- Silverman, K., e J. Pierrehumbert (1990): "The timing of prenuclear high accents in English", en J. Kingston e M. E.

Beckman (eds.): *Papers in Laboratory Phonology I. Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 72-106.

Sobrinho, María dos Anxos (1997): *A entoación do galego do Baixo Miño. Aproximación descriptiva*. Memoria de licenciatura. Universidade de Santiago (inérita).

11. Anexos

Anexo 1: CORPUS FIXO

1. Frases sen extensión

Enunciativas	Interrogativas
O ferrolán falaba co capitán O ferrolán falaba co cabalo O ferrolán falaba co cómico O soldado falaba co capitán O soldado falaba co cabalo O soldado falaba co cómico O médico falaba co capitán O médico falaba co cabalo O médico falaba co cómico	O ferrolán falaba co capitán? O ferrolán falaba co cabalo? O ferrolán falaba co cómico? O soldado falaba co capitán? O soldado falaba co cabalo? O soldado falaba co cómico? O médico falaba co capitán? O médico falaba co cabalo? O médico falaba co cómico?

2. Frases con extensión no suxeito

Enunciativas	Interrogativas
O ferrolán folgazán falaba co capitán O ferrolán pequeno falaba co capitán O ferrolán rápido falaba co capitán O ferrolán folgazán falaba co cabalo O ferrolán pequeno falaba co cabalo O ferrolán rápido falaba co cabalo O ferrolán folgazán falaba co cómico	O ferrolán folgazán falaba co capitán? O ferrolán pequeno falaba co capitán? O ferrolán rápido falaba co capitán? O ferrolán folgazán falaba co cabalo? O ferrolán pequeno falaba co cabalo? O ferrolán rápido falaba co cabalo? O ferrolán folgazán falaba co cómico?

O ferrolán pequeno falaba co cómico O ferrolán rápido falaba co cómico	O ferrolán pequeno falaba co cómico? O ferrolán rápido falaba co cómico?
O soldado folgazán falaba co capitán O soldado pequeno falaba co capitán O soldado rápido falaba co capitán	O soldado folgazán falaba co capitán? O soldado pequeno falaba co capitán? O soldado rápido falaba co capitán?
O soldado folgazán falaba co cabalo O soldado pequeno falaba co cabalo O soldado rápido falaba co cabalo	O soldado folgazán falaba co cabalo? O soldado pequeno falaba co cabalo? O soldado rápido falaba co cabalo?
O soldado folgazán falaba co cómico O soldado pequeno falaba co cómico O soldado rápido falaba co cómico	O soldado folgazán falaba co cómico? O soldado pequeno falaba co cómico? O soldado rápido falaba co cómico?
O médico folgazán falaba co capitán O médico pequeno falaba co capitán O médico rápido falaba co capitán	O médico folgazán falaba co capitán? O médico pequeno falaba co capitán? O médico rápido falaba co capitán?
O médico folgazán falaba co cómico O médico pequeno falaba co cómico O médico rápido falaba co cómico	O médico folgazán falaba co cómico? O médico pequeno falaba co cómico? O médico rápido falaba co cómico?

3. Frases con extensión no complemento

Enunciativas	Interrogativas
O ferrolán falaba co capitán folgazán	O ferrolán falaba co capitán folgazán?
O ferrolán falaba co capitán pequeno	O ferrolán falaba co capitán pequeno?
O ferrolán falaba co capitán rápido	O ferrolán falaba co capitán rápido?
O ferrolán falaba co cabalo folgazán	O ferrolán falaba co cabalo folgazán?
O ferrolán falaba co cabalo pequeno	O ferrolán falaba co cabalo pequeno?
O ferrolán falaba co cabalo rápido	O ferrolán falaba co cabalo rápido?
O ferrolán falaba co cómico folgazán	O ferrolán falaba co cómico folgazán?
O ferrolán falaba co cómico pequeno	O ferrolán falaba co cómico pequeno?
O ferrolán falaba co cómico rápido	O ferrolán falaba co cómico rápido?
O soldado falaba co capitán folgazán	O soldado falaba co capitán folgazán?
O soldado falaba co capitán pequeno	O soldado falaba co capitán pequeno?
O soldado falaba co capitán rápido	O soldado falaba co capitán rápido?
O soldado falaba co cabalo folgazán	O soldado falaba co cabalo folgazán?
O soldado falaba co cabalo pequeno	O soldado falaba co cabalo pequeno?
O soldado falaba co cabalo rápido	O soldado falaba co cabalo rápido?
O soldado falaba co cómico	O soldado falaba co cómico

folgazán O soldado falaba co cómico pequeno O soldado falaba co cómico rápido O médico falaba co capitán folgazán O médico falaba co capitán pequeno O médico falaba co capitán rápido O médico falaba co cómico folgazán O médico falaba co cómico pequeno O médico falaba co cómico rápido	folgazán? O soldado falaba co cómico pequeno? O soldado falaba co cómico rápido? O médico falaba co capitán folgazán? O médico falaba co capitán pequeno? O médico falaba co capitán rápido? O médico falaba co cómico folgazán? O médico falaba co cómico pequeno? O médico falaba co cómico rápido?
---	--

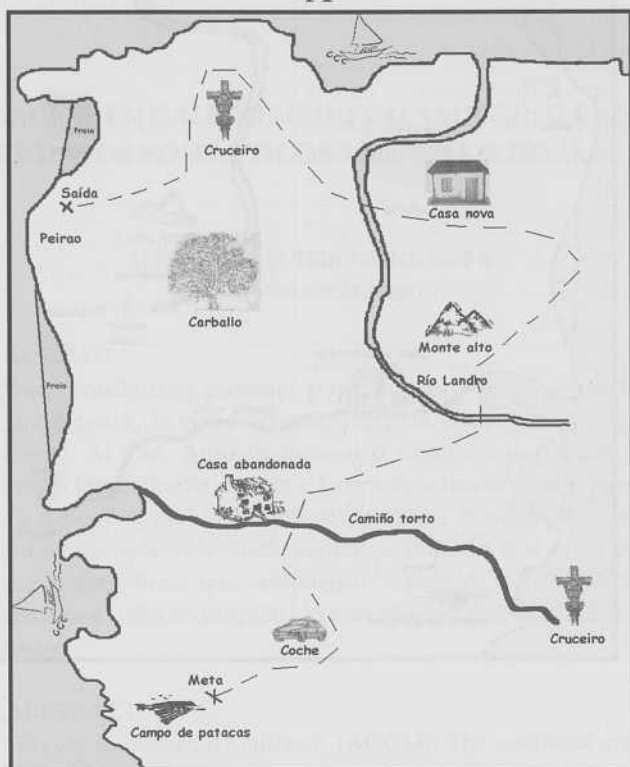
Anexo 2: CORPUS LIBRE DE USO COTIÁN

CONTEXTUALIZACIÓN	RESPOSTAS POSIBLES DO INFORMANTE
1- Non levas reloxo, necesitas saber a hora. Pasa un señor ao teu lado, que lle dis?	Ten hora? / Pódeme dicir a hora? / Que hora ten? / Que hora é? / Sabe que hora é?
2- Pasas pola rúa e un neno que non ten reloxo pregúntache pola hora. Contéstalle.	Son as doce e media / Doce e dez / É a unha e cuarto / Non levo hora / Non teño reloxo
3- Sabes que un veciño de oitenta anos estivo unha semana na cama con gripe. Tropezas con el. Pregúntalle polo seu estado.	Que tal vai da gripe? / Que tal está? / Como está? / Como se encontra? / Como anda? / Xa está mellor? / Vexo que xa se levantou da cama.
4- Estiveches unha semana enfermo de gripe. Atopas un veciño que che pregunta como estás. Contéstalle.	Vou mellor, grazas / Moi ben, grazas / Imos indo / Imos tirando / Mellor, grazas / Máis ou menos / Non hai queixa
5- Estás comendo cun amigo e tes o pan na outra punta da mesa, como lle dis que cho achegue?	Pódesme pasar o pan? / Pásame o pan / Cólleme o pan / O pan, por favor.
6- Un amigo co que estás comendo pídeche que lle achegues pan. Dille que efectivamente llo pasas.	Por suposto, toma / Si, claro, aquí tes / Toma/ aí vai / Home, iso non se pregunta.
7- Levas todo o día na casa. Chega o teu irmán de fóra. Pregúntalle polo tempo.	Que tempo vai aí fóra? / Chove? / Como anda o día? / Que tal tempo vai hoxe?
8- Estás na casa, pero acabas de chegar de fóra. O teu irmán pregúntache polo tempo. Infórmao.	Pois, agora mesmo loce o sol / Ten traza de chover / Barruza un pouco / Está chovendo / Vai un bonito día para pasear
9- Atopas na rúa un amigo despistado que non te ve. Que lle dis para que repare en ti?	Ola! / Ei, que tal? / Non saúdas?

10- Oes na rúa un saúdo que vai dirixido a ti e reparas na presenza ao teu lado dun amigo que hai tempo que non ves. Que lle dis?	Home! / Anda! / Non te vía / Canto tempo.
11- O amigo que encontras na rúa vai coa súa filla pequena. Pregúntalle á nena pola súa idade.	Cantos anos tes? / Que idade tes? / Que tempo tes?
12- Estás nunha vila que non coñeces e queres saber por onde se vai á Praza Maior. Pregúntalle a alguén.	Sabe onde queda a Praza Maior? / Está por aquí a Praza Maior? / A praza Maior, por favor?
13- Estás na Praza Maior. Un turista, que se dirixe a ti en galego, pregúntache onde está a Praza Maior. Respóndelle.	Pois é esta / Esta! / Aquí! / Non sei onde é, non son de aquí.
14- Es camareiro nun restaurante. Nunha das mesas, uns clientes acabaron xa o menú do día. Pídelles se queren café.	Queren café? / Van querer café? / Café?

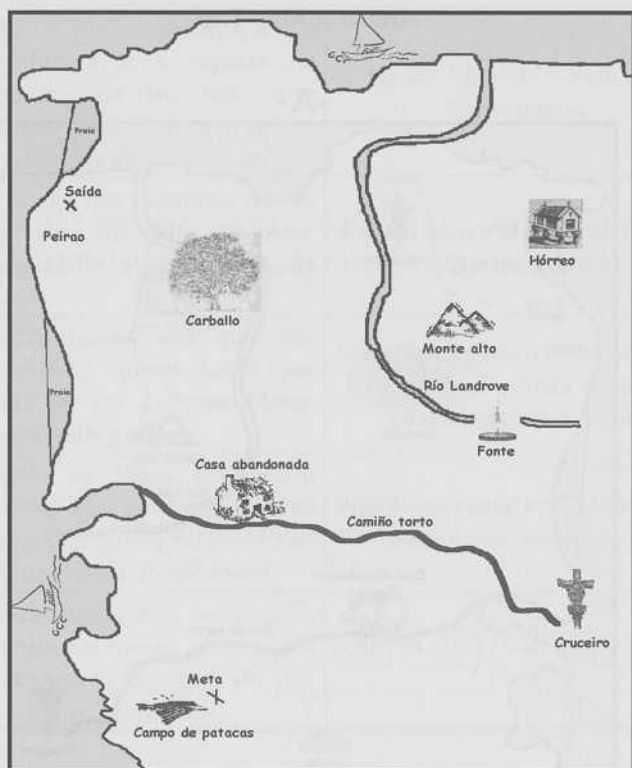
Anexo 3: Map Task

-A-



(Mapa do informante)

-B-



A ORACIÓN EN GALEGO MEDIEVAL (AOGM): O BANCO DE DATOS SINTÁCTICOS MEDIEVAIS DO ILG*

ALEXANDRE RODRÍGUEZ GUERRA
Universidade de Vigo

RESUMO

Esta contribución presenta por primeira vez un proxecto que xira ó redor do estudo da estrutura da cláusula en galego medieval: AOGM. Aproximámonos ó camiño que AOGM percorreu ó longo dunha década e facémolo a través dunha perspectiva descritiva moi xeral, orientada cara a tres vértices: os dous primeiros asócianse, basicamente, a cada un dos dous programas informáticos que xestionaron a base de datos sintáctica; o terceiro é unha recompilación e avaliación dos resultados alcanzados.

ABSTRACT

"Prayer in medieval Galician" (AOGM): The medieval syntactic database of ILG.

For the first time, this contribution presents a project centered around the study of the clause structure in medieval Galician:

* O presente artigo (versión moi modificada e ampliada da nosa intervención no *Seminario de la Lengua Española: presente y futuro*, que organizou o Departamento de Lengua Española da Universidade de Santiago de Compostela baixo a dirección de Emilio Montero Cartelle) insírese no proxecto de investigación interuniversitario *Diseño e implementación dun servidor de recursos integrados para o desenvolvemento de tecnoloxías da lingua galega (RILG)*, financiado polo MEC (código HUM2006-11125-C02-01). Pertence, en concreto, á parte do TMILG (Tesouro Medieval Informatizado da Lingua Galega), que dirixe no Instituto da Lingua Galega Xavier Varela Barreiro. Os xa mencionados, X. Varela e E. Montero, e mais a Ramón Mariño e Carme Pazos, os meus agradecementos.

AOGM. By means of a very general descriptive perspective we approach the line that AOGM went throughout a decade, which is directed towards three main points: the first two ones are basically associated to both computer programs in charge of manage the syntactic database; the third one is a summary and evaluation of the achieved results.

PALABRAS CHAVE

Galego medieval - Sintaxe histórica - Proxecto AOGM

KEY WORDS

medieval Galician, historical syntax, AOGM project.

1. Limiar

Nas páxinas que seguen tentaremos deseñar a grandes trazos a historia interna que, dende as orixes, conduciu á creación da base de datos sintáctica denominada AOGM. A presente achega, a pesar de proceder dun froito xa maduro con algúns lustros de existencia, ten aínda sabor inaugural pois é a primeira vez que reflexionamos publicamente sobre dela. Adaptamos, xa que logo, unha perspectiva xenérica e esencialmente descritiva que permita enxergar os alicerces esenciais sobre os que se foi construíndo e asentando AOGM. En primeiro lugar, AOGM non se podería entender á marxe dos tres grandes proxectos de investigación que, de maneira consecutiva, se estenderon ó longo da década que vai de 1993 a 2002¹:

1993-1996: *Gramática histórica da lingua galega.*

1997-1999: *A oración en galego medieval.*

2000-2002: *AOGM: os verbos de movemento.*

Foron tres proxectos independentes pero que sempre pivotaron ó redor dunhas coordenadas centrais, esencialmente

¹ Todos tres obtiveron o patrocinio institucional da Consellería de Educación e Ordenación Universitaria da Xunta de Galicia a través de convocatorias públicas.

humanas, que non mudaron: o director (Xavier Varela Barreiro), o centro de investigación (o Instituto da Lingua Galega da Universidade de Santiago de Compostela) e os tres coordinadores (Ramón Mariño para fonética e fonoloxía, Xavier Varela para a morfoloxía e Alexandre Rodríguez para a sintaxe). AOGM nace, como tal, no ano 1997 pero para entendelo cómpre situar as súas raíces nas reunións e traballos que no seo do proxecto de investigación *Gramática histórica da lingua galega* se desenvolveron no ILG dende o ano 1993 ata 1996.

Este último proxecto xirou, inicialmente, ó redor de tres eixes: fonética e fonoloxía, morfoloxía e sintaxe. Vai ser esta terceira póla a que, co tempo, constituirá o albo específico de AOGM. Á marxe do anterior, non se pode esquecer que a importancia dun proxecto como a *Gramática histórica da lingua galega* xustifica non só os traballos particulares nalgúns dos eidos devanditos, senón tamén outros proxectos paralelos, algúns dos cales son dunha enorme envergadura. Así, foi a elaboración da *Gramática histórica da lingua galega* a que "motivou a posta en marcha da construción dun corpus do galego medieval", mais tamén "durante bastantes anos quedou relegada a un segundo plano e os esforzos centráronse sobre todo nos traballos do TMILG" (ILG 2007: 24)².

2. Antecedentes

Sen ningunha dúbida, o principal motivo polo que se puxo en marcha AOGM foi a constatación do atraso que sufrían os estudos de sintaxe histórica galega a mediados da última década do s. XX. Este feito non nos ten que sorprender, pois Emilio

² Varela Barreiro (2004) proporciona unha revisión dos estudos de gramática histórica galega en xeral e do proxecto *Gramática histórica da lingua galega* en particular. Ademais, na páxina en rede do TMILG pódese atopar toda a información verbo do proxecto *Tesouro medieval galego-portugués* así como de todas as súas pólas; das cinco que o conforman, "a liña máis avanzada e de maior proxección externa" na parte galega é, precisamente, a do TMILG (cfr. <http://ilg.usc.es/tmilg>).

Ridruejo (1994: 587) refírese en idénticos termos á situación da sintaxe histórica do castelán. Para avaliarmos obxectivamente o devandito atraso, tomaremos como referente dúas obras de bibliografía lingüística galega publicadas nese período. A complementariedade de ambas as dúas (F. García Gondar 1995 - BILEGA a partir de agora- e X. L. Regueira 1996 -GBLG no que segue-), pois parten de presupostos metodolóxicos distintos (exhaustiva a primeira, selectiva a segunda), garante aínda máis a exactitude dos datos bibliográficos.

En BILEGA a sección dedicada á “Morfoloxía e sintaxe” posúe 266 referencias (5,2% do total); especificamente a “Sintaxe” presenta só 33 entradas (12,4%) -sorprende a cantidade de ítems que quedan en branco, sen ningunha referencia bibliográfica (pénsese que estamos a falar de 72 subepígrafes)-. Desas 33 entradas hai 13 (39,4%) que estudan a situación lingüística de etapas pretéritas (esencialmente a medieval). A “Lingüística histórica” merece unha atención específica e está conformada por 739 títulos (14,5% do total). A “Historia interna” dispón dun subparágrafo propio que supón, con 160 referencias, o 77,7% do total de “Lingüística histórica”. A parte da historia interna dedicada ós “Cambios sintácticos” conta só con 3 referencias.

En GBLG a epígrafe dedicada á “Morfoloxía e sintaxe” posúe 202 entradas e supón un 8,5% do total de referencias bibliográficas do libro. No seu interior realízanse cinco subdivisións e a quinta, “Sintaxe”, presenta 36 entradas, isto é, un 17,8% do total de traballos dedicados á morfosintaxe. Desas 36 publicacións, son 14 (38,9%) as que se sitúan no ámbito histórico³.

A táboa seguinte recompila a situación que acabamos de describir:

³ Francisco Antonio Cidrás Escáneo (2002) xa chamaba a atención sobre este feito: “na selectiva *Guía bibliográfica de lingüística galega* (que recolle a produción ata o remate de 1995) unicamente se inclúen 202 entradas de morfosintaxe (de características moi desiguais) dun total de case 2400 fichas recollidas”.

	Referencias totais	Fonética e fonoloxía	Morfoloxía	Sintaxe	Sintaxe histórica
BILEGA	5109	4,8% (244)	5,8% (294)	0,7% (36)	0,3% (16)
GBLG	2365	4,7% (112)	7% (166)	1,5% (36)	0,6% (14)

Táboa 1⁴

Os datos son o suficientemente significativos: do total de referencias bibliográficas escolmadas, a sintaxe histórica representaba -con 16 e 14 entradas, respectivamente- uns exiguos 0,3 e 0,6%. Á vista da situación precedente, a mediados da década dos noventa do século pasado semellaba máis que xustificado e necesario centrarse no traballo en sintaxe histórica galega debido, sobre todo, (i) á pouca presenza que supoñía dentro dos estudos de lingüística histórica, e (ii) á acusada desproporción, en xeral, do eido sintáctico fronte ó fonético/fonolóxico e ó morfolóxico. A situación a comezos da devandita década, sendo infinitamente mellor do que en décadas ou centurias pasadas, amosaba claramente unhas lagoas que cumpría encher. Non esquezamos que algúns anos despois, Ramón Mariño (2002) aínda sinalaba que “especialmente queda moito por facer nos campos da sintaxe histórica e do estudio histórico do léxico galego”.

Dende a óptica portuguesa non tiñamos modelos directos nos cales beber, pero si tentamos imitar, por exemplo, o magnífico labor desenvolvido pola investigadora brasileira Rosa Virgínia Mattos e Silva. Esta investigadora non só se achegou á sintaxe histórica do portugués a través dalgunhas das páxinas de estudos específicos (Silva 1989), senón que, a finais de 1990, fundou o grupo de investigación PROHPOR ‘Programa para a historia da lingua portuguesa’ (<http://www.prohpor.ufba.br/>).

Polo que se refire ás obras xerais, as gramáticas históricas galegas (ou portuguesas) adoitaron centrarse exclusivamente nos eidos fonético/fonolóxico e morfolóxico (por moito que se

⁴ Con BILEGA sumamos as referencias das epígrafes que se inclúen en “3 Morfoloxía e sintaxe” e mais en “6 Lingüística histórica”. En GBLG só contamos coas referencias explícitas (exclúense as cruzadas).

botase man da denominación 'morfosintaxe'). A sintaxe, tanto frástica coma oracional, ocupa, no mellor dos casos, só unhas poucas citacións, espalladas por distintos capítulos e que só en contadas ocasións constitúen un fin en si mesmas. Adoitan xirar, ademais, ó redor dun tema recorrente: a orde de palabras (sobre todo dos clíticos).

É certo que para achegarse á sintaxe histórica galega con garantías non abundaba unicamente coa constatación das lagoas. Precisábase a conxunción de distintos factores: o primeiro, contar cun grupo de lingüistas con ganas de traballar en gramática e, especialmente, en sintaxe histórica; o segundo, dispoñer dun modelo teórico e dunha metodoloxía de análise e interpretación sintáctica o suficientemente sólidos como para poder acceder, coa plena comprensión dos feitos sintácticos, á estrutura da cláusula galega medieval; o terceiro, alén da cobertura institucional, posuír un espazo físico propio e adecuado ás necesidades investigadoras dun proxecto destas características⁵; e o cuarto, non por último menos importante, a presenza dunha persoa que organizase e soubese aglutinar esforzos e encarreirar o conxunto (este papel correspondeulle a Xavier Varela, que sinalou e abriu o vieiro, evitou traballos baldíos e formou discípulos)⁶. Nesta encrucillada, houbo unha coincidencia de sinerxías: o grupo humano, o marco teórico e a metodoloxía de traballo en sintaxe histórica.

Verbo do primeiro, conseguíuse reunir un conxunto de investigadores voluntariosos e con coincidencia nas liñas de investigación. O grupo inicial estaba integrado por Xavier Varela Barreiro (director), Amparo Cereixo Silva, M^a Sol López Martínez, Eduardo Moscoso Mato, Carme Silva Domínguez e Alexandre Rodríguez Guerra. Así, concentrados esencialmente a comezos dos noventa do s. XX e cunha orientación basicamente funcionalista, agroman os primeiros traballos individuais de sintaxe histórica galega, ben como teses de doutoramento, ben

⁵ Neste sentido, o clima intelectual con que nos acolleu o ILG da universidade compostelá foi inmellorable.

⁶ Queremos que esta contribución sirva de humilde homenaxe a alguén que sempre regalou sabenza sen pedir nunca nada a cambio.

como memorias de licenciatura: Xavier Varela Barreiro (1992) céntrase nos cuantificadores, M^a Sol López Martínez (1993) no CDIR preposicional⁷, Marta Dacosta Alonso (1992) nas estruturas coordinadas, Eduardo Moscoso Mato (1992) na pasiva, Carme Silva Domínguez (1993) nos posesivos e, poucos anos despois, Amparo Cereixo Silva (1997) na sintaxe dos adverbios en *-mente*.

O segundo, polo menos no relativo ó marco teórico, alcanzouse plenamente dende o momento en que a tradición sintáctica de orientación funcionalista presente na universidade compostelá impregnou o período de formación e especialización de todos eles⁸. Non esquezamos que este aspecto constitúe a cerna de calquera empresa lingüística destas características. Emilio Ridruejo expón claramente cando afirma que “la dificultad más ardua radica en la elección de un modelo teórico que se pueda aplicar a la investigación de la sintaxis” (1994: 591).

O terceiro, indisociable do marco teórico elixido, é, canda o anterior, o aspecto máis complexo e senlleiro dos comentados. O éxito ou o fracaso de AOGM, coma en xeral de calquera base de datos, pasa pola precisión e potencialidade explicativa do modelo teórico e da metodoloxía empregada. Chegados a este punto cobra importancia a figura de Guillermo Rojo, cabeza visible dun magnífico equipo de lingüistas e responsable do proxecto “Base de Datos Sintácticos del Español Actual” (BDS; cfr. <http://www.bds.usc.es/>). Xa na última década do s. XX, foron os traballos orientados dende unha óptica funcionalista os que, no ámbito peninsular, afondaron máis e mellor no eido da sintaxe, sincrónica ou diacrónica, do castelán. Gravitando ó redor dos polos da escola funcionalista ovetense pero, sobre todo, da compostelá, os distintos investigadores (Antonio J. Meilán García 1991, Marta Pérez Toral 1992, Carlos Folgar 1993, Mercedes Suá-

⁷ Estes dous autores realizaron, xa a comezos da década anterior, cadansúa tese de licenciatura de natureza histórica: M^a Sol López Martínez (1981) e Xavier Varela Barreiro (1981).

⁸ Cfr. con dúas obras fundamentais para a fundación e consolidación da escola funcionalista compostelá: Guillermo Rojo (1978) e Guillermo Rojo e Tomás Jiménez Juliá (1989).

rez Fernández 1996, Elena Rivas e M^a José Rodríguez Espiñeira 1997, Inmaculada C. Báez Montero 1998, Manuel Mosteiro Louzao 1999, entre outros) foron debullando diversas peculiaridades de sintaxe histórica centradas sobre todo no ámbito oracional. Acaráanse, practicamente todos, ó devandito proxecto da BDS.

Guillermo Rojo, coñecedor dos nosos degoiros sintácticos, forneceunos todo o material, información e ánimos para que tentásemos aplicar no galego o modelo que tan boa recolleita lle estaba a dar á sintaxe (histórica ou non) do español⁹. Mercé ó que poderíamos denominar osmose entre áreas, puidemos adoptar para o galego medieval a metodoloxía de traballo da BDS. Dispoñer deste sistema de análise permitiunos aforrar un tempo e uns esforzos preciosos, posibilitounos comezar de inmediato a debullar na sintaxe galega medieval, forneceunos a aparataxe adecuada para realizar o noso labor con garantías e unha marxe de erro relativamente pequena, e obrigounos, unicamente, a realizar -polo menos nunha primeira fase- uns poucos axustes para que se puidese adaptar satisfactoriamente ás características sintácticas dos textos galegos medievais.

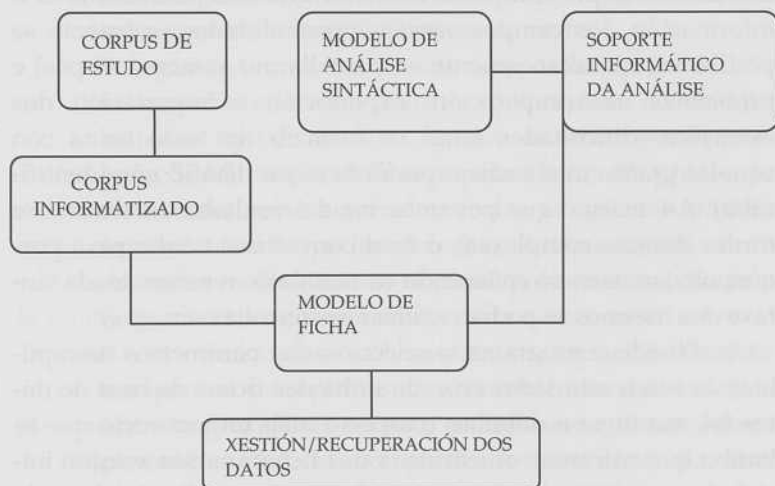
3. As fases do traballo

AOGM é, en esencia, unha base de datos informatizada de sintaxe galega medieval. Este xeito de traballar estaba tamén aconsellado, por exemplo, polo propio Emilio Ridruejo, quen lembra que, para a realización dunha obra xeral sobre sintaxe histórica, deberíamos contar "con un corpus, (...) y acceder a él, en lo posible, mediante los instrumentos informáticos adecuados" (E. Ridruejo 1994: 590). Nas páxinas que seguen describiremos, brevemente, as dúas grandes etapas en que podemos dividir a historia do proxecto de AOGM. A fronteira esencial entre ambas as dúas, aínda que tamén subxazan diferenzas de maior

⁹ Dende estas breves liñas, agradecemoslle publicamente a Guillermo Rojo a súa xenerosidade.

calado, responde a un criterio básico: o programa informático empregado (neste nivel non temos en conta as actualizacións ou versións novas de cada un) e, subsidiariamente, o modelo de estrutura da base de datos manexada.

Nas reunións iniciais (nas que interviñeron Xavier Varela Barreiro, M^a Sol López Martínez e Alexandre Rodríguez Guerra) deseñouse a estratexia metodolóxica concreta, fixéronse obxectivos xerais e específicos e propúxose unha periodización. As lagoas eran tantas que o traballo semellaba inabarcable: non había bases de datos galegas sintácticas, non existían córpora textuais galegos informatizados, non contabamos con ferramentas informáticas específicas...¹⁰ Isto sen entrar no aumento exponencial das dificultades de análise que implica traballar con textos medievais: a natureza distinta/específica dos datos, a súa escaseza ou parcialidade (cos problemas derivados de comprobar a (in)existencia de diversas posibilidades sintácticas/construtivas), a falta dun dicionario histórico, mesmo as dificultades de carácter ecdótico... Por medio da seguinte representación esquematizamos as distintas fases do traballo:



¹⁰ En X. Varela (2004: 668 e ss.) descríbese detalladamente todo o esforzo e dedicación que implicou esta fase.

3.1. A base de datos AOGM en dBASE

O primeiro programa informático utilizado por AOGM foi dBASE IV (de Ashton-Tate). Elixíuse porque, de entre os dispoñibles no mercado naquela altura, era o que mellor se adecuaba ás nosas necesidades, non tanto pola estrutura e características que presentaba senón polas posibilidades á hora de xestionar e recuperar a información. Del di Xosé Soto (2001: 281) que “foi a raíña das bases de datos baixo o sistema DOS” e “os expertos cualificana como unha base de datos de gran flexibilidade e inmensamente potente”. A interface que empregaba este programa naquela altura era a do sistema operativo MS-DOS e os nosos coñecementos informáticos, grazas tamén ó dominio no campo da programación en dBASE, permitiánnos xestionar adecuadamente e con certa axilidade a asunción de calquera tipo de buscas, simples e complexas, así como a recuperación selectiva dos contidos (introducindo directamente a orde na liña de comando).

Certamente, o programa tiña limitacións (por exemplo, os 254 caracteres por campo ou os inconvenientes para xestionar a información dos campos *memo*), incomodidades (soamente se podían ver simultaneamente en pantalla uns poucos campos) e problemas na recuperación, exportación e importación dos exemplos (dificultades tanto co formato do texto coma con aquelas grafías medievais específicas e que dBASE non identificaba). Ademais, o que por unha banda resultaba unha vantaxe (ordes directas complexas), ó final convertíase nunha pexa porque, obviamente, só coñecendo os mandatos e a non doada sintaxe dos mesmos se podían axilizar as consultas.

Elixido o programa, a selección dos parámetros susceptibles de seren estudados en cada unha das fichas da base de datos foi, sen lugar a dúbidas, o aspecto máis importante ó que se houbo que enfrontar. A estrutura das fichas, na súa versión inicial do ano 1996, baséase no verbo como predicado de cada unha das cláusulas obxecto de estudo e incorpora todos os aspectos sintácticos, potencialmente relevantes, das mesmas. Nesta primeira versión, o modelo de ficha constaba de 58 cam-

pos, todos eles alfanuméricos (non se empregaron campos denominados *memo*) e cun tamaño variable que ía de 1 a 254 caracteres. Sempre se traballou directamente a través do soporte informático da base de datos: non se empregaron soportes intermedios (papel, cartolina ou similares) na recolla da información porque había que rendibilizar ó máximo os (limitados) recursos de que dispuxo o proxecto.

En calquera caso, estamos diante da(s) versión(s) máis apegada(s) ó orixinal (BDS) que nos serviu de modelo. Os parágrafos que seguen describen sucintamente os campos de que constaba cada ficha, pero non o fan de xeito lineal, un a un, senón de maneira conxunta, a través da agrupación en tres claros módulos:

- En primeiro lugar, hai un grupo de campos que serve para localizar e identificar a forma verbal obxecto de estudo. Neles recóllense os seguintes datos: obra, autor, volume, capítulo, páxina, liña, ano e século. Incorpóranse tamén a cadanseu campo o contexto precedente, a forma verbal obxecto de análise específica, o contexto subsecuente e o lema (infinitivo) do verbo.
- En segundo termo, existen oito campos nos que se recolle información relativa á cláusula contemplada na súa globalidade: tipo de cláusula, función sintáctica que desempeña, modalidade, diátese e polaridade; do predicado identifícase cal é a forma verbal e advírtese a posible existencia dunha perífrase verbal; sinálase tamén a orde de todas as funcións sintácticas presentes na cláusula (argumentais ou periféricas).
- A continuación desenvólvense todas as funcións sintácticas que son susceptibles de se encontraren nas cláusulas obxecto de análise: suxeito, complemento directo, complemento indirecto, complemento preposicional, complemento predicativo, atributo, complemento circunstancial e complemento axente. De todas elas fornécese a información relativa ó tipo de unidade e, de case todas, márcase a animación, a determinación e o número;

cando é pertinente engádese tamén a posibilidade de sinalar a marca preposicional; dos complementos directo e indirecto tense en conta a posible existencia de redobro; co predicativo identifícase a función sintáctica a que se refire e, finalmente, co complemento circunstancial sinalase o tipo semántico a que pertence.

Existe, ademais, un campo de libre disposición para o investigador que lle permite realizar calquera comentario ou observación.

Con que información se cobre cada un deles? Podemos distinguir, de inicio, aqueles campos que se enchen empregando unha codificación previamente estipulada daqueles outros que se cobren de maneira directa. Entre os primeiros, hainos que admiten a escolla entre unha dupla ou tripla opción pechadas/fixadas de antemán (é o que acontece cos campos que desenvolven a animación, a determinación e o número, por exemplo) pero hainos tamén que, presentando unha lista máis ou menos fixa de opcións, admiten respostas máis complexas (por exemplo, 6 na modalidade, 7 no tipo de unidade, 10 no suxeito, 15 na función da cláusula ou 17 no tipo de oración). A media é de 5 ítems por campo e, de contemplarmos separadamente os grupos sinalados nos parágrafos precedentes, o relativo á contorna da cláusula amosa unha media de case dez ítems por campo fronte ós case catro do relativo á contorna das funcións sintácticas. Os segundos, que supoñen un total de dez campos, cóbrense directamente, por exemplo, coa preposición que actúa de marca das distintas funcións sintácticas, co redobro clítico nos complementos directo e indirecto ou coa identificación da perífrase verbal concreta.

Como xa se amosou no esquema precedente, procedemos simultaneamente á constitución dun corpus obxecto de estudo. Alén de polo factor cronolóxico, guiámonos por un criterio básico: a non exclusión de ningunha tipoloxía textual porque, tal e como sostén Rafael Cano, para conseguirmos os nosos obxectivos "cualquier texto es válido" (1994: 585) e, como lembra Emilio Ridruejo, o corpus cómpre que sexa "si no muy extenso al

menos suficientemente representativo" (1994: 590). En todo caso, sempre tivemos en conta o parámetro da tipoloxía textual á hora de interpretar os resultados pois, como pon de manifesto J. F. García Santos (1996: 300), "con mucha frecuencia el contenido del texto determina un tipo o unos tipos concretos de estructuras sintácticas", alén de que "es especialmente peligrosa la mezcla indiscriminada de fuentes en verso y en prosa", como perigoso é "extraer conclusiones de carácter general basadas en datos tan heterogéneos como los que pueden proporcionar uno y otro tipo de fuente". Daquela, con todas as precaucións posibles, realizamos unha escolma de cláusulas de distintos textos medievais que, na súa versión de xaneiro de 1996, quedou composta inicialmente por unhas 16000 cláusulas das cales só se analizaron pouco máis da cuarta parte (exactamente 4239)¹¹. Houbo que dixitalizar os exemplos, revisalos, seleccionar as oracións que ían ser estudadas e dispoñelas en cadansúa ficha.

Ó final do proceso, as cláusulas analizadas abrangueron un período que ía do século XIII ó XV (só residualmente aparecen exemplos de comezos do s. XVI) e, na busca da maior representatividade posible, recolleron produción lírica (profana e mariana), sete obras de prosa non notarial e outras tantas obras de prosa notarial. A distribución precisa segundo a tipoloxía anterior é a seguinte:

	Lírica		Prosa non notarial					Prosa notarial
XIII	LP (12,6%)	CSM (8,6%)						(8%)
XIV			TC (8,6%)	XH (10,1%)	HT (7,2%)	CT (9,4%)	LT (3,9%)	(7,5%)
XV			MS (12,5%)	CI (5%)				(4,5%)
XVI								(2,1%)

Táboa 2¹²

¹¹ Nesta primeira fase o traballo de análise desenvolvérono M^a S. López Martínez e A. Rodríguez Guerra.

¹² Non desenvolvemos as abreviaturas das obras por seren moi coñecidas. En todo caso, aparecen recollidas na lista de obras do TMILG en <http://ilg.usc.es/tmilg/obras.html>.

Como se pode comprobar, de agruparmos os datos anteriores por grandes xéneros, é a prosa non notarial a que posúe unha maior presenza no corpus (co 56,7% dos exemplos)¹³, prosa notarial e lírica contan practicamente coa mesma cantidade de casos analizados (22,1% e 21,2%, respectivamente). De agruparmos as cláusulas analizadas por séculos advírtese unha clara concentración no século XIV co 46,7%; a bastante distancia está o s. XIII co 29,2% dos exemplos, o s. XV co 22% e, de maneira residual, hai algún exemplo de prosa notarial datado a comezos do s. XVI (2,1%). En §4 fornécese unha recompilación dos estudos publicados mercé á análise de todas estas cláusulas de AOGM.

3.2. *A base de datos AOGM en Access*

Seis anos despois, os compoñentes do proxecto de AOGM enfrontáronse á necesidade de anovar (i) o soporte informático de AOGM e (ii) a estrutura da base de datos¹⁴. Son bastantes as novidades que acompañan o cambio do programa xestor da base de datos AOGM. Neste sentido, dBASE vai ser substituído por Access; este é, probablemente, o xestor de datos máis popular baixo a contorna de Windows. Del di X. Soto que “é na actualidade un dos sistemas máis recomendados para almacenar e manexar datos” (2001: 281). A unha interface nidiamente mellorada (moito máis agradable e cómoda á vista) hai que engadirlle todas as posibilidades que permite a opción dos formularios: con facilidades á hora de cubrir os campos mercé ós menús despregables (que evitan ter que memorizar unha longa serie de convencións e posibilitan unha maior complexización das respostas) e cunha axuda na barra inferior que explica o campo en que está situado o cursor. Todo o anterior facilita enormemente

¹³ Por subxéneros o reparto é o seguinte: prosa literaria LT, CT, HT, prosa histórica TC, XH, prosa relixiosa CI, MS.

¹⁴ O grupo de investigadores que se vai envorcar nesta segunda etapa de AOGM queda constituído por: Xavier Varela Barreiro (director), Manuel Blanco García, Amparo Cereixo Silva e Alexandre Rodríguez Guerra.

a utilización por parte de calquera usuario. Coa estrutura específica do formulario que xestiona a presentación organizada de AOGM conséguese que, aínda que non caiban todos os campos simultaneamente nunha mesma pantalla (precísanse tres), haxa uns que sempre estean visibles na parte superior (módulo *a*), no eixe vertical da esquerda (campos do módulo *b*) e á dereita os do módulo *e*. Isto permítelle ó investigador dispoñer ininterrompidamente da información básica da cláusula que está a analizar. Os anteriores están acompañados na primeira pantalla polos campos das funcións SUX, CDIR, CIND, CPREP, CAX e Dativo de Solidariedade, na segunda polas de CCIR e PTVO e, finalmente, na terceira polo campo denominado 'pendente' (reservado para que o investigador reflexione sobre as dúbidas ou comentarios verbo da análise realizada). Na seguinte reprodución pódese comprobar como é a primeira das pantallas:

Co cambio de soporte informático, alén do xa comentado, tentouse acomodar máis e mellor a base de datos á realidade, lingüística e textual, do galego medieval. Paralelamente ó proce-

so de debate e construción do formulario anterior, procedeuse á elaboración dun manual de apoio tanto para o manexo da base de datos coma para a interpretación sintáctica das cláusulas galegas medievais¹⁵. Nel realizouse un percorrido lineal por todos os campos da estrutura da base de datos, con explicacións das motivacións e obxectivos que guían cada un, cun repaso de cáles son as opcións para enchelos e cáles son os criterios xerais que cómpre empregar. Así mesmo, propuxéronse análises-tipo para determinados exemplos dubidosos ou problemáticos e fornecéronse, en todos os casos, exemplos abondos. Tratouse de facilitar ó máximo o labor de análise tentando minimizar a diverxencia, favorecendo a posta en común, axilizando o proceso de revisión e permitindo o intercambio de información entre os colaboradores.

Tal e como xa se adiantou, no novo formato de ficha prodúcese un cambio na estrutura da base de datos, que pasa a posuír 127 campos organizados en cinco módulos¹⁶:

- (a) Módulo do exemplo (13 campos).
- (b) Módulo do predicado e da cláusula (25 campos).
- (c) Módulo das funcións clausais (79 campos).
- (d) Módulo do modificador e dúbidas (4 campos).
- (e) Módulo do(s) investigador(es) (6 campos).

Entre as principais novidades podemos escolmar as seguintes:

- (a) Módulo do exemplo:
 - (i) incorpórase a identificación exclusiva da cláusula na base;

¹⁵ A cerna do cal débémolla a X. Varela Barreiro.

¹⁶ Para podermos avaliar mellor a nosa proposta, podémola confrontar coa de X. Soto para estudar a intransitividade. Este autor (2001: 282-290) establece tres grandes bloques que configuran a estrutura da súa base de datos sintáctica: 'atributos do predicado e da construción' (con 20 campos), 'atributos do suxeito' (con 6 campos) e 'atributos dos complementos' (con 5 campos).

- (ii) o investigador ten ó seu dispor tanto a cláusula obxecto de análise coma un contexto amplo da anterior.
- (b) Módulo do predicado e da cláusula:
 - (i) márcanse as construcións complexas (lexías, mandato, percepción...);
 - (ii) xerarquízouse a escolla da forma verbal da cláusula relacionada -bipolares, subordinadas, principais-;
 - (iii) especificanse os casos de predicado ausente -sobre todo na apódose das bipolares-;
 - (iv) reformuláronse os campos de tipo/función da cláusula:
 - tipo: independencia, coordinación, subordinación (substantiva, adverbial típica/atípica -prótase-, adxectiva).
 - función: independente, oracional (coordinación), clausal (SUX, CDIR, CIND...), frástica
 - incrustación: principais de substantivas, adxectivas e adverbais (típicas/atípicas -apódose-);
 - (v) redes xerárquicas: márcase a función desempeñada na estrutura inmediatamente superior (máximo de 4 niveis);
 - (vi) a orde de constituíntes clausais recóllese nunha versión completa e noutra reducida (exclusivamente coas funcións argumentais).
- (c) Módulo das funcións clausais:
 - (i) márcase o dativo de solidariedade;
 - (ii) especificanse tipos de CIND (semántica), de CPREP (suplemento, modal, adverbial...) e de CCIR (ata 22 tipos);
 - (iii) aumenta a precisión tanto dos tipos de unidade (contéplanse 58 diferentes) coma da determinación (16 subtipos).
- (d) Módulo do modificador e dúbidas:
 - (i) incorpórase información verbo do modificador da frase (bipolaridade ou coordinación).
- (e) Módulo do(s) investigador(es):

- (i) identifícanse nel a data e o nome do primeiro investigador que realizou a análise así como o do revisor e a necesidade, ou non, de repasar unha ficha concreta.

Con esta nova proposta, non só se tentou ampliar o número de parámetros que poidan amosar a súa relevancia para comprendermos mellor a sintaxe da cláusula galega medieval, aumentar a coherencia estrutural na análise sintáctica ou adecuar a proposta de análise á casuística real de exemplos medievais, senón que tamén se procurou aumentar o grao de fondura da información que se recolle en cada un deles e a tipoloxía das unidades sintácticas que se poidan describir. Dende esta perspectiva, interesounos alongar a liña do horizonte sintáctico pois, deste xeito, aumenta o abano de posibles traballos parciais que poidan xurdir ó abeiro de AOGM. Nesta dirección, non desbotamos a posibilidade de que a partir desta base de datos se obtengan estudos de natureza non só clausal senón tamén frástica ou mesmo pragmático-discursiva e textual.

Como resultado do traballo nesta segunda fase de AOGM case se alcanzaron as 33.000 cláusulas medievais analizadas¹⁷. O peso específico das obras parcialmente incluídas repártese do xeito seguinte: LP (32,9%), TC (19,8%), CT (15,2%), CSM (12,9%), GE (6,4%), MS (3%), HT (1,8%), CI (1,5%), FT (0,9%); PN (5,6%; sete obras). De establecermos unha clasificación por grandes xéneros obtemos unha distribución cunha forte, e equilibrada, presenza da prosa non notarial (48,6%) e da lírica (45,8%), fronte a unha representación máis cativa da prosa non literaria (5,6%). Se a perspectiva que se adopta é a cronolóxica, obsérvase que son os séculos XIII e XIV os que concentran a práctica totalidade dos exemplos (co 47,7 e 45,9%, respectivamente); ó s. XV só pertencen o 5,9% das cláusulas e ó s. XVI o 0,5%.

¹⁷ Cómpre salientar que as dúas terceiras partes delas son obra de Amparo Cereixo.

4. Os logros

Na fin dos dous últimos puntos (§3.1. e §3.2.) sinalamos cal foi o número total de cláusulas que se analizaron así como a súa distribución segundo distintos parámetros. Chegou o momento de verificar directamente os resultados obtidos por AOGM nunha dupla dirección: (i) cales son os estudos concretos que xurdiron ó abeiro desta base de datos e (ii) cal é o seu peso específico á vista dos recursos bibliográficos dispoñibles.

De consultarmos a información presente en BILEGA¹⁸ compróbase que na epígrafe "643 Historia interna do galego: cambios sintácticos" hai cento vinte e seis referencias bibliográficas¹⁹. Delas, o 81,7% non son especificamente galegas pois a vertente galega só agroma a carón dunha perspectiva románica, iberorromance, portuguesa ou castelá²⁰. Das referencias enfocadas exclusivamente dende a nosa lingua, o 56,5% publicáronse ata o ano 1999 (e só dúas delas xurdiron da man dalgún membro do noso proxecto). Obviamente, o 43,5% restante publicouse no ano 2000 ou seguintes e, fronte ó dato anterior, a práctica totalidade (80%) acaróase a investigadores que nun momento ou outro pertenceron ó proxecto AOGM (cfr. *supra*).

De todos os xeitos convén ser realistas: o camiño aberto aínda é corredoiro, pois moitas son as lagoas que cómpre encher e relativamente poucos e parciais os obxectivos alcanzados. Neste sentido, se comparamos no mesmo BILEGA a epígrafe dedicada ós cambios sintácticos coas outras epígrafes de historia interna do galego obtemos os datos seguintes:

¹⁸ Na súa versión 2.1; a data de consulta foi o 31 de maio de 2008.

¹⁹ Ningunha das cinco epígrafes que desenvolven a máis xeral "64 Historia interna do galego" conta con ningunha subdivisión interna. Porén, no capítulo "3 Gramática, pragmática, lingüística textual e análise do discurso", só na subepígrafe dedicada á sintaxe, "32 Sintaxe: traballos xerais", localízanse máis de sesenta subdivisións.

²⁰ O 31,1% publicouse no ano 2000 ou posteriores.

Cambios fonéticos e fonolóxicos	Cambios morfolóxicos	Cambios sintácticos	Hª interna en xeral	Hª do léxico e etimoloxía	Outros cambios
45% (421)	35,9% (335)	13,5% (126)	3,3% (31)	2,1% (20)	0,2% (2)

Táboa 3

Como se pode apreciar, o eido sintáctico aínda está moi afastado dos outros dous grandes conxuntos: o fonético/fonolóxico (a máis de trinta puntos) e o morfolóxico (a máis de vinte puntos). Agora ben, se diferenciamos entre os traballos publicados ata 1999 daqueles publicados dende o ano 2000 ata a actualidade, compróbase que a tendencia anterior está a mudar:

Cambios fonéticos e fonolóxicos		Cambios morfolóxicos		Cambios sintácticos		Hª interna en xeral		Hª do léxico e etimoloxía		Outros cambios	
>1999	2000>	>1999	2000>	>1999	2000>	>1999	2000>	>1999	2000>	>1999	2000>
89,8%	10,2%	82,4%	17,6%	67,5%	32,5%	83,9%	16,1%	70%	30%		100%

Táboa 4

Fóra da epígrafe "Outros cambios" (que soamente conta con dúas referencias bibliográficas), os "Cambios sintácticos" son os que concentran unha maior proporción de publicacións a partir do ano 2000 (case a terceira parte do total), unicamente está seguido moi de preto pola "Historia do léxico e etimoloxía" (a dous puntos e medio); as outras referencias oscilan entre a décima parte (dos "Cambios fonéticos e fonolóxicos") e algo máis da quinta parte dos "Cambios morfolóxicos". Ademais, se contemplamos en xeral as publicacións do ano 2000 e posteriores, as relativas ós "Cambios sintácticos" supoñen algo máis da cuarta parte (26,3%) de todas as referencias publicadas neste período (e que se inclúen en BILEGA dentro da epígrafe dedicada á "Historia interna do galego").

Concentrados especificamente nos 126 títulos citados na epígrafe "643 Historia interna: cambios sintácticos", alén da distribución xa vista no cadro anterior (un terzo deles publicouse no ano 2000 ou en anos posteriores), convén diferenciar entre

aqueles que están feitos dende unha perspectiva da lingua galega, daqueles outros que o fan dende outra perspectiva lingüística distinta ou, simplemente, máis xeral:

BILEGA (2008)	Perspectiva non galega	Perspectiva galega
>1999 (84: 66,7%)	84,5% (71)	15,5% (13)
2000> (42: 33,3%)	76,2% (32)	23,8% (10)

Táboa 5

Os datos corroboran que entre un bloque e outro houbo case un aumento porcentual de dez puntos a favor da perspectiva galega. De adoptarmos esta perspectiva, confírmase que as publicacións xeradas no seo de AOGM supuxeron o 15,4% das publicadas antes de 1999 e o 80% das publicadas a partir de 2000.

Os resultados alcanzados, sendo importantes, aínda se sitúan nunha fase inicial pois “as achegas do equipo neste terreo foron aparecendo como traballos individuais sobre aspectos puntuais do programa conxunto da obra” (ILG 2007: 24). Con todo, é certo que estes traballos individuais tocaron os máis diversos temas, como se pode comprobar na seguinte escolma: Amparo Cereixo (2000a e 2000b) céntrase na análise da función do CIND no galego medieval; Marta Dacosta (2004) estuda a coordinación medieval en exemplos tirados da lírica e da prosa non notarial; M^a Sol López Martínez (2004) realiza unha descrición dos esquemas sintácticos medievais do verbo *chamar*; Carme Silva (2002) debulla distintos aspectos dos valores das frases con posesivo no período medieval, (2001) concéntrase nas lexías complexas ou (2004) nos determinantes na CT; Alexandre Rodríguez (2000) realiza unha aproximación á xerarquía da animación dende a perspectiva galega, estuda os esquemas causais (2001b) e temporais (2001a) das cláusulas galegas medievais e fai o propio coa orde de constituíntes oracionais (2004 e 2005); Eduardo Moscoso (2000) realiza unha aproximación ás cláusulas galegas medievais con tempos compostos, en (2004) selecciona as que teñen só antepretérito e en (2005) céntrase nas que pre-

sentan os verbos *aver* e *ter* en contexto perifrástico; Ana Vidal (2004) reflexiona sobre as oracións causais e consecutivas tamén no galego medieval²¹.

Hai, por suposto, traballos de natureza exclusivamente académica que non foron publicados; é o caso do estudo que realizou Manuel Blanco (2002) de máis de 3.000 oracións medievas co verbo *ir* como traballo de investigación titorado. Igualmente, hai outras moitas investigacións en curso que se centran no período medieval (funcións clausais específicas, orde de constituíntes, estrutura da cláusula medieval, verbos concretos...), pero non en exclusiva. Así, por exemplo, Sonia Sobral estuda 2.000 cláusulas galegas (ss. XVII a XX) co verbo *andar* empregando o mesmo modelo de AOGM.

Aínda que a escolma anterior é, en efecto, moi limitada no que se refire ós eidos de sintaxe histórica galega xa cubertos, a potencialidade de AOGM permitiría, por exemplo, responder con exactitude todos os puntos enumerados por Emilio Ridruejo no punto 6 do artigo que vimos manexando (1994: 593): as categorías gramaticais, os procedementos de referencia, a modalidade oracional, as estruturas predicativas, as clases funcionais de adscrición de argumentos a predicados, os predicados complexos, a orde de palabras e os procedementos gramaticais de organización discursiva. Mesmo, posibilitaría afrontar o estudo doutros aspectos seleccionados tamén por E. Ridruejo (1994: 594-595) noutras epígrafes como, por exemplo, a evolución das clases léxicas de nomes, todos os tipos de oracións, as funcións comunicativas ou a historia da elipse, entre outros.

O paso do tempo, a experiencia do uso e a consecución de obxectivos parciais axudáronnos a mellorar a estrutura, precisión e calidade da información sintáctica, e a facilitar o xeito de a xestionar. Así, conseguíuse aumentar a fiabilidade e as posibilidades de explotación da base: porque a partir dela pódense desenvolver distintos traballos parciais, non só en niveis clausais (funcións sintácticas) senón tamén no nivel inferior (frase) e no

²¹ Na páxina <http://ilg.usc.es/tmilg/resultados.html> pódese consultar a lista completa de traballos realizados polos membros do proxecto.

superior (texto)²². Agardamos que no futuro a suma destes traballos individuais permita elaborar unha obra xeral, homoxénea, con visión de conxunto e que posibilite entender cales foron e como mudaron os distintos elos que configuraron a sintaxe na súa deriva dende a lingua latina cara ó romance galego (medieval).

Bibliografía

- Báez Montero, Inmaculada C. (1998): *La construcción con predicativo del complemento directo en español medieval*. Vigo: Universidade.
- BILEGA = Francisco García Gondar (dir.) (1994-): *Bibliografía Informatizada da Lingua Galega*. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades. Versión 2.1. [<http://www.cirp.es/bdo/bil/>].
- Blanco López, Manuel (2002): *Estructuras do verbo ir no galego medieval*. Traballo de investigación titorado. Universidade de Santiago de Compostela.
- Cano Aguilar, Rafael (1994): "Perspectivas de la sintaxis histórica española: el análisis de los textos", in *Actas del congreso de la lengua española (Sevilla, 7-10 octubre 1992)*. Madrid: Instituto Cervantes, 577-586.
- Cereixo Silva, Amparo (1997): *Os adverbios en -mente no galego medieval*. Tese de licenciatura. Universidade de Santiago de Compostela.
- Cereixo Silva, Amparo (2000a): "O complemento indirecto cos verbos de dicción. Marcaxe e colocación", in D. Kremer (ed.): *Actas do VI Congreso Internacional de Estudos Galegos: Un século de estudos galegos. Galicia fóra de Galicia*. Sada, A Coruña: Ediciós do Castro / Galicien-Zentrum der Universität Trier, vol. I, 143-155.

²² Neste sentido, é especialmente adecuada a precisión de R. Cano Aguilar (1994: 578) cando sinala que o que debemos perseguir é unha sintaxe 'dos' textos e non 'sobre' os textos. Non esquezamos que para este autor os textos son "actos comunicativos plenos" (1994: 581).

- Cereixo Silva, Amparo (2000b): "A marcaxe do complemento indirecto na *Crónica Troiana*", in I. de Riquer *et alii* (eds.): *Prof. Basilio Losada: Ensinar a pensar con liberdade e risco*. Barcelona: Universitat de Barcelona, 253-259.
- Cidrás Escáneo, Francisco Antonio (2002): "Os estudos de gramática", *Canal Estudos Galegos* 31; <http://www.vieiros.com/>
- Dacosta Alonso, Marta (1992): *A coordinación no galego medieval*. Tese de licenciatura. Universidade de Santiago de Compostela.
- Dacosta Alonso, Marta (2004): "A coordinación na lingua galega medieval, a través das *Cantigas de Amor* e de *Escarnio*, os *Miragres de Santiago*, a *Crónica Troiana* e a *Coronica de Iria*", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade (16-20 de setembro de 1996)*. Santiago de Compostela: ILG / Consello da Cultura Galega, vol. III, 157-176.
- Folgar, Carlos (1993): *Diacronía de los objetos directo e indirecto (del latín al castellano medieval)*. Santiago de Compostela: Universidade. Anexo 37 de *Verba*.
- García Gondar, Francisco (dir.) *et alii* (1995): *Repertorio bibliográfico da lingüística galega: Desde os seus inicios ata 1994 inclusive*. Santiago de Compostela: Centro de Investigacións Lingüísticas e Literarias "Ramón Piñeiro".
- García Santos, Juan Felipe (1996): "La investigación en sintaxis histórica", in A. Alonso González *et alii* (eds.): *Actas del III congreso internacional de historia de la lengua española (Salamanca, 22-27 noviembre 1993)*. Madrid: Arco Libros, vol. I, 293-300.
- Instituto da Lingua Galega (2007): *Memoria de actividades 2006*. Santiago de Compostela: Universidade.
- López Martínez, M^a Sol (1981): *O acusativo con preposición no galego medieval*. Tese de licenciatura. Universidade de Santiago.

- López Martínez, M^a Sol (1993): *O complemento directo con preposición a en galego*. Santiago de Compostela: Universidade. Anexo 36 de *Verba*.
- López Martínez, M^a Sol (2004): "Os verbos denominativos na lingua medieval: *chamar*", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade* (16-20 de setembro de 1996). Santiago de Compostela: ILG / Consello da Cultura Galega, vol. III, 195-208.
- Mariño Paz, Ramón (2002): "A lingüística histórica galega na actualidade", *Canal Estudos Galegos* 31; <http://www.vieiros.com/>
- Meilán García, Antonio José (1991): *La oración simple en la prosa castellana del siglo XV*. Oviedo: Universidad de Oviedo.
- Moscoso Mato, Eduardo (1992): *A pasiva en galego medieval. A construción "ser + participio" nos Miragres de Santiago*. Tese de licenciatura. Universidade de Santiago de Compostela.
- Moscoso Mato, Eduardo (2000): *Os tempos compostos no galego medieval*. Santiago de Compostela: Universidade. Anexo 46 de *Verba*.
- Moscoso Mato, Eduardo (2004): "Antepretérito sintético e analítico en galego medieval", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade* (16-20 de setembro de 1996). Santiago de Compostela: ILG / Consello da Cultura Galega, vol. III, 227-233.
- Moscoso Mato, Eduardo (2005): "Perífrases de infinitivo con *aver* e *ter* no galego medieval", in A. Boullón / X. L. Couceiro / F. Fernández (eds.): *As tebras alumeadas. Estudos filolóxicos ofrecidos en homenaxe a Ramón Lorenzo*. Santiago de Compostela: Universidade, 239-252.
- Mosteiro Louzao, Manuel (1999): *Las conjunciones de causa en castellano medieval. Origen, evolución y otros usos*. Santiago de Compostela: Universidade. Anexo 45 de *Verba*.
- Pérez Toral, Marta (1992): *Sintaxis histórica funcional del español. El verbo hacer como impersonal*. Oviedo: Universidad de Oviedo.

- Regueira Fernández, Xosé Luís (coord.) et alii (1996): *Guía bibliográfica de lingüística galega*. Vigo: Xerais.
- Ridruejo, Emilio (1994): "Sintaxis histórica", in *Actas del congreso de la lengua española (Sevilla, 7-10 octubre 1992)*. Madrid: Instituto Cervantes, 587-596.
- Rivas, Elena e M^a José Rodríguez Espiñeira (1997): *La cláusula en castellano medieval: constituyentes funcionales*. Santiago de Compostela: Universidade.
- Rodríguez Guerra, Alexandre (2000): "A xerarquía da animación do latín ó galego e ó portugués", in J. M. Carrasco et alii (ed.): *Actas del Congreso Internacional de H^a y Cultura en la Frontera*. Cáceres: Universidad de Extremadura, vol. II, 845-863.
- Rodríguez Guerra, Alexandre (2001a): "Conxuncións e locucións conxuntivas temporais en galego medieval", *Verba* 28, 207-258.
- Rodríguez Guerra, Alexandre (2001b): "Os esquemas causais en galego medieval", in X. L. Regueira / A. Veiga (eds.): *Da gramática ó diccionario. Estudos de lingüística galega*. Santiago de Compostela: Universidade. Anexo 49 de *Verba*, 225-247.
- Rodríguez Guerra, Alexandre (2004): "Verbo da orde relativa dos constituíntes oracionais non-clíticos do latín ó galego medieval", *Verba* 31, 151-186.
- Rodríguez Guerra, Alexandre (2005): "Parámetros (co)implicados na ordenación de constituíntes oracionais en latín", in G. M^a Rio-Torto et alii (coord.): *Estudos em Homenagem ao prof. Dr. Mário Vilela*. Porto: Fac. de Letras da Universidade do Porto, vol. I, 51-70.
- Royo, Guillermo (1978): *Cláusulas y oraciones*. Santiago de Compostela: Universidade.
- Royo, Guillermo e Tomás Jiménez Juliá (1989): *Fundamentos del análisis sintáctico funcional*. Santiago de Compostela: Universidade.
- Silva Domínguez, Carme (1993): *Os posesivos no galego moderno*. Tese de licenciatura. Universidade de Santiago de Compostela.

- Silva Domínguez, Carme (2001): "*Aver seu acordo, fazer sua oração: a participación do posesivo na construción de lexías complexas no galego medieval*", in X. L. Regueira / A. Veiga (eds.): *Da gramática ó dicionario. Estudos de lingüística galega*. Santiago de Compostela: Universidade. Anexo 49 de *Verba*, 249-64.
- Silva Domínguez, Carme (2002): *Frases nominais con posesivo en galego. Estructura e valores referenciais*. Santiago de Compostela: Universidade. Anexo 50 de *Verba*.
- Silva Domínguez, Carme (2004): "Procedementos de determinación no galego medieval. Estudio dos determinantes na *Crónica Troiana*", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade (16-20 de setembro de 1996)*. Santiago de Compostela: ILG / Consello da Cultura Galega, vol. III, 277-298.
- Silva, Rosa Virgínia Mattos e (1989): *Estruturas trecentistas. Elementos para uma gramática do Português Arcaico*. Lisboa: Imprensa Nacional-Casa da Moeda.
- Soto Andión, Xosé (2001): "Unha posible base de datos para analiza-la *intransitividade* no galego actual", in X. L. Regueira / A. Veiga (eds.): *Da gramática ó dicionario. Estudos de lingüística galega*. Santiago de Compostela: Universidade. Anexo 49 de *Verba*, 279-291.
- Suárez Fernández, Mercedes (1996): *El complemento predicativo en castellano medieval. (Época prealfonsí)*. Santiago de Compostela: Universidade. Anexo 42 de *Verba*.
- TMILG = Xavier Varela Barreiro (dir.) (2004-): *Tesouro Medieval Informatizado da Lingua Galega*. Santiago de Compostela: Instituto da Lingua Galega. [<http://ilg.usc.es/tmilg>].
- Varela Barreiro, Xavier (1981): *Contribución ó estudio dos cuantificadores en galego medieval*. Tese de licenciatura. Universidade de Santiago de Compostela.
- Varela Barreiro, Xavier (1992): *Os cuantificadores en galego moderno. A cuantificación pronominal*. Santiago de Compostela: Universidade. Tese de doutoramento en microficha.

- Varela Barreiro, Xavier (2004): "Un proxecto do ILG no abalo da gramática histórica da lingua galega", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade* (16-20 de setembro de 1996). Santiago de Compostela: ILG / Consello da Cultura Galega, vol. II, 649-684.
- Vidal Meixón, Ana (2004): "Aproximación ás causais e consecutivas en galego medieval", in R. Álvarez / F. Fernández / A. Santamarina (eds.): *Actas do I Congreso Internacional A Lingua Galega: Historia e Actualidade* (16-20 de setembro de 1996). Santiago de Compostela: ILG / Consello da Cultura Galega, vol. III, 317-327.

AVALIACIÓN DUN ETIQUETADOR AUTOMÁTICO ESTATÍSTICO PARA O GALEGO ACTUAL: XIADA

EVA DOMÍNGUEZ NOYA

Centro Ramón Piñeiro para a Investigación en Humanidades
edomin@cirp.es

FCO. MARIO BARCALA RODRÍGUEZ

Centro Ramón Piñeiro para a Investigación en Humanidades e Universidade de
Vigo
fbarcala@cirp.es

MIGUEL ÁNGEL MOLINERO

Universidade da Coruña
mmolinero@udc.es

RESUMO

Neste traballo avaliamos, dende o punto de vista lingüístico, un etiquetador automático estatístico, desenvolto conxuntamente polo Centro Ramón Piñeiro para a Investigación en Humanidades e o Grupo COLE das Universidades de Vigo e A Coruña, destinado a etiquetar os documentos do *Corpus de Referencia do Galego Actual* co obxecto de proporcionar recursos e ferramentas para a análise lingüística computacional do galego actual.

PALABRAS CHAVE

Corpus, galego, etiquetación, lematización, avaliación.

ABSTRACT

"Evaluation of an statistical automatic labelling machine for present Galician: XIADA"

We evaluate, from a linguistic point of view, a statistical automatic labelling machine, which is explained together by the

Center Ramón Piñeiro on Humanities Research and the COLE Group of Vigo and La Coruña Universities, and which also set aside for labelling the papers of Present Galician Reference Corpus so as to provide tools and resources for the computational linguistic analysis of Present Galician.

KEYWORDS

Corpus, Galician, tagging, lemmatization, evaluation

1. Introducción

1.1. O corpus

O *Corpus de Referencia do Galego Actual* (CORGA)¹ é un corpus documental que en decembro de 2008 contén 23 millóns de formas, integrado por distintos tipos de textos, -xornais, semanarios, revistas, ensaios e textos de ficción (novela, relato curto e teatro)-, que abrangue temporalmente dende o ano 1975 ata a actualidade.

Todos os documentos do corpus están codificados segundo o estándar XML² (*eXtensible Markup Language*) e cada un deles está constituído por unha *cabeceira* e un *corpo*. Na *cabeceira* incluimos a información bibliográfica pertinente xunto cunha caracterización temática, e no *corpo* aparece o texto estruturado formalmente segundo o tipo de documento de que se trate. Por exemplo, o XML permítenos considerar un xornal un único arquivo XML que está organizado en múltiples noticias as cales, á súa vez, conteñen obrigatoriamente un corpo e opcionalmente un titular, resumo e/ou pé de foto. A maiores, cada un destes elementos está constituído por parágrafos (texto comprendido entre dous puntos e á parte) e estes son segmentados en oracións (secuencia textual separada do resto do texto por un punto, punto e coma, dous puntos, etc.). Esta disposición deta-

¹<http://corpus.cirp.es/corga>

² <http://www.w3.org>

llada posibilita realizar consultas sobre a totalidade da noticia ou sobre unha unidade estrutural concreta (titular, resumo, pé de foto ou corpo). Ademais, ao ter asignada como mínimo unha área temática que caracterice cada noticia, é posible delimitar os ámbitos nos que unha palabra ou expresión se emprega.

O formato XML facilitou diferenciar formalmente os distintos tipos de textos que integran o corpus pero tamén permitiu a marcación de fragmentos que aparecen nunha lingua distinta ao galego -deste xeito evitamos indexalos no sistema de buscas e impedimos que engorden os datos totais do corpus no referente a número de palabras-; marcar a presenza de poemas e de cada un dos seus versos³; marcar a existencia de entrevistas e as intervencións dos distintos interlocutores⁴ ou sinalar a aparición dun determinado vocábulo nunha gráfica ou nunha táboa, etc.

Dende o 2007 o sistema de buscas do CORGA que está dispoñible na rede incorpora unha nómina de autores e obras que lle permite ao usuario:

1) Obter datos específicos sobre a estrutura textual do corpus. Para iso pódese empregar un ou varios criterios de busca, sendo posible combinar estes ou os seus valores segundo as necesidades do usuario: autor, título da obra, área temática, ano, etc.

2) Obter a frecuencia normalizada dunha ocorrencia fronte á global; deste xeito poden obterse datos relativos á frecuencia dunha expresión de busca con respecto ao subconxunto de documentos sobre os que se realiza a consulta.

³ Decidimos non incluír no deseño do CORGA poemarios, porén non é infrecuente a aparición dalgún poema, sobre todo en ensaios literarios e noticias de opinión e cultura. A caracterización de certas oracións ou fragmentos como pertencentes a un poema pódelle servir ao usuario para desbotar ou explicar certas ocorrencias.

⁴ Non só nas entrevistas e coloquios senón tamén nas obras teatrais remítimos cada alocución ao personaxe ou interlocutor que lle corresponde. Deste xeito, os nomes non suman voces para o cómputo de formas totais e proporcionámoslle ao usuario información que lle pode ser útil na súa consulta (por exemplo, determinadas ocorrencias poden aparecer só nun interlocutor).

Conscientes de que para realizar consultas máis avanzadas e dar un salto cualitativo nas posibilidades de busca é fundamental que os textos do CORGA estean lematizados e etiquetados, paralelamente á construción do corpus e á incorporación a este de novos documentos, no Centro Ramón Piñeiro para a Investigación en Humanidades estase desenvolvendo unha ferramenta denominada *Etiquetador e lematizador do galego actual*⁵ cuxos resultados iniciais queremos presentar aquí. Con ela pretendemos asignarlle a cada unidade léxica o lema e etiqueta que lle corresponde segundo o contexto no que se localice.

Dende o punto de vista lingüístico, para a construción do analizador e etiquetador automático foi necesario seguir os pasos seguintes: determinar un sistema de etiquetas, deseñar a estrutura dun lexicón e construílo en consonancia con esta e elaborar certas regras lingüísticas que axudasen no recoñecemento e segmentación das distintas unidades léxicas.

1.2. O conxunto de etiquetas

O etiquetario ou *tagset* co que traballamos foi desenvolvido no marco do proxecto *Construción, etiquetación e lematización do Corpus de Referencia do Galego Actual* como unha ferramenta para a análise computacional e a anotación automática dos textos do CORGA. Para a súa elaboración seguimos as recomendacións de EAGLES [Leech e Wilson, 1994] e os textos básicos da lingüística galega⁶. Así, este sistema de etiquetación presenta unha estrutura xerárquica na que se identifica, no primeiro nivel, a categoría morfolóxica, e nun segundo, os atributos gramaticais pertinentes que constitúen o conxunto básico de atributos para cada categoría.

En primeiro lugar delimitamos as clases de palabras existentes en galego e, a continuación, establecemos os atributos gramaticais pertinentes que permiten o recoñecemento mor-

⁵ Ferramenta á que denominamos XIADA. Véxase <http://corpus.cirp.es/xiada>

⁶ Véxase o apartado de Bibliografía complementaria.

folóxico de calquera palabra. Primamos a descrición morfolóxica e reducimos a información sintáctica á caracterización de determinadas categorías cuxos elementos integrantes poden funcionar como determinantes, non determinantes, nucleares ou modificadores.

Como se pode observar no etiquetario⁷ de XIADA que reproducimos na táboa 1, na primeira columna recóllense as diferentes clases de palabras; nas seguintes columnas recóllense os atributos e os valores destes e, por último, nas celas incorpórase un número que indica a posición que o atributo, se é pertinente, ocupa dentro da etiqueta final.

A distribución dos atributos dentro de cada categoría é simétrica, de tal xeito que todos os elementos pertencentes a unha categoría dada posúen o mesmo número de atributos e os seus valores ocupan a mesma posición dentro da cadea resultante. Con todo, non todos os elementos pertencentes a unha mesma categoría morfolóxica responden positivamente a todos os atributos que a caracterizan. Así, na categoría verbal, o atributo xénero só é un trazo pertinente para a caracterización do participio. Para o resto de elementos do paradigma verbal a aplicación dese atributo non é pertinente. Por este motivo creamos un valor "0" co significado de "Non aplica". Hai tamén ocasións nas que a atribución dun valor inequívoco é imposible polo que debemos neutralizar o valor dalgún ou de varios atributos. En XIADA marcamos esta ambigüidade co valor "a" que representará o xénero, número, persoa, posuidor ou valor dependendo da clase de palabra e da posición que ocupe na cadea de etiquetaxe.

Combinando as clases de palabras delimitadas e os atributos que as caracterizan, este sistema de etiquetación presenta arredor de 400 etiquetas posibles aínda que polo momento no subcorpus desambiguado só constatamos 375 distintas.

⁷ Disponible en <http://corpus.cirp.es/xiada>

1.3. O lexicón

Unha vez tivemos claro cal ía ser o noso etiquetario necesitabamos un lexicón onde almacenar a información de etiqueta e lema que lle correspondía a cada palabra.

Dado que a introdución e xestión dos datos debía ser práctica, estimamos que maioritariamente as entradas do lexicón tiñan que ser caracterizadas morfoloxicamente desagregándoas en *lema*⁸, *raíz*⁹, *subetiqueta*¹⁰ e *grupo de derivación*¹¹. Esta estrutura esixiu un estudo formal detallado dos integrantes de cada unha das clases gramaticais variables para agrupalos en modelos e deste xeito construír a gramática formal que nos permitise analizar e tamén xerar as formas flexionadas e conxugadas do galego moderno.

O lexicón de XIADA, ademais de conter a información morfolóxica que permite a identificación e caracterización morfolóxica plena de calquera palabra galega, responde aos pará-

⁸ É o representante das distintas formas flexionadas ou conxugadas que se integran nun paradigma. No noso sistema a clase morfolóxica é un elemento constitutivo do lema polo que todo lema está asociado a unha clase de palabra. Así, para "galego", por exemplo, teremos dous lemas: un, asociado á categoría substantivo e outro á de adxectivo.

⁹ Basicamente podemos considerar que a raíz é o lema sen morfemas gramaticais. A raíz e o lema coinciden nas entradas das clases invariables e nas dos substantivos e adxectivos que só presentan flexión no número. En ningún caso se pode identificar esta raíz co lexema da gramática tradicional posto que nós, para o seu establecemento, non temos en conta os morfemas derivativos senón só os flexivos.

¹⁰ Caracterización morfolóxica básica da entrada no lexicón.

¹¹ Conxuntos de terminacións para as que se proporcionan uns valores.

Por exemplo o grupo G1 consta de:

o	-> masculino singular
a	-> feminino singular
os	-> masculino plural
as	-> feminino plural

Deste xeito, para todos os substantivos e adxectivos que teñen flexión de xénero e forman o plural engadindo -s introducimos só o lema e a raíz correspondente, sen necesidade de ter que inserir todas as formas do paradigma.

metros seguintes: indicación sobre a normativa oficial, ámbito léxico e procedencia da forma:

a) Normatividade. Para un correcto recoñecemento das formas existentes nos textos reais contemporáneos galegos necesitamos que o lexicón sexa amplo e que non só estea integrado por lemas normativos. Se este fose o caso, moitos dos documentos do CORGA non poderían ser etiquetados ou serían a medias. É fundamental, polo tanto, darlles cabida tamén a dialectalismos, hiperenxebrismos, castelanismos, etc. porque estes aparecen nos textos. Debemos, así a todo, poder diferenciarlos e para iso introducimos este criterio na caracterización de lemas, formas e incluso algunha desinencia.

b) Tipo de léxico. Cremos que podía ser útil diferenciar na estrutura do lexicón módulos segundo o tipo de léxico que conteñan que poidan ser combinados. Para iso dispuxemos no deseño da estrutura, e obviamente implementamos na construción do lexicón, dun campo máis no que se caracteriza o lema como pertencente ao léxico común ou ao técnico-científico, especificando neste último o ámbito no que se clasifica: administración, economía, medicina, etc.

c) Fonte. Nos textos actuais preséntanse formas que non aparecen nos dicionarios-vocabularios preceptivos da lingua galega ben porque son termos técnicos para os que se acaba de propoñer unha denominación ben porque se documentan cun uso categorial distinto, non obstante, a súa introdución no lexicón é imprescindible para o recoñecemento e caracterización dos textos. Coa indicación da fonte achegamos datos concretos sobre a localización do lema e a fiabilidade da forma.

Recapitulando, o lexicón de XIADA, ademais de conter a información morfolóxica que permite a identificación e caracterización morfolóxica plena de calquera palabra galega actual, proporciona información sobre a normativa oficial, o ámbito léxico e a procedencia da forma¹².

¹² Así, por exemplo, cando introducimos unha unidade como "galego" no lexicón temos que proporcionar a seguinte información:

Unha vez deseñada a estrutura do lexicón, implementámolo coas entradas do *Vocabulario Ortográfico da Lingua Galega* (VOLG) e coas 100 000 formas máis frecuentes do CORGA, o que se traduce en algo máis de 55 000 lemas.

Na seguinte táboa amosamos a información correspondente ás tres grandes clases de palabras variables -substantivos, adxectivos e verbos- sobre o número de lemas que figura no lexicón xeral de XIADA e a cantidade de formas que é posible xerar con eses lemas. Debemos aclarar que as formas dos diminutivos, aumentativos ou superlativos sintéticos regulares son remitidos ao lema base, non conforman un lema á parte. Así, “gravísimo” ou “caladiños” están asociados respectivamente aos lemas “grave” e “calado”. Polo que se refire ao cómputo das formas, nas verbais, non están incluídas as posibles combinacións con pronomes enclíticos.

CATEGORÍA	LEMAS	FORMAS
Substantivos	29 806	64 646
Adxectivos	13 715	45 449
Verbos	6 779	461 185

1.4. O preprocesador - etiquetador

Lingüísticamente, unha vez elaborados os modelos formais que acollen as raíces e grupos de derivación e implementado o lexicón computacional, estamos en disposición de etiquetar automaticamente calquera texto galego contemporáneo. Trátase de executar o programa que o equipo informático do

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría- lema	normativa- lema	fonte
xeral	principal	galeg	Sc	G1	galego	S	si	volga 2004
xeral	principal	galeg	A0	G1	galego	A	si	volga 2004

proxecto está desenvolvendo para traballar con arquivos codificados en XML, formato no que están os textos do CORGA.

Para proceder a etiquetar un documento, primeiro é necesario delimitar as distintas unidades léxicas que este contén, tarefa que lle corresponde ao preprocesador. Aínda que poida parecer sinxelo, non é tarefa trivial indicarlle a un programa como segmentar formas verbais con pronomes enclíticos ou separar os formantes dunha contracción. Seguimos traballando para que cometa menos erros pero é inevitable que nalgúns casos de ambigüidades falle de vez en cando. É sobre todo un traballo informático pero no que o equipo lingüístico participa coa construción de regras que gobernan o comportamento do preprocesador. Por exemplo, a caracterización como pronome átono acusativo de terceira persoa para "no, na, nos, nas" só será válida se vai enclítico a unha forma verbal rematada en ditongo.

No noso caso, ademais, despois de delimitar as unidades léxicas presentes nos textos asígnalle a cada unha os lemas e etiquetas posibles. Así, a saída do documento preprocesado proporcionaranos o texto analizado contendo para cada palabra gráfica ou *token* todas as posibles análises morfolóxicas. Non obstante, para realizar unha consulta nun corpus etiquetado non necesitamos todas as análises potenciais dunha forma senón só a que presenta en cada contexto, tarefa esta que realiza o etiquetador escollendo a etiqueta e o lema correctos.

Para poder empregar con certas garantías o etiquetador automático estatístico, necesitamos adestralo desambiguando manualmente un subconxunto de textos. O subcorpus escollido para o adestramento inicial foi extraído do *Corpus de Referencia do Galego Actual* e constou de 200 000 palabras procedentes de noticias xornalísticas do ámbito económico.

Sabemos que as probabilidades de acerto do etiquetador aumentan se o corpus de adestramento é amplo, pero tamén estamos seguros de que os resultados positivos se incrementarán en proporción á riqueza e variedade temática do subcorpus desambiguado manualmente, de xeito que o noso obxectivo máis próximo é desambiguar á man unhas 200 000 formas para

cada un dos grandes bloques do CORGA: prensa, ensaio e ficción.

A mediados do 2008 iniciamos unha segunda fase de desambiguación manual dun subcorpus, extraído tamén do CORGA, que estaba constituído polo número 1166 do ano 2005 do semanario *A Nosa Terra* e polas noticias da versión electrónica do día 17 de marzo de 2007 do xornal *Galicia Hoxe*. En total, 58 766 palabras.

Durante a revisión e corrección da versión etiquetada automaticamente do semanario apreciamos unha alta porcentaxe de acerto na etiquetación, motivo polo que valoramos a eficacia do etiquetador co seguinte documento escollido para desambiguar. As 89 noticias e as 22 212 palabras das que consta o arquivo GH2007-03-17.xml son suficientes para realizar unha proba significativa sobre a efectividade do etiquetador empregado.

Antes de proceder a expoñer os resultados obtidos, queremos chamar a atención sobre dous feitos que consideramos fundamentais para valorar na súa xusta medida o etiquetador utilizado:

1) As porcentaxes de acerto que presentan os etiquetadores existentes para unha lingua como o castelán ou o inglés -para o galego non hai estatísticas- son do 95%/97% [Graña, 2000: 137-165]. Con todo, debe terse en conta que nos resultados que presentan esoutros traballos inclúense soamente erros de etiquetación -non de lematización-, e, ademais, en ningún caso eses etiquetadores realizan segmentacións das unidades léxicas presentes no texto. Estes dous factores, non lematización e non segmentación textual, cómpre telos presentes na comparativa final.

2) O tamaño do etiquetario que empregamos -arredor de 400 etiquetas distintas- é bastante máis grande ca o empregado pola maior parte de etiquetadores. Segundo Jurafsky [Jurafsky, 2000: 296-297], os etiquetarios máis usados para o inglés son o

*Penn Treebank*¹³ que consta de 45 etiquetas e é o que se usou no corpus *Brown*¹⁴; o C5 de 61 etiquetas empregado polo programa CLAWS¹⁵ para etiquetar o *British National Corpus* (BNC)¹⁶ e o C7, de 146 etiquetas, utilizado para etiquetar un subconxunto do BNC.

Só uns poucos sistemas de etiquetación teñen un etiquetario equiparable en tamaño ao que aquí presentamos [Graña, 2000: 139 e Sampson, 1994] (Véxase táboa 1).

2. *Avaliación da etiquetación automática realizada sobre as 22 212 formas gráficas das noticias do xornal GH2007-03-17*

Cómpre advertir que fomos rigorosos na análise efectuada e recollemos e clasificamos todos aqueles casos nos que a solución proporcionada polo etiquetador automático non coincidiu coa solución escollida polo lingüista revisor. Tamén é preciso aclarar que os erros cometidos polo etiquetador debido á presenza de erratas, faltas de acentuación ou outros erros lingüísticos, existentes todos eles nos documentos orixinais, foron recollidos e rexistrados no apartado que lles correspondía segundo o tipo de fallo detectado.

Todos os erros que se produciron no documento etiquetado automaticamente intégranse nalgunha das epígrafes seguintes:

a) Non se lle asigna lema á unidade léxica aínda que si se propón unha caracterización morfolóxica. Non se pode, con todo, responsabilizar ao etiquetador dos fallos que se producen por non asignar lema -de momento non ten tarefas específicas de lematización de formas descoñecidas, simplemente, para cada forma, ten asociado un lema a unha etiqueta-. Os erros ti-

¹³ Véxase <http://www.cis.upenn.edu/~treebank>

¹⁴ http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html

¹⁵ <http://ucrel.lancs.ac.uk/claws/>

¹⁶ <http://www.natcorp.ox.ac.uk>

pográficos, presentes nos documentos analizados, que provocan a aparición de palabras inexistentes contabilizáronse neste apartado.

b) Prodúcese unha segmentación errónea á hora de separar os constituíntes da unidade.

c) Caracterízase erroneamente un elemento (*token*) cun atributo que non lle corresponde.

d) Érrase ao clasificar a unidade categorialmente.

e) Engánase ao atribuír o lema.

f) Non se proporciona a análise morfolóxica axeitada porque esta non figura no lexicón. O lematizador só lle asigna lema a aquelas formas que están implementadas nos lexicóns cos que traballa e para que o etiquetador proporcione unha análise determinada dunha forma gráfica calquera é preciso que na gramática formal que este emprega estea recollida dita análise.

g) Non se produce un recoñecemento morfolóxico atinado porque o texto orixinal presenta erros ortográficos ou lingüísticos, entendendo por estes últimos, basicamente, desviacións da norma de tipo morfosintáctico.

Un dos problemas que atopamos ao traballar con textos reais é a presenza, en bastantes ocasións, de formas ou construcións morfosintácticas que non seguen a normativa oficial¹⁷, por exemplo, en canto ao xénero dos substantivos, ao uso do til diacrítico ou ao emprego dos pronomes átonos. Outro dos problemas derivados de non manipular o texto orixinal xórdenos coa presenza dos erros tipográficos que dan lugar a unha unidade -imposible nese contexto- coincidente con outra xa existente no lexicón. Pois ben, estes dous tipos de erro provocan que unha unidade, á que lle correspondería unha etiqueta *x*, coincida no lexicón formalmente con outra unidade cuxa etiqueta é *z*, polo que o etiquetador non vai acertar coa análise.

Na táboa 2 (véxase final do artigo) amosamos todos os erros detectados na etiquetación automática das 22 212 formas

¹⁷ Nunhas ocasións percíbese que ese uso é intencional e noutras, que é froito do descoñecemento do galego normativo.

gráficas das que consta a versión electrónica do xornal *Galicia Hoxe* do día 17 de marzo de 2007 introducido en CORGA, distribuídos por sección e organizados segundo a anterior clasificación.

Só en tres seccións -Cultura-Sociedade, Opinión e Televisión- a porcentaxe de acerto descende do 94,5% ata situarse no seu punto máis baixo no 92,1%. Unha das razóns que explican esta diferenza é a abundancia de nomes propios multpalabra -sobre todo na sección de TV, títulos de películas, series, etc.- que non son correctamente segmentados e, en consecuencia, provocan unha concatenación de erros.

Outro dos factores diferenciais radica no estilo e lingua propios das noticias de cultura e opinión que se asemellan máis á narración que á crónica xornalística. É probable que cando dispoñamos do corpus de adestramento correspondente á ficción, se o empregamos para etiquetar este tipo de noticias, vexamos unha melloría considerable¹⁸.

A continuación examinaremos con detalle algúns dos erros cometidos polo etiquetador segundo a clasificación antes proposta. Comezaremos polos que non son achacables ao programa.

2.1. Non se lle asigna lema á unidade léxica

Os *tokens* incluídos baixo esta epígrafe -122 en total que representan o 0,54% do total- presentan o lema baleiro aínda que son etiquetados correctamente na maioría dos casos. Esta circunstancia dáse porque, a pesar de que o lematizador-etiquetador non é capaz de asignar un lema se este non figura no lexicón ben por omisión, ben por ser errata, castelanismo, estranxeirismo

¹⁸ O etiquetador que utilizamos foi adestrado cun corpus xornalístico de 200.000 palabras da temática económica formado, sobre todo, por noticias das seccións de Economía, Galicia, España e Internacional. A nosa idea é constituír un subcorpus desambiguado manualmente por cada xénero presente no CORGA de xeito que os poidamos combinar e escoller aquel que máis se adecúa ao tipo de texto que queremos etiquetar automaticamente.

mo, etc., si é capaz de aventurar unha caracterización morfolóxica tendo en conta o contexto no que ese *token* se insire.

O etiquetador non é quen de detectar nin de corrixir unha errata -tampouco é ese o seu cometido- polo que ante os erros tipográficos cometidos polo(s) autor(es) da(s) noticia(s) vaise comportar como se non o fosen e vai analizar o texto que se lle proporciona. Posto que un acento mal colocado ou unha letra que sobra ou falta converten a palabra pretendida noutra, existente ou non na nosa lingua, o lematizador-etiquetador vai actuar de modo distinto segundo o caso. Se a unidade resultante coincide con outra que teña no lexicón, outórgalle a etiqueta e o lema que a el lle consta; se non a atopa, non lle vai asignar ningún lema. Por este motivo algunhas erratas, como se poderá observar nos exemplos, foron catalogadas neste apartado.

Nas seguintes oracións o contexto determina que a unidade que destacamos subliñándoa fose analizada correctamente:

<expresión>Tahoces acusou a Xunta de “crear inseguridade xurídica” e de “atentar contra as regras do xogo”, xa que considera que o ofrecemento de presenza pública “actuou como criterio convalidante” de propostas que “non cumprían outros requisitos”.</expresión> [A0ms] [GH2007-03-17/19; Galicia]

<expresión>Non obstante, o país está envolto nunha “pseudo” guerra civil por mor da violencia interconfesional con atentados terroristas a diario e máis de 650.000 iraquís mortos en catro anos.</expresión> [A0fs] [GH2007-03-17/70; Internacional]

<expresión>Algo vai mal cando unha sociedade se infantiliza ata estes extremos e chamamos para todo a papa Estado (ou papa Xuíz).</expresión> [Vpi30s]; [GH2007-03-17/86; Opinión]

Algunhas outras formas descoñecidas para o lematizador pero ben analizadas son as seguintes: *chanchullos*; *vertebrador*; *reapertura*; *promedia*; *enfaticizou*; *galescolas*; *xugosos*; *PDC*; *desbaldimento*; *vozarrón*; *comisión-homenaxe*; *catástrofe*; *emprendemento*; *FEFN*; *INWA*; *pedofilia*; *pedófilo*; *chats*; *armonización*; *bici*; *ornanismos*; *linguística*; *eúscaro*; *viceconselleiro*; *plurilinguismo*; *anfitrón*; *NOAA*; *existe*; *catoliscismo*; *RG*; *blog*; *histórico-contemporánea*; *performer*, etc.

No entanto, equívocase en formas como *garantista*, *CIG-Ensino*, *reponsables*, *FEFN*, *Réplica-rede*, *calatán*, *plurilingue*, *monolingue*, *biodiversidad*, *represaliados*, *socio-cultura*, *PS3*, *SETI@Home*, *XrossMediaBar*, *pasquines*, etc.

2.2. Erros ortográficos, lingüísticos ou desviacións morfolóxicas da norma existentes nos textos orixinais

A incidencia deste apartado é das menores -12 casos que representan o 0,05% dos erros totais-. Con todo, para ser rigorosos, debemos reflectila aínda que, máis tarde, cando sexa o momento de avaliar verdadeiramente o etiquetador, deberemos eliminar estas análises do seu debe pois son erros do autor textual, non da ferramenta.

Baixo esta epígrafe recollemos só os fallos ortográficos ou lingüísticos cometidos polos autores das noticias que provocan que, por coincidencia gráfica con algunha forma existente no lexicón, o etiquetador actúe de xeito erróneo. Non nos referimos ás grallas ortográficas que causan a aparición de novas unidades léxicas (*reponsables* é claramente un erro de teclado por *responsables*) senón soamente a aqueles erros ortográficos ou desviacións morfosintácticas da norma que fan que o *token* resultante coincida na súa grafía con un xa existente no lexicón pero para o cal se lle outorgan análises diferentes.

De todos eles, o que máis chama a atención é o que, por partida dobre, se produce nunha noticia da sección *Campus*, en concreto na noticia GH2007-03-17/29.etq.xml do CORGA:

<expresión>Así, en Infantil matricularanse uns 22.000 nenos, que abranguen unhas 14.000 familias; uns 2.140 en Primaria, dunhas 1.430 familias; uns 13.260 niso, dunhas 8.800 familias; e uns 5.110 en Bacharelato, de 3.400 familias.</expresión>

<expresión>Os prazos para formalizar a matrícula prolongaranse do 20 ao 30 de xuño en Infantil e Primaria, e do 25 de xuño ao 10 de xullo niso e Bacharelato.</expresión>

O noso etiquetador segmenta e caracteriza ben os dous elementos integrantes da contracción pero, como é lóxico, non

consegue dar coa análise correcta. Obviamente, quería-se facer referencia á etapa educativa ESO pero a casualidade de que antes houbera unha preposición "en" e logo a sigla ESO provoca que a secuencia "en ESO" se convirte na contracción da preposición e mais o demostrativo neutro, resultado probable da aplicación dun tradutor ou un corrector automático.

O etiquetador tampouco consegue atinar cando o xénero do substantivo empregado difire do que a el lle consta no lexicon:

<expresión>Despois tivo lugar a mesa redonda no Ateneo da capital ourensá, presentada polo humorista Xose Lois González O Carrabouxo, quen asegurou que se trata de homenaxe "necesario e bo", porque "aínda a calor da presenza de Peña na rúa do Paseo non se foi".</expresión> [GH2007-03-17/26; Galicia]

cando se emprega unha forma singular por unha plural:

<expresión> "Hai unha pregunta que o presidente do Goberno debe responder porque é o que lle interesa aos navarros:</expresión> [GH2007-03-17/63; España]

ou cando, en vez dun substantivo, se utiliza unha unidade que formalmente se integra no paradigma verbal:

<expresión> Despois de todo o mundo policial é un xenero en si mesmo, o chamado thriller e ata ten entidade propia no mundo das letras con ese xénero chamado novela detectivesca pero... ¿o dos médicos?</expresión> [GH2007-03-17/89; TV]

2.3. Non figura no lexicon con ese valor

O 0,47% dos erros -106 dos 1260 totais- débense a que o etiquetador non proporciona entre as alternativas posibles a etiqueta que lle corresponde ao *token* segundo o contexto no que está inserido. En realidade, non deberiamos contabilizar os fallos recollidos baixo esta epígrafe no debe do etiquetador -senón no do equipo lingüístico- posto que se no lexicon non consta unha determinada análise o etiquetador non nola pode proporcionar.

Estas deficiencias morfolóxicas son emendables, basta con implementar para cada unha das entradas afectadas as características formais das análises das que carece.

Vexamos a continuación como estas carencias inciden en diversos aspectos da caracterización global.

Afecta á identificación da clase de palabra en casos como:

<expresión>A Consellería de Medio Rural cuadruplicará este ano os cursos para mellorar a preparación dos máis de 4.600 efectivos do operativo de extinción de incendios forestais, co obxectivo de mellorar a súa eficacia, despois de que o propio departamento recoñecese que a actual formación era "insuficiente e deficitaria". [analizouno como adxectivo no sitio de substantivo] [GH2007-03-17/1; *Galicia*]

<expresión>A creación do Cegadi e da figura do asistente persoal, estratexias pioneiras</expresión> [considerounos substantivos en vez de sigla e adxectivo, respectivamente] [GH2007-03-17/32; *Cultura-Sociedade*]

<expresión>Este foro internacional permitiu tamén recordar que a explotación sexual infantil está moi ligada á situación de desigualdade da que seguen sendo vítimas moitas mulleres en todo o mundo.</expresión> [Identificouno como participio en vez de adxectivo] [GH2007-03-17/37; *Cultura-Sociedade*]

atinguelles aos atributos:

<expresión>No entanto, admitiu que esta modernización non se levará a cabo "na súa totalidade" este ano, cun sistema de "alta eficiencia".</expresión> [Clasificouna como locución conxuntiva en lugar de adverbial] [GH2007-03-17/6; *Galicia*]

<expresión>Máis de 650.000 iraquís faleceron catro anos despois da famosa foto de Bush, Blair e Aznar que marcou o inicio do conflito polas "armas de destrución masiva".</expresión> [Tratouno como adverbio nuclear en vez de modificador] [GH2007-03-17/70; *Internacional*]

repercute no lema:

<expresión>A produtora do inimitable "Loco de la colina" alega que non poderá seguir "desenvolvendo a súa profesión" por

unha doenza</expresión> [Remitiuno ao lema “produtor” en vez de a “produtora”] [GH2007-03-17/87; *Opinión*]

e finalmente pode tamén implicar a segmentación, a etiquetación e o lema da unidade, como nos dous exemplos que seguen¹⁹:

<expresión>O Consello de Ministros deu onte o visto e prace a obras de urxencia, por valor de case 1,5 millóns de euros, que foron aprobadas polo Ministerio de Medio Ambiente para restaurar zonas afectadas polos incendios forestais que se produciron o pasado verán.</expresión> [GH2007-03-17/3; *Galicia*]

<expresión>O descubrimento dunha placa conmemorativa nunha céntrica rúa de Ourense e unha mesa redonda centraron onte os actos conmemorativos en honra ó nacionalista ourensán Xosé Enrique Rodríguez Peña, que faleceu o verán pasado e que foi ex conselleiro no goberno tripartito de Fernando González Laxe, ex deputado e ex candidato do BNG á alcaldía ourensá.</expresión> [GH2007-03-17/26; *Galicia*]

2.4. Segmentacións erróneas

Os 228 casos recollidos -1,02%- converten este apartado no terceiro máis deficiente do etiquetador. Porén, se lembramos a existencia das contraccións, das formas verbais con pronomes enclíticos, os nomes propios e numerais multipalabra e, sobre todo, a grande cantidade de locucións que presenta o galego, podemos afirmar que esa porcentaxe é irrisoria. Con todo, a medida que o corpus de adestramento se vaia incrementando, dispoñamos dunha gramática formal de propios e dunha regra de construción dos numerais multipalabra irán tamén diminuindo as segmentacións fallidas.

Como cremos que é interesante e que dá unha idea bastante aproximada da valía deste etiquetador, deseguido clasifi-

¹⁹ Consideramos que “visto e prace” e “mesa redonda” son substantivos multipalabra pero ata o momento non introducimos no lexicón nin estudamos a gramática formal de máis unidades multipalabra que as locucións.

caremos e comentaremos as diferentes clases de fallos segmentais que se producen.

2.4.1. *Con numerais*

Ante un numeral multipalabra o preprocesador consulta o lexicón, comproba se algunha das regras introducidas lle afecta e, ao ver que non, segmentao en tantas unidades como numerais distintos haxa implicados. É dicir, para un caso como o seguinte, en vez de identificar o numeral "vinte e dous", considera que son dous numerais coordinados pola conxunción "e" e ofrece unha análise para "vinte" e outra para "dous"²⁰.

<expresión>Un mozo de vinte e dous anos, residente en Dos Hermanas (Sevilla), levou ó seu pai ó Xulgado porque este se negaba a subirlle a súa paga para extras, que chegaba ós 150 euros ó mes (a familia xa lle cubría ademais os gastos básicos).</expresión> [GH2007-03-17/86; *Opinión*]

2.4.2. *Con propios*

De entrada, queremos indicar que ata agora no lexicón de XIADA non hai nomes propios nin nos documentos do CORGA se marcan de ningún xeito, polo que o etiquetador só dispón dunha pequena lista externa de antropónimos e topónimos e dunha serie de propios que foron detectados automaticamente procesando os textos, extraéndoo a partir de unidades que aparecen en maiúscula no interior das oracións. Temos previsto modificar e completar o módulo dos propios pero polo momento non acometemos esta tarefa.

A maioría dos erros producidos neste apartado débense á presenza de elementos de enlace -preposicións, contraccións, conxuncións e signos de puntuación- no interior do propio multipalabra que provocan que a unidade se rompa ante a súa aparición. Vexamos algúns exemplos:

Asociación pola Defensa Ecolóxica de Galicia

²⁰ Xa se fixeron progresos para que o sistema reconeza os numerais multipalabra, polo que en versións posteriores do analizador apreciarase unha melloría considerable neste apartado.

George W. Bush

Creación e Difusión Cultural

Pacto Local en Galicia; xestión dos municipios no futuro

Outro dos factores que impiden unha correcta identificación e caracterización do propio é o formato no que este aparece. Se as iniciais do propio multpalabra están en maiúsculas e non hai elementos do apartado anterior, o recoñecemento é atinado pero se só presenta maiúsculas o primeiro elemento, a segmentación, e en consecuencia a etiquetación, vai ser defectuosa. Exs.:

Lei orgánica de educación

Un conto de inverno

Lei orgánica para a igualdade efectiva de mulleres e homes

2.4.3. Con contraccións

Nas dúas secuencias seguintes recolleemos os dous únicos casos do subcorpus avaliado nos que o etiquetador erra na segmentación do token “á”, e en consecuencia na súa caracterización morfolóxica. Considera que é o substantivo en vez da contracción. A primeira oración aparece no titular e a segunda no corpo da noticia.

<expresión>As temperaturas no hemisferio norte de decembro a febreiro foron case un grao superiores á media do século XX</expresión> [GH2007-03-17/45; Cultura-Sociedade]

<expresión>O organismo do Goberno de Estados Unidos indicou na súa páxina web que no período que vai de decembro a febreiro, as temperaturas foron de 0,72 graos centígrados superiores á media do século XX.</expresión> [GH2007-03-17/45; Cultura-Sociedade]

Na próxima oración dáse o caso contrario; o etiquetador deu como válida a alternativa da contracción en lugar do substantivo:

<expresión>¿Fíxoselle un gran favor ó xénero feminino, ou cor-táronselles as ás para chegar ós postos máis altos como está de-mostrando a diario coa súa puxanza, tesón, traballo e vonta-de?</expresión> [GH2007-03-17/81; *Opinión*]

A confusión entre o pronome átono proclítico de primeira plural e a contracción da preposición e o artigo plural, *token* “nos”, só se produce nunha ocasión:

<expresión>Por iso, se non se consegue maioría absoluta nas eleccións e se aspira a ser un partido no goberno “non nos queda máis remedio” que pactar, aínda que “nuns poderemos e nou-tros non”.</expresión> [GH2007-03-17/23; *Galicia*]

Esta análise en concreto, chamounos especialmente a atención e decidimos indagar un pouco máis. A nosa sorpresa foi considerable ao ver que a análise era impecable se retiraba-mos as comiñas. Confirmamos algo que xa intuíamos con outros erros: a presenza ou ausencia de comiñas altera o recoñece-mento. Este signo ortográfico altera a gramática “natural” da oración e iso dificulta as tarefas de recoñecemento e caracteriza-ción.

Finalmente, debemos constatar neste apartado outra defi-ciencia na etiquetación automática realizada. Trátase da forma “daquela” para a que dá como válida a análise de contracción e non a de adverbio que lle corresponde:

<expresión>O caso é que, ironicamente, ese pai ausente preocu-pábase daquela polo incerto futuro que nos esperaba e agora pensa que as mulleres temos xa gañada a batalla.</expresión> [GH2007-03-17/80; *Opinión*]

Tomamos nota dos erros e en próximas versións veremos se este comportamento se corrixe ao aumentar o corpus de adestramento ou se é necesario estudar outras alternativas para a súa solución.

2.4.4. Con formas verbais con pronomes enclíticos

Os seis casos que imos tratar aquí son os únicos, de todo o xornal, nos que se produce unha segmentación errónea e, en consecuencia, tamén etiquetación fallida onde está implicada unha posible forma verbal cun pronome enclítico.

Na metade dos casos desglosa a unidade en compoñentes cando correspondería un só elemento. No primeiro deles, a inusitada aparición de formas verbais en primeira persoa singular nos textos xornalísticos quizais explique a escolla a favor da terceira persoa do singular do presente de indicativo do verbo ler co pronome de acusativo singular de terceira -[Vpi30s ler] / [Raa3ms o]- en vez da primeira persoa do presente de indicativo [Vpi10s ler]:

<expresión>Onde leo a nova non se ofrecen máis datos:</expresión> [GH2007-03-17/86; Opinión]

No segundo exemplo deste subapartado a confusión prodúcese entre dúas posibles análises: unha como infinitivo de "rever" seguido do enclítico de terceira acusativo feminino e outra como terceira persoa do singular do presente de indicativo de "revelar". Obviamente, escolleu a alternativa non válida.

<expresión>Un libro da EGAP revela que nas deputadas priman as divorciadas e as solteiras</expresión> [GH2007-03-17/52; Cultura-Sociedade]

No último caso, a brevidade da secuencia e o non haber nada antes -é un titular- confunde o etiquetador que se decanta por un imperativo seguido de clítico de acusativo -[V0p20s ler] / [Raa3fp o]- no sitio da análise correcta como substantivo -[Scfp lea]-:

<expresión>Leas interminables</expresión> [GH2007-03-17/17; Galicia]

Só nun caso o etiquetador opta pola análise simple da unidade en lugar de considerar dous elementos. Así, na seguinte expresión a unidade "fana" é considerada forma verbal de "fa-

nar" -[Vpi30s fanar]- en vez de terceira do plural do presente de indicativo de "facer" e mais o pronome de terceira acusativo feminino singular -[Vpi30p facer] / [Raa3fs o]-.

<expresión>O secretario xeral da Asociación Unificada da Garda Civil (AUGC), Joan Miquel Perpinyá, mostrou onte a súa satisfacción polos dous proxectos de lei aprobados polo Consello de Ministros xa que, na súa opinión, "cambian absolutamente a Garda Civil, modernízana e fana máis eficaz".</expresión> [GH2007-03-17/67; España]

No seguinte exemplo as categorías implicadas son as mesmas, verbo e pronome, e na forma verbal só difire o atributo de número. É na caracterización do pronome onde se dan máis diferenzas. O etiquetador ten que escoller se está diante de "recorda" seguido de "nos" ou de "recordan" e mais "os". A análise do propio que precede á unidade en cuestión non lle proporciona ao etiquetador ningunha información sobre o número do substantivo, fundamental neste caso, e optou pola segmentación equivocada.

<expresión>Caxigueiro recórdanos a escasa protección das nosas costas a través de A linguaxe da memoria, un gravado cos nomes dos máis perigosos afundimentos do litoral galego: Prestige, Urquiola, Mar Exeo.</expresión> [GH2007-03-17/12; Galicia]

No último caso desta serie prodúcese unha cadea de erros: o etiquetador decide que "dixome" é un substantivo porque antes del cataloga "médico" como adxectivo e, en consecuencia, a palabra seguinte considéraa relativo en vez de conxunción:

<expresión>"O médico dixome que debín de perder o DIU, que como é unha cousa pequena puiden perdelo polo servizo sen decatarme", recordou.</expresión> [GH2007-03-17/41; Cultura-Sociedade]

Cómpre facer aquí unha puntualización. A estrutura sintáctica recoñecida polo programa non é inusual no corpus de adestramento:

Artigo + Adxectivo + Substantivo + Relativo: 64 casos

Artigo + Substantivo + Verbo + Conxunción: 183 casos

con todo, a análise non é a axeitada polo que será preciso estudar que factores determinan este comportamento e ver como poderemos corrixilo.

2.4.5. Con locucións

No lexicón de XIADA constan 583 locucións, das cales só 252 están caracterizadas como seguras, é dicir, a súa aparición non presenta ambigüidades segmentais (ex. *non obstante, a carranchapernas, a oito, etc.*). Nas 331 restantes existe ambigüidade e o etiquetador ten que decidir se está ante unha locución ou non (ex. *en breve, con todo, etc.*)

No subcorpus avaliado figuran 235 ocorrencias de locucións. De todas elas, o etiquetador equivocouse en 51 casos, o que supón o 21,7% de erro. Vexamos a súa distribución:

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV
Nº de palabras	5859	1015	6489	2523	821	761	3795	949
Nº de locucións totais	52	11	74	23	11	5	47	12
Nº de locucións falladas	8	1	13	4	0	2	21	2
Porcentaxe de erro	15,3%	9,09%	17,56%	17,39%	0%	40%	44,68%	16,66%

Unha vez máis comprobamos como na sección de Opinión conseguimos os peores resultados. De todos xeitos, as cifras obtidas estannos indicando que debemos actuar de inmediato sobre este campo, creando regras precisas que desfagan o maior número posible das ambigüidades segmentais existentes e facilitarlle así o labor ao etiquetador. É dicir, incidir, a través de re-

gras, naquelas locucións que foron caracterizadas como inseguras.

No exemplo seguinte o etiquetador desagregou a locución nos seus constituíntes e non a tratou como unha unidade multipalabra:

<expresión>Barreras seguirá á espreita por se Navantia se abre ó civil en breve</expresión> [GH2007-03-17/13; Galicia]

Este erro podería corrixirse cunha regra na que se indique que só se a continuación aparece un substantivo non a considere locución adverbial: en breve + S -> non La0

No caso seguinte o erro producido é xustamente o contrario: considera locución o que debería ser desglosado en distintas unidades:

<expresión>Din os analistas que o camiño do bilingüismo efectivo faise máis curto día a día.</expresión> [GH2007-03-17/84; Opinión]

Como locución prepositiva, "camiño de" equivale a "cara a, en dirección a" e para que teña este valor non pode ir precedido de determinante [D], adxectivo [A] ou calquera outra forma que funcione como determinante. É dicir, o substantivo "camiño" non pode aparecer precedido de artigo, adxectivo ou ningunha outra forma cuxa etiqueta teña na segunda posición un "d" correspondente ao valor "determinante" do atributo valor. A regra complicase cos numerais onde debemos lembrar que ese atributo ocupa a terceira posición: [D, A, Ed, Td, Md, Id, N?d, Gd] + camiño de -> non Lp0

2.5. Erros na asignación dos atributos

No referido aos atributos é interesante destacar que a confusión prodúcese maioritariamente só sobre un deles, sendo os valores do xénero (masculino, feminino, ambiguo, non aplica) os que máis problemas lle causan ao etiquetador mentres que no caso de errar en dous atributos, estes son xénero e número (singular, plural, ambiguo, non aplica).

Na táboa seguinte aclaramos a nosa afirmación coas cifras resultantes da extracción e clasificación dos fallos referentes aos atributos:

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV
Nº de palabras	5859	1015	6489	2523	821	761	3795	949
Nº de erros totais	295	51	375	134	40	39	257	75
Nº de erros nos atributos	107	20	113	40	26	11	63	22
Nº de erros en 1 atributo	89	17	97	33	21	7	51	17
Nº de erros en 2 ou máis atributos	18	3	16	7	5	4	12	5
Porcentaxe de erro sobre o total de erros	36,27%	39,21%	30,13%	29,85%	65%	28,2%	24,51%	29,33%
Porcentaxe de erro nos atributos	1,82%	1,97%	1,74%	1,58%	3,16%	1,44%	1,66%	2,31%

A táboa que segue recolle os atributos en conflito, distribuídos por sección:

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV
valor (nuclear modificador)	6	-----	4	8	4	-----	10	3
valor (determi- nante non determinante)	7	5	4	1	-----	-----	9	4
xénero	38	9	60	13	9	5	22	7
número	9	2	10	4	5	1	4	2
número do posuidor	12	1	12	3	3	1	1	---
propio común	10	-----	3	2	-----	-----	-----	---
caso [pronome]	4	-----	1	-----	-----	-----	-----	---
tipo [locución ou pronome]	2	-----	1	-----	-----	-----	1	---
persoa [3ª por 1ª]	-----	-----	2	-----	-----	-----	2	---
tempo	-----	-----		1	-----	-----	2	---
modo	1	-----	-----	1	-----	-----	-----	1
xénero e número	15	3	14	7	5	4	6	3
tempo e modo	2	-----	2	-----	-----	-----	2	1
caso e xénero	1	-----	-----	-----	-----	-----	3	---
tempo, modo e persoa	-----	-----	-----	-----	-----	-----	1	1

Deseguido recolleemos todos os *tokens* que se corresponden con fallos nos atributos correspondentes a *número do posuidor*, *caso*, *tipo*, *persoa*, *tempo*, *modo*, *tempo e modo*, *caso e xénero* e, finalmente *tempo*, *modo e persoa*. Para os demais atributos -*valor*,

xénero, número, tipo no substantivo e *xénero e número*- eliximos unha mostra representativa por non ser posible incluír todos os casos.

Como se pode deducir polos exemplos recollidos, os erros na asignación de atributos só se producen cando a forma implícada presenta na súa etiqueta algún valor ambiguo, para cuxa caracterización só o contexto –e non sempre– nos aclara que valor, dos posibles, lle corresponde.

Atributo implicado	Exemplos de formas gráficas nas que se comete o fallo
valor (nuclear modificador)	máis, non, antes, mesmo, menos, mellor, seriamente, etc.
valor (determinante non determinante)	aquel, outros, calquera, oito, dúas, primeira, 2.140, 5.110, etc.
xénero	axentes, importante, que, lle, portavoces, dezaioito, eu, demais, etc.
número	que, PXOM, 80, Galicia, 2002, José Saramago, prevén, etc.
nº do posuidor	seu, seus, súa, súas
propio común	Alcaldía, Barreras, Barreiro, Bugallo, Saramago, Falanxe, etc. ²¹
caso [pronome]	nos, me
tipo [locución ou pronome]	ela, nós, mentres tanto
persoa [3ª por 1ª]	tiña, era, tería
tempo	ampliamos, lía, cría
modo	este
xénero e número	que, ningún, PXOM, Navantia-Fene, Méndez Romeu, etc.
tempo e modo	privatizara, advertira, fora, pedira, nacera
caso e xénero	me, te
tempo, modo e persoa	mira, parece

²¹ Todas estas unidades aparecen a comezo de oración, o que implica a presenza de maiúscula inicial e a ambigüidade no tipo, polo que o etiquetador ten que decidir se o substantivo é propio ou común.

2.6. Erros na atribución de categoría

Se nos centramos nos problemas de recoñecemento automático da clase de palabra morfolóxica, comprobamos que as confusións producidas son, na maioría dos casos, lóxicas dada a proximidade de certas clases ou explicables pola limitación que presenta o etiquetador, baseado nun modelo de Markov [Markov, 1913: 153-162] de grao 2, ao ter en conta conceptualmente só a forma e etiqueta das dúas palabras anteriores e posteriores. A seguinte táboa recolle os erros producidos por asignar unha categoría incorrecta. Na primeira columna preséntanse as clases que entran en conflito²² e nas seguintes as seccións onde se producen eses erros e exemplos extraídos dos documentos:

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV	Exemplos
adverbio conxunción							4		nin, pois
adverbio indefinido	2		2			1	7	1	algo, máis, tal, mesmo, pouco, etc.
adverbio numeral			1						primeiro
adverbio pronomes	3								el
adxectivo adverbio	1						2		libre, medio
adxectivo verbo (participio)	14	2	6	5	2	1	1	1	postas, afectados, achegadas, etc.
adxectivo substantivo	21		31	7		2	27	6	efectivos, tácticas, técnicas, etc.
artigo indefinido	1					1	1		uns, un
artigo numeral	1	1	3	1			1	1	unha, un

²² Anotamos as dúas clases de palabras implicadas, a incorrecta -asignada polo etiquetador- e a correcta -asignada polo lingüista- pero a orde de aparición non significa que todos os casos se dean na dirección apuntada senón que integran fallos nos dous sentidos. Por exemplo, a confusión entre conxunción e relativo presenta algún caso en todas as seccións do xornal, non obstante nunhas ocasións caracterízase como conxunción o que é relativo e noutras como relativo cando é conxunción.

artigo preposición	8	2	6		1	1	4	a
artigo pronome	4		2	1			2	o, os, as, a
conxunción interrogativo	3		1			2	2	1 que, como
conxunción numeral			1					segundo
conxunción preposición			2	1				segundo, senón
conxunción relativo	22	3	30	18	1	3	19	2 que, como, cando
conxunción pronome	7	1	3	3			5	1 se
indefinido adxectivo								1 propia
indefinido numeral			1					un
numeral sigla	1							A-8
preposición adverbio			1				1	ata
preposición pronome	1		1					a
preposición substantivo			1					cabo
relativo interrogativo	1		2	1		3		que, como, quen, onde
sigla numeral			2				1	2 XX, Leb2
substantivo adverbio			1	2				ben, si, non
substantivo indefinido							1	propios
substantivo numeral							1	segundos
substantivo sigla			8					3 Cegadi, Igape, MLR, CSL, etc.
substantivo verbo	5		7	2			10	1 perda, poder, lixo, conta, etc.
verbo adxectivo			1				1	turbia, digna

2.7. Erros na atribución do lema

Como podemos observar na táboa xeral de erros, os 6 casos recollidos baixo esta epígrafe só constitúen o 0,02% de fallos, non obstante o noso obxectivo é acadar o 100% de efectividade.

Debemos distinguir neste apartado de erros na adxudicación do lema dous tipos ben diferenciados:

1) O etiquetador falla na atribución do lema pero acerta coa etiqueta. Estes erros foron contabilizados no apartado e) da

táboa xeral onde se recollen os erros producidos ao atribuír o lema.

2) O etiquetador equivócase ao asignar o lema porque fallou na etiquetación. Estes erros incluímos nesta sección para a súa explicación -non nas estatísticas, onde van ao apartado c), correspondente a erros nos atributos- porque son os únicos casos nos que o erro na etiquetación provoca un erro na asignación do lema.

Vexamos as diferencias entre estes dous tipos de equivocacións cos casos reais extraídos do corpus avaliado.

1) O etiquetador falla na atribución do lema pero acerta na etiqueta: a etiqueta é idéntica e a caracterización só difire no lema²³.

Un *token* pode corresponder a máis dun lema pertencente á mesma categoría gramatical, -factor que non teñen en conta os etiquetadores-, pero o normal é que o contexto nos aclare a cal corresponde.

Na noticia 12 da sección *Galicia* do subcorpus avaliado temos un caso no que o contexto non nolo aclara:

<expresión>Algunhas das creacións que conforman a mostra 'ArtNatura', exposta no claustro do Rosalía de Castro</expresión> [GH2007-03-17/12; Galicia]

É unha das ocorrencias de moitas formas verbais coincidentes dos derivados de *poñer* e *pór* nos que o problema de asignación de lema non é do etiquetador senón do lingüista. Por que lema decantarnos? A solución tomada foi considerar que o lema non marcado ía ser o de *poñer* e que só se no resto do documento aparecía algunha das formas distintivas dos derivados de *pór* optariamos por este.

Nos seguintes casos non atopamos ningunha explicación racional e comprobamos que o etiquetador ante unha forma

²³ O etiquetador para as súas escollas só ten en conta a forma e a etiqueta, nunca os lemas.

cunha etiqueta unívoca remisíble a máis dun lema escolle un aleatoriamente.

Isto ocorre coas formas do tema de pretérito *fomos* e *foi*, coincidentes para os verbos *ser* e *ir* nas que opta sempre por atribuílas ao lema *ser*.

<expresión>Pola súa parte, Miguel Anxo Fernández Lores quixo destacar que o BNG “é a única forza deste concello que ten un proxecto claro de cidade, que fomos executando nestes últimos oito anos de goberno, con resultados positivos á vista de todos”.</expresión> [GH2007-03-17/20; Galicia]

<expresión>Despois tivo lugar a mesa redonda no Ateneo da capital ourensá, presentada polo humorista Xose Lois González O Carrabouxo, quen asegurou que se trata de homenaxe “necesario e bo”, porque “aínda a calor da presenza de Peña na rúa do Paseo non se foi”.</expresión> [GH2007-03-17/26; Galicia]

Paradoxalmente, para *fun* remítea ao verbo *ir* e non a *ser*:

<expresión>fun coordinador desa “normalización lingüística”, durante anos.</expresión> [GH2007-03-17/81; Opinión]

Por último, dúas mostras na clase dos substantivos. Na primeira, a forma “cores”, coa etiqueta substantivo común feminino plural -Scfp- pode remitir ao lema “cor” ou a “core” e o etiquetador decide, sen ningún criterio, que é “core”. Na segunda, para a forma “raíña”, presente nun refrán, o etiquetador considerou que era o feminino de “rei” e non o sinónimo de “raiola”:

<expresión>Caen as árbores da Amazonía, desécanse as lagoas azuis e os encoros apertados entre montañas, morren plantas coñecidas, apáganse as cores da infancia.</expresión> [GH2007-03-17/85; Opinión]

<expresión>Marzo, marzán, cara de can: pola mañá sol da raíña e pola noite afeitan as barbas cunha fouciña.</expresión> [GH2007-03-17/79; Opinión]

2) O etiquetador equivócase ao asignar o lema porque fallou na etiquetación.

Na seguinte oración atribúe a forma *prevén* ao lema *previr* e non a *prever* como lle corresponde:

<expresión>Non obstante, a decisión non satisfai a todos e onte mesmo CCOO xa advertiu de que existe “preocupación e temor” entre os traballadores de Fadesa, polo seu futuro laboral, pois prevén que “en dous ou tres anos” os novos donos desmantelen os 450 empregos da Coruña.</expresión> [GH2007-03-17/36; Cultura-Sociedade]

A explicación do fallo radica en primeira instancia na distancia existente entre o verbo e o seu suxeito e en segunda na estatística: no corpus de adestramento figuran 11.752 formas verbais en terceira persoa do singular fronte a 3.998 de terceira plural.

Algo semellante aconteceu coas formas “*cría*” e “*lí*”, ambas terceira persoa do singular do copretérito de indicativo dos verbos “*crer*” e “*ler*”, para as que o etiquetador optou pola etiqueta que as caracterizaba como presente dos verbos “*criar*” e “*liar*”, respectivamente. Consultamos o corpus de adestramento e cremos que as 5.521 ocorrencias de Vpi30s fronte ás 467 de Vii30s influíron no resultado.

<expresión>Vicente Risco cría que os debuxos de vexetais que antes facían os nosos canteiros nas tampas dos defuntos, poden ter algunha relación coa resurrección, pois as árbores teñen varias vidas, tantas como anos...</expresión> [GH2007-03-17/79; Opinión]

<expresión>Aquela nena que lía os xornais da época coa sensación de estar a vivir unha páxina da historia sabía que o camiño estaría condicionado por nacer muller, e non poucas veces imaxinou as vantaxes de ser home.</expresión> [GH2007-03-17/80; Opinión]

É a mesma hipótese que consideramos para as tres ocorrencias de *estea*, presente de subxuntivo de *estar*. Para todas elas remite ao lema *estear* (=apoiar; escampar; secar) porque a eti-

queta correspondente ao presente de indicativo -Vpi30s- aparece en 5.521 ocasións mentres que a de presente de subxuntivo -Vps30s- só o fai en 544:

<expresión>Medio Rural considerou necesario que esta formación estea acompañada dunha divulgación das medidas postas en marcha por este departamento na loita contra os incendios forestais: o Decreto de xuño e o proxecto de Lei de Prevención e Defensa.</expresión> [GH2007-03-17/4; Galicia]

<expresión>Para a AUGC, segundo declarou a Efe Perpinyá, "é un día feliz" por "estas dúas normas, que farán que o persoal estea máis motivado".</expresión> [GH2007-03-17/67; España]

<expresión>Telecinco acaba de lograr que unha teleserie que estaba moi mal de saúde, M.I.R., estea a recuperarse logo dun cambio de día e de subministrar varias doses de paciencia.</expresión> [GH2007-03-17/89; TV]

3. Conclusións

Se descartamos os erros non atribuíbles ao etiquetador -por non proporcionar lema ou equivocarse ao asignalo, por erros lingüísticos textuais e por non figurar unha análise no lexicon- vemos como as porcentaxes de acerto do etiquetador soben, situándose a media nun 95,44%.

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV	Total
palabras	5859	1015	6489	2523	821	761	3795	949	22 212
erros totais	246	37	303	108	35	36	195	54	1014
segmentacións erróneas	43	7	79	27	5	11	44	12	228
erros nos atributos	107	20	113	40	26	11	63	22	402
erros na categoría	96	10	111	41	4	14	88	20	384
porcentaxe de erro	4,19%	3,64%	4,66%	4,28%	4,26%	4,73%	5,13%	5,69%	4,56%
porcentaxe de acerto	95,81%	96,36%	95,34%	95,72%	95,74%	95,27%	94,87%	94,31%	95,44%

Esta cifra permítenos ser moi optimistas sobre a porcentaxe de acerto que acadará o noso etiquetador no futuro posto que é moi alentador que, cunha morfoloxía bastante complexa e un etiquetario de 375 etiquetas distintas reais, consigamos xa un 95% de acerto para o galego actual, sabendo, ademais, que dispoñemos dunha marxe de mellora considerable nos numerais e propios multipalabra e no recoñecemento das locucións.

Finalmente, queremos comparar os resultados do noso etiquetador cos doutros etiquetadores. Para iso debemos descartar todos os fallos relacionados coa lematización e a segmentación. Os primeiros, porque nos resultados doutros etiquetadores non se recollen os erros de lematización; os segundos, porque eses etiquetadores non realizan ningún tipo de segmentación das distintas unidades léxicas presentes nun texto.

Por iso, para empregar os mesmos criterios de validación na comparativa, só podemos ter en conta os erros producidos en calquera punto da etiqueta, incluídos os que se cometen por non figurar no lexicón con ese valor.

Como se pode observar na última táboa, onde reflectimos só os erros relacionados coa etiquetación, a porcentaxe final de acerto que acadamos con esta versión do etiquetador é semellante ás dispoñibles para o inglés ou o castelán: 95,99%.

	Galicia	Campus	Cultura	España	Internacional	Economía	Opinión	TV	Total
palabras	5859	1015	6489	2523	821	761	3795	949	22 212
erros totais	228	30	247	95	32	27	181	52	892
erros nos atributos	107	20	113	40	26	11	63	22	402
erros na categoría	96	10	111	41	4	14	88	20	384
por non estar con ese valor	25	---	23	14	2	2	30	10	106
porcentaxe de acerto	96,11%	97,05%	96,2%	96,24%	96,11%	96,46%	95,24%	94,53%	95,99%

Referencias bibliográficas

- Graña Gil, J. (2000): *Técnicas de Análisis Sintáctico Robusto para la Etiquetación del Lenguaje Natural*, Tese doutoral, Departamento de Computación, Universidade da Coruña, 2000.
- Jurafsky, D. e J.H. Martin (2000): *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, 2000.
- Leech, G. e A. Wilson (1994): "Morphosyntactic Annotation", Draft-Work In Progress, EAGLES Document EAG-CSG/IR.T3.1. en N. Calzolari e J. M. McNaught (eds.), *EAGLES Interim Report EAG-EB-IR-2*, 1994.
- Sampson, G. (1994): *The SUSANNE corpus, release 3, 04/04/1994*. School of Cognitive & Computing Sciences, University of Sussex, Falmer, Brighton (England).
- Markov, A. A., (1913): "An example of statistical investigation in the text of Eugene Onyegin illustrating coupling of tests in chains" en *Proceedings of the Academy of Sciences, St. Petersburg*, vol. 7 of VI, 1913, pp. 153-162.

Bibliografía complementaria

- Alarcos Llorach, E. (1994): *Gramática de la lengua española*, Madrid, Espasa Calpe, 1994.
- Álvarez, R., H. Monteagudo e X.L. Regueira (1986): *Gramática Galega*, Vigo, Galaxia, 1986, 4ª ed.
- Álvarez, R. e X. Xove (2002): *Gramática da lingua galega*, Vigo, Galaxia, 2002.
- Álvarez, R. (1994): "Secuencias de pronomes átonos en galego moderno" en Ramón Lorenzo (ed.), *Actas do XIX Congreso de lingüística e filoloxía románicas*, pp. 247-265, vol. VI, A Coruña, Fundación Pedro Barrié de la Maza Conde de Fenosa, 1994.
- Bosque, I. (1991): *Las categorías gramaticales. Relaciones y diferencias*, Madrid, Síntesis, 1991.

- Bray, T.; J. Paoli e C.M. Sperberg-McQueen (eds.): *Extensible Markup Language (XML)* dispoñible en <http://www.w3.org/TR/REC-xml> [Consulta: 16/11/2006]
- Burnage, G. e D. Dunlop (1993): "Encoding the British National Corpus", en J. Aarts, P. de Haan e N. Oostdijk (eds.), *English Language Corpora: Design, Analysis and Exploitation*, Amsterdam, Rodopi, 1993.
- Burnard, L. (1997): "The Text Encoding Initiative's. Recommendations for the Encoding of Language Corpora: Theory and Practice" [Seminario de Industrias de la Lengua, Soria, 14-18 de xullo de 1997]
- Chacón Calvar, R. e M. Rodríguez Alonso (1993): *Diccionario crítico de dúbidas e erros da lingua galega*, Sada, Ediciós do Castro, 1993.
- Costa Casas, X.X., M. Anxos González Refoxo, C.C. Morán Fraga e X.C. Rábade Castiñeira (1998): *Nova Gramática para a aprendizaxe da lingua*, A Coruña, Vía Láctea, 1988.
- Díaz Regueiro, M. (1992): *Os verbos galegos*, Consellería de Educación e Ordenación Universitaria. Dirección Xeral de Política Lingüística, Santiago de Compostela, 1992.
- Freixeiro Mato, X.R. (2000): *Gramática da lingua galega. Morfosintaxe*, A Coruña, A Nosa Terra, 2000.
- Ide, N. (coord): "Corpus Encoding Standard. Versión 1.4." dispoñible en <http://www.cs.vassar.edu/CES> [Consulta: 16/11/2006]
- Instituto da Lingua Galega e Real Academia Galega (2003): *Normas ortográficas e morfolóxicas do idioma galego*, Santiago de Compostela, ILG/RAG, 1993, 11ª ed.; 1997, 16ª ed.; 2003, 18ª ed.
- Instituto da Lingua Galega e Real Academia Galega (1989): *Vocabulario Ortográfico da Lingua Galega. Versión provisional*, [Edición fotocopiada], Santiago de Compostela, 1989.
- González González, M. e A. Santamarina Fernández (coords.) (2004): *Vocabulario Ortográfico da Lingua Galega*, Santiago de Compostela / A Coruña, ILG/RAG, 2004.
- Morel, J; S. Torner; J. Vivaldi; Ll. de Yzaguirre e M.T. Cabré (1998): *El Corpus de l'IULA: Etiquetaris*, Barcelona, Univer-

- sitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 1997; 2ª ed. revisada e corrixida, 1998.
- Nicolás Rodríguez, R. (1993): *Diccionario dos verbos galegos*, Pontevedra, Edicións do Cumio, 1993.
- Reis, R. e J.J. Dias de Almeida (1998): "Etiquetador morfo-sintáctico para o Português" en M.R. Braga Marquillas e M.A. Mota (eds.), *Actas do XIII Encontro Nacional da Associação Portuguesa de Linguística*, pp. 209-221, vol. II, APL-Colibri, Lisboa, 1998.
- Sánchez León, F. (1997): "El etiquetado del Corpus de Referencia del Español Actual (CREA)" [Seminario de Industrias de la Lengua, Soria, 14-18 de xullo de 1997]

TÁBOAS

			m-masculino			1-primeira
			f-feminino	s-singular		2-segunda
		c-coordinante	a-masculino ou feminino	p-plural		3-terceira
		s-subordinante	n-neutro	a-singular ou plural	c-comparativo s-superlativo	a-primeira ou terceira
		0-non aplica	0-non aplica	0-non aplica	0-non aplica	0-non aplica
Categoría	Tipo	Subtipo	Xénero	Número	Grao	Persoa
Substantivo (S)	c-común p-proprio		m / f / a / 0	s / p / a / 0		
1	2		3	4		
Adxectivo (A)			m / f / a	s / p / a	c / s / 0	
1			3	4	2	
Verbo (V)			m / f / 0	s / p / 0		1 / 2 / 3 / a / 0
1			5	6		4
Preposición (P)						
1						
Conxunción (C)	c-coordinante s-subordinante					
1	2					
Adverbio (W)	n-nuclear m-modificador a-nuclear ou modificador r-relativo g-interrogativo-exclamativo					
1	2					
Artigo (D)	d-determinado i-indeterminado		m / f	s / p		
1	2		3	4		
Demonstrativo (E)			m / f / n	s / p		
1			3	4		
Relativo (T)			m / f / a	s / p / a		
1			3	4		
Posesivo (M)			m / f / a	s / p / a		1 / 2 / 3
1			5	6		3
Indefinido (I)			m / f / a / 0	s / p / a / 0		
1			3	4		
Numeral (N)	c-cardinal o-ordinal		m / f / a	s / p		
1	2		4	5		
Pronome (R)	t-tónico a-átono		m / f / n / a	s / p / a		1 / 2 / 3
1	2		5	6		4
Exclamativo-Interrogativo (G)			m / f / a	s / p / a		
1			3	4		
Locución (L)	a-adverbial p-prepositiva c-conxuntiva	c / s / 0				
1	2	3				
Intersección (Y)						
1						
Signo de puntuación (Q)						
1						
Categoría Periférica (Z)	f-fórmula a-abreviatura g-sigla s-símbolo o-outros tipos		m / f / a / 0	s / p / a / 0		
1	2		3	4		

Táboa 1: Etiquetario de XIADA

	Galicia	Campus	Cultura	España
palabras	5859	1015	6489	2523
erros	295	51	375	134
erros ao non dar lema	18 casos 0,30%	12 casos 1,18%	47 casos 0,42%	10 casos 0,39%
segmentacións erróneas	43 casos 0,73%	7 casos 0,68%	79 casos 1,21%	27 casos 1,07%
erros nos atributos	107 casos 1,82%	20 casos 1,97%	113 casos 1,74%	40 casos 1,58%
erros na categoría	96 casos 1,63%	10 casos 0,98%	111 casos 1,71%	41 casos 1,62%
erros ao atribuír lema	3 casos 0,05%	-----	-----	-----
por non figurar no lexicón con ese valor	25 casos 0,42%	-----	23 casos 0,35%	14 casos 0,55%
por erros lingüísticos ou desviacións da norma	2 casos 0,03%	2 casos 0,19%	2 casos 0,03%	1 caso 0,03%
porcentaxe de erro	5,03%	5,02%	5,77%	5,31%
porcentaxe de acerto	94,97%	94,98%	94,23%	94,69%

Táboa 2: Distribución de erros clasificados por tipo e sección

Internacional	Economía	Opinión	TV	Total
821	761	3795	949	22 212
40	39	257	75	1260
3 casos 0,36%	-----	24 casos 0,63%	8 casos 0,84%	122 casos 0,54%
5 casos 0,60%	11 casos 1,44%	44 casos 1,15%	12 casos 1,26%	228 casos 1,02%
26 casos 3,16%	11 casos 1,44%	63 casos 1,66%	22 casos 2,31%	402 casos 1,8%
4 casos 0,48%	14 casos 1,83%	88 casos 2,31%	20 casos 2,10%	384 casos 1,72%
-----	-----	3 casos 0,07%	-----	6 casos 0,02%
2 casos 0,24%	2 casos 0,26%	30 casos 0,79%	10 casos 1,05%	106 casos 0,47%
-----	-----	3 casos 0,07%	2 casos 0,21%	12 casos 0,05%
4,87%	5,12%	6,77%	7,9%	5,67%
95,13%	94,88%	93,23%	92,1%	94,33%

TABLE I			
Year	Population	Area	Population Density
1900	1,000,000	10,000	100
1910	1,500,000	12,000	125
1920	2,000,000	15,000	133
1930	2,500,000	18,000	139
1940	3,000,000	20,000	150
1950	3,500,000	22,000	159
1960	4,000,000	24,000	167
1970	4,500,000	26,000	173
1980	5,000,000	28,000	179
1990	5,500,000	30,000	183
2000	6,000,000	32,000	188
2010	6,500,000	34,000	191
2020	7,000,000	36,000	194

TABLE I. Population and Area of the United States, 1900-2020.

The population of the United States has increased steadily over the past century, from 1,000,000 in 1900 to 7,000,000 in 2020. This growth has been accompanied by a corresponding increase in the area of the country, from 10,000 square miles in 1900 to 36,000 square miles in 2020. The population density of the United States has also increased over time, from 100 people per square mile in 1900 to 194 people per square mile in 2020. This increase in population density is a result of the concentration of the population in the eastern half of the country, where the population density is much higher than in the western half. The increase in population density is also a result of the increase in the number of people living in urban areas, which have a much higher population density than rural areas. The increase in population density is a reflection of the growth of the United States as a nation, and it is a sign of the progress that has been made in the country.

Joan Veny: *Petit atlas lingüístic del domini català*, volum 1. Barcelona: Institut d'Estudis Catalans, 2007.

Camiñan as linguas cara á eliminación das diferenzas diatópicas? Sabemos desde hai tempo que o espazo, con todo o que este implica, é un dos factores de diferenciación lingüística. Probablemente a maioría das linguas non conseguirán nunca evitar as diferenzas que veñen dadas pola diversidade do territorio en que son faladas.. Pero a pesar de ser isto así, resulta evidente que na sociedade actual existen factores que están actuando dunha maneira moi poderosa na dirección de eliminar moitas das características dialectais típicas de cada territorio. Hoxe a escolarización é un fenómeno universal nos países desenvolvidos, e a través do sistema escolar transmítese un modelo lingüístico que tende á uniformidade e á expansión dunha variante estándar das linguas. No mesmo sentido actúan os medios de comunicación, a través dos que se difunde unha variedade lingüística común, que non ten en conta as características dialectais dos receptores. É probable que no futuro adquiran maior peso as variedades diatráticas ou sociais que as de tipo diatópico.

Pero, quizais porque a poboación empeza a ser consciente deste fenómeno, obsérvase un aprecio cada vez maior polas características lingüísticas propias do microcosmos ao que cada persoa se atopa máis vinculado, especialmente por aquelas que están máis afastadas da variedade normativa.

Esta realidade realza aínda máis o valor de traballos coma o Pequeno Atlas Lingüístico do Dominio Catalán, obra que sen perder nin un ápice da seriedade científica de que fai gala o *Atlas Lingüístic del domini català*, do que se publicaron ata este momento tres volumes, pretende sacar a preocupación e a interpretación dos fenómenos dialectais do ámbito dos lingüistas especialistas, para achegalos dunha maneira amigable ao conxunto da poboación.

A principal característica metodolóxica que diferencia este pequeno atlas do seu irmán maior é o estudo que acompaña a cada mapa. O profesor Joan Veny presenta nesta obra cada fenómeno comentado e explicado dunha maneira clara, atractiva e intelixible para calquera lector culto.

Que contén este primeiro volume do *Petit Atlas Lingüístic del domini català*?

O volume ábrese cunha introdución do propio J. Veny na que fai unha presentación do proxecto, encádrao dentro da xeografía lingüística europea, poñéndoo en paralelo con outras obras similares, como o *Petit Atlas Linguistique de la Wallonie* ou o *Kleiner Bayerischer Sprachatlas*, e explica tanto a macroestrutura como a microestrutura da obra. Nesta parte de presentación da obra, a introdución vai seguida dun apartado dedicado ás convencións gráficas (no que se describen os signos fonéticos e as abreviaturas utilizadas ao longo da obra) e doutro no que se ofrece a relación dos puntos de enquisa.

Unha serie de mapas introdutorios ofrécenos a localización de cada un dos puntos de enquisa, a división comarcal, unha clasificación dialectal, a división eclesiástica tradicional e actual, e a división política e administrativa. De especial interese é o mapa da clasificación dialectal -realizado, como todos os mapas deste atlas, cunha claridade e visibilidade extraordinarias-, que vai acompañado dunha descrición moi ordenada e moi completa dos trazos máis característicos de cada variedade dialectal. Clasificación revisada tendo en conta a vitalidade actual dos trazos lingüísticos atribuídos tradicionalmente a cada unha destas variedades, o que explica que se prescindise do *salat* e do *xipella*, precisamente pola recesión das características propias destas variedades, e que, en cambio, se introduza, dentro do valenciano meridional, un subdialecto *alicantí* con características propias.

A parte nuclear da obra está dedicada aos mapas lingüísticos, o primeiro deles de carácter comparativo, destinado, valéndose das realizacións da palabra *mudo*, a "situar" o catalán fronte ás linguas que o rodean: o castelán, o aragonés e o occitano.

Trinta e catro mapas céntranse no estudo da fonética (10 dedicados ao vocalismo e 24 ao consonantismo, complementados por outros dous de carácter fonosintáctico). A través destes mapas podemos examinar dunha maneira moi plástica fenómenos coma o da distribución da vogal final *-e* nos masculinos, a variación do timbre da vogal átona final *-a* nos femininos, o timbre da vogal do morfema dos femininos plurais, a diferenciación ou non de *o/u* átonos, a conservación ou non do *v* labiodental, o enordecemento do [z] intervocálico, a iodización ou ieísmo, a realización ou non do *r* final nos substantivos e nos verbos, a epéntese consonántica ou non no grupo *-n'r-* etc.

Á morfoloxía unicamente se lle dedica neste volume un mapa, o da distribución do pronome *es/se* na frase portadora "el nen *es xucla* el dit".

A maioría dos mapas refírense ao léxico. Neste volume o léxico que se estuda pertence todo ao campo semántico do corpo humano. Son 66 mapas e 66 comentarios que fan as delicias de calquera lector, mesmo do non lingüista, e que nos introducen dunha maneira amena no coñecemento da distribución das áreas léxicas, no coñecemento dos étimos de cada denominación e nas motivacións que subxacen debaixo da maior parte das denominacións, da penetración de castellanismos en determinadas áreas nas que compiten coas formas autóctonas etc. Todo isto pode observarse en calquera mapa, valla como testemuño do que digo o comentario que acompaña o número 60, que estuda os nomes do nocello en catalán, e que me permito reproducir pola súa concisión:

"Al nord del domini s'estén la isoglossa de *clavilla*, amb els seus derivats *clavillar* i *claviller*, procedent del llatí *CLAVICULA* 'claueta', d'evident origen metafòric. Al sud, ocupa una àrea important *turmell*, der. de *torm*, probablement preromà ('bony'), amb la variant *turnell*, pel seu cràcter giratori. A l'oest, contigu a l'aragonès, i en relació amb aquesta mateixa idea de 'girar', es troba *tortol*, diminutiu de *torter* (<lat. *tortorius*), que allà també ha pres el sentit de 'cassoleta del genoll, ròtula' i 'espiral de fum'. En el val. septentrional tenim *garró*, diminutiu de *garra*, que de 'part inferior de la cama' (<cèltic *garra, i no del cast. *garra* 'urpa' < àrab) ha passat a designar el *turmell* per contigüitat. Finalment, una gran part del valencià ha introduït el castellanisme *tobillo* (a voltes amb la variant *toவில்*), flanquejat al nord i sud per supervivents del genuí *turmell*."

O volume complétase cunha bibliografía selecta, na que se recollen os traballos máis importantes relacionados cos temas e fenómenos que aparecen no volume, un pequeno vocabulario de terminoloxía lingüística de grande utilidade para o lector non especialista, e un índice alfabético de palabras que aparecen nos comentarios dos mapas, tanto palabras catalás coma étimos ou préstamos doutras linguas.

Os romanistas felicitámonos por dispoñer dunha obra coma esta que conxuga dunha maneira maxistral o rigor científico coa sinxeleza, a exactitude coa presentación agradable dos fenómenos lingüísticos, o que demostra que é posible outra maneira de presentar a investigación lingüística, de encontrarlle unha cara amable que a faga accesible para a maior parte da poboación. Claro que isto, que pode parecer sinxelo, só está ao alcance de verdadeiros sabios como é o caso de Joan Veny, que ten na cabeza perfectamente organizada toda a complexa realidade dialectal e histórica da lingua catalá e a súa incardinación no conxunto románico.

Cando temos a fortuna de atopar unha obra así, só nos cabe aproveitarnos dela, das súas ensinanzas tanto nos contidos coma na metodoloxía, e felicitar tanto o seu autor coma os colaboradores e a Institución que a promoveu e lle deu soporte. Parabéns, e desexamos que non se fagan esperar moito os seguintes volumes.

MANUEL GONZÁLEZ GONZÁLEZ
Universidade de Santiago

Mariño Paz, R. (2008). *Historia de la lengua gallega*. München: Lincom. 291 páx.

Sen dúbida un dos libros chave dentro dos estudos da lingüística galega dos últimos anos foi a obra de Ramón Mariño titulada *Historia da lingua galega* (Santiago de Compostela: Sotelo Blanco. 1998). Ata aquel momento non se dispoñía de ningún manual que recompilase de xeito exhaustivo información suficiente sobre as vicisitudes históricas sufridas pola lingua galega desde a súa prehistoria ata o momento actual. Desde entón as persoas interesadas no coñecemento desta lingua puidemos atopar nesta obra importantes datos sobre a súa historia externa de xeito ordenado e crítico, pero non só iso, tamén tivemos moita información sobre gramática histórica do galego, sobre historia da escrita en galego, sobre historia social da lingua e mesmo sobre literatura galega.

Dez anos despois o profesor Mariño volve retomar este mesmo tema, ofrecéndonos a obra que estamos a analizar. Non se trata dunha simple actualización da publicación anterior, senón que máis ben estamos ante unha reescrita sobre o mesmo tema, pero con pretensións diferentes. A diferenza básica é o posible público ó que vai dirixida. Creo que, mentres que a obra do 98 pretendía dotar a sociedade galega (sobre todo a curiosos, estudantes e especialistas sobre a lingua galega) dun manual básico sobre a historia da lingua, agora, dez anos despois, a pretensión é moito máis ambiciosa: trátase de darlle a coñecer as vicisitudes históricas da nosa lingua tamén á comunidade filolóxica europea e mesmo universal. E isto queda patente con só observar que, nesta ocasión, é unha prestixiosa editorial alemá con ampla presenza en toda Europa (pois publica obras en inglés, alemán, francés, español, portugués e italiano) a que se interesou por editar este estudo. Este é o motivo de que nesta ocasión a redacción do traballo estea feita en castelán.

É sen dúbida a calidade deste traballo de Ramón Mariño a que fixo que esta editorial se interesase pola súa publicación e iso debe ser un orgullo para o autor, pero tamén debe selo para todos os galegos o feito de que esta casa editora estea convencida do interese que a nosa

lingua esperta no mundo filolóxico europeo. Creo que é importante deixar ben patente este dato nun momento no que dentro da sociedade galega se está a facer moi visible un pequeno grupo social que, desde a máis vergoñenta ignorancia en aspectos históricos, culturais e lingüísticos, e desde unha total insensibilidade social, teñen o atrevemento de enarborar como bandeira a mentira, a aldraxe e o desprestixio da lingua que uniu durante séculos os habitantes do noroeste hispánico e que nos une hoxe a todos os galegos. Creo que sería bo para estes grupos sociais non só a lectura desta obra, na que sen dúbida aprenderían datos básicos sobre o noso pasado que eles ignoran totalmente, senón tamén levar a cabo unha reflexión sobre as razóns que poden facer que o mundo científico internacional se sinta interesado polo coñecemento dunha lingua que eles denigran e desprezan, a pesar de formar parte da súa propia historia e do seu presente como individuos pertencentes a un pobo de moi rica cultura.

Dicía máis arriba que non estamos ante unha simple actualización e tradución da obra do 98, senón que o enfoque é distinto. É máis, a obra do 98 continúa tendo total vixencia no presente, xa que os datos que se ofrecían seguen sendo totalmente actuais. Basicamente podemos incluír os cambios e diferenzas en tres grandes grupos:

a) Xa a primeira vista se observa que hai cambios na macroestrutura, isto é, na organización de capítulos e seccións. Nalgúns casos estas mudanzas veñen condicionadas por novos coñecementos e achados científicos recentes, e noutras ocasións é o distinto enfoque do libro o que parece aconsellar esta renovada estrutura. Nalgunhas ocasións mesmo se suprimen certos aspectos estudados na obra do 1998, probablemente por seren considerados demasiado de detalle para as pretensións da obra actual. Nos capítulos terceiro e cuarto, sen deixar de tratarse basicamente os mesmos aspectos, é onde os cambios da estrutura expositiva son máis importantes.

b) Por outro lado, nesta recente obra aparecen datos novos, froito dos coñecementos e achegas que foron xurdindo a partir do 1998. Igual que sucedía no caso anterior, o profesor Mariño dá mostras de coñecer en profundidade todos os estudos publicados sobre os temas que trata e de manter constantemente actualizados os seus coñecementos por medio de novas conclusións recollidas de publicacións recentes. Non en van, das máis de 550 referencias bibliográficas ofrecidas, case un terzo teñen data de publicación posterior a 1997.

c) Finalmente, tamén hai cambios de organización na microestrutura e na redacción dos subcapítulos e das seccións. Aínda

naqueles casos en que aparece unha mesma sección en ambas as obras, a organización interna acostuma variarse e a redacción é diferente, indo moito máis aló dunha simple tradución.

Xa entrando na estrutura interna da obra, esta aparece organizada en catro grandes capítulos, ademais dunha introdución previa e dunha bibliografía final. Merece especial mención a introdución, pois, a pesar da súa brevidade, constitúe unha magnífica reflexión sociolingüística sobre os procesos de dominación lingüística. Como exemplo vivo de especial importancia para coñecer certos procesos deste tipo é como o autor presenta a lingua galega, incitando deste modo a coñecer a historia do noso idioma. Trátase sen dúbida dunha atractiva e intelixente presentación que fará que aumente o interese por achegarse a este libro e por debuzarse sobre a lingua propia dos galegos.

Cada un dos catro capítulos da obra, organizados cronoloxicamente, trata un período da historia da lingua. O segundo (Idade Media), terceiro (séculos XVI, XVII e XVIII) e cuarto (séculos XIX, XX e XXI) teñen unha extensión semellante, mentres que o primeiro (formación da lingua) supón en extensión aproximadamente a metade de calquera dos anteditos. En cada un destes capítulos hai unha ou varias seccións introdutorias nas que se fai un repaso polos principais acontecementos históricos da época, especialmente por aqueles que de xeito máis directo influíron no devir da situación da lingua nesa altura. Tamén se detén o autor a miúdo a informarnos sobre a estrutura social da poboación galega nun determinado período e sobre os cambios que se van producindo na rede social.

Como xa indicamos, o primeiro dos grandes capítulos trata o período de formación da lingua galega: substratos, latinización e superestratos. Aínda que obviamente o autor se refire ó noso idioma, este primeiro capítulo vai moito máis aló destes límites, e moi ben podería servir (agás nalgúns aspectos concretos) como presentación da situación lingüística da Península Ibérica en todo este período (desde os tempos máis recuados dos que temos datos ata a época das invasións árabes) e ser, xa que logo, aproveitado para explicar a etapa de formación de calquera das linguas iberorromances actuais.

O segundo dos capítulos fai referencia á emerxencia e decadencia da lingua galega na Idade Media. Non trata, xa que logo, todo o período medieval, pois de feito no primeiro capítulo xa se chegara ata o período de invasións árabes (século VIII). Nun estrito sentido cronolóxico dáse unha certa superposición do capítulo primeiro (que abrangue os primeiros séculos da Idade Media) con este segundo.

Non obstante, o enfoque é ben diferente en ambos os casos. No primeiro faise referencia ós acontecementos que teñen que ver coa formación da lingua (latinización e superestratos), mentres que neste segundo se tratan outros aspectos, que aínda que poden ser simultáneos ós anteriores, teñen que ver co proceso de conversión do latín en romance.

Esta temática que acabo de sinalar (o proceso de evolución de latín a romance) é a primeira que se explica no segundo capítulo, pero a continuación tamén se abordan outras como o proceso de individualización do galego dos outros romances que o rodean, as diferenzas lingüísticas que van xurdindo entre galego e portugués nos últimos séculos da Idade Media, ou o cultivo da lingua galega na época medieval.

Obsérvase claramente neste segundo capítulo a intención do autor de ofrecernos moito máis ca unha historia externa. Aínda que sexa de forma indirecta, un lector curioso vai obtendo abundantes datos sobre os cambios que van outorgando personalidade propia á lingua medieval da *Gallaecia*, como froito da especial evolución do latín nesta zona, así como das diferenzas que se van observando entre o novo romance galego e os outros romances que o rodean. Cómpre deixar claro, de todas formas, que estes acertados datos sobre a evolución da lingua que nos ofrece Ramón Mariño non son exclusivos deste capítulo. Xa no primeiro aparecían apuntamentos deste tipo, e nos dous seguintes tamén figuran seccións dedicadas a cambios e características da lingua galega nos diferentes períodos cronolóxicos.

Ademais diso, cando o autor fai referencia ó cultivo da lingua galega neste período medieval, está a ofrecernos datos que máis ben pertencen ó plano da historia da literatura galega e da historia da escrita en galego, como tamén se volverá observar en gran medida nos capítulos sucesivos. Neste sentido resultan moi abundantes as seccións nas que o autor presenta estudos lingüísticos de textos da época: "El romance gallego en la documentación latina anterior al siglo XIII", "La lengua de los primeros textos escritos en romance gallego", "Los textos gallego-leoneses".

Será no capítulo terceiro onde se trate o período que o profesor Mariño chama de *dialectalización* da lingua galega (séculos XVI, XVII e XVIII), e faino desde diferentes perspectivas:

-Despois dunha introdución histórica estuda os inicios do proceso de penetración do castelán en Galicia como consecuencia da procedencia foránea de certos grupos sociais. É este o inicio da

decadencia da lingua galega ata ser considerada un dialecto do castelán.

-Desde un punto de vista de historia da cultura, estúdanse os procesos de menosprezo e de defensa da lingua galega.

-Desde unha perspectiva que podería considerarse máis literaria, tamén nos informa o autor da situación do cultivo escrito da lingua galega nos distintos xéneros literarios.

-Podería formar parte dunha historia da escrita en galego a sección que fai referencia ás variedades do galego escrito.

-Baixo a perspectiva da historia da lingüística galega, abórdase o estudo das obras do padre Sarmiento, do padre Sobreira e de José Cornide, entre outros autores menores.

-Unha vez máis aparecen datos de gramática histórica na sección dedicada ós cambios lingüísticos.

Ata hai ben pouco estabamos afeitos a que, cando se trataba o tema da historia da lingua ou da literatura galegas, se pasase moi por riba esta etapa que acostumaba chamarse "séculos escuros". Xa no 1998 Ramón Mariño acabou con esta tendencia, dedicándolle unha especial atención a estes tres séculos pouco coñecidos e con máis transcendencia do que se pensaba. Tamén nesta nova obra o autor se detén neste período e ofrece abundantes datos, moitos deles de primeira man, xa que é un grande especialista en historia da lingua galega e do seu cultivo non só na Idade Media, senón tamén nos séculos posteriores.

Finalmente, no capítulo cuarto expónsenos a situación da lingua galega nos séculos XIX, XX e comezos do XXI, tamén enfocada desde diferentes perspectivas. Aínda que a estrutura deste capítulo podería parecer moi diferente á do anterior e se introduza nova terminoloxía (*gramaticalización, ampliación funcional*), en realidade ofrécensenos os mesmos contidos: introdución histórica con especial referencia ós cambios políticos, administrativos, sociais, económicos e funcionais; lenta caída do monolingüismo en galego; cultivo escrito do galego e características que presenta este galego escrito; estudos lingüísticos sobre o galego; características novidasas do galego falado e estudo da conciencia que se ten sobre a lingua.

Finalmente, merece ser destacada a utilidade da sección bibliográfica, xa que a partir dela podemos afondar máis en calquera aspecto dos tratados neste libro. Destaca a súa exhaustividade, ademais dun bo filtro selectivo que sempre é de agradecer. Como xa indicamos máis arriba, estamos ante unha bibliografía totalmente actualizada na que se dá noticia dos últimos traballos existentes sobre os aspectos tratados nesta obra.

Como dixemos, son moitos os datos que nos ofrece Ramón Mariño a partir de achados propios. Porén, o groso dos contidos que se presentan nunha obra destas características teñen que basearse en abundantes traballos doutros lingüistas. Gustaríame destacar que, nestes casos, o autor procura informarnos das perspectivas de diferentes estudosos, expoñendo tamén a súa visión particular, acompañada dos argumentos que o levan a defendela. Isto demostranos a madureza da obra, así como as horas de estudo e de reflexión que hai detrás do tratamento dos numerosos aspectos presentes neste traballo. Deste xeito resulta doado que o lector extraia as súas propias conclusións ou que afonde nun determinado aspecto a partir dos datos bibliográficos aquí ofrecidos.

Xa para concluír, considero que estamos ante o que é e será unha obra chave na lingüística galega, tanto polo tema en si como polo seu rigor científico. Se tivese que manifestar a miña queixa por un aspecto concreto, faríao sen dúbida polo feito de non ser fácil de atopar esta publicación nas nosas librerías, debido á cativa distribución desta editorial no noso país. Porén, sempre nos queda o recurso da adquisición por Internet, algo máis que doado nos tempos que corren.

EDUARDO MOSCOSO MATO

Normas para a presentación de colaboracións

1. Os *Cadernos de Lingua* están dedicados á publicación de traballos de investigación sobre calquera aspecto da lingua galega, así coma á difusión de notas e informacións lingüísticas relevantes para o coñecemento e a promoción da lingua na sociedade. Tamén se publican recensións de obras de interese para eses mesmos fins.

2. Os traballos, notas ou recensións recibidos serán revisados por especialistas do Consello de Redacción ou do Comité Científico. En calquera caso a decisión final sobre a publicación correspóndelle ó Consello de Redacción.

3. A lingua de publicación é o galego, aínda que excepcionalmente poderán aparecer traballos escritos en portugués ou noutras linguas.

4. Os editores poderán efectuar correccións ortográficas, de puntuación ou de estilo sempre que sexan de pouca entidade e non afecten ó contido. Doutro xeito solicitarase a autorización do autor.

5. Sempre que sexa posible, o autor entregará dúas copias en papel e unha copia electrónica, compatible cos procesadores de textos máis usuais para PC. Os gráficos, de os haber, estarán inseridos no lugar correspondente do texto.

6. Nos artigos e notas aparecerá en primeiro lugar o título, seguido do nome do autor e do centro ou institución no que desenvolve o seu traballo. As recensións deben estar encabezadas pola referencia bibliográfica completa da obra comentada; o nome do autor incluírase ó final.

7. Recoméndase que o texto teña, como máximo, tres niveis de encabezamentos, numerados correlativamente 1., 2... 1.1., 1.2... 1.1.1., 1.1.2., etc., segundo o seu nivel na xerarquía.

8. As notas a pé de páxina deben evitarse na medida do posible. As citas inclúiranse no texto indicando entre parénteses o apelido ou

apelidos do autor, o ano de publicación da obra e, cando sexa conveniente, as páxinas; p.e., "... (Lorenzo 1996)" ou "... Kabatek (1996: 108)...". De seren necesarias, as notas ó pé deben aparecer indicadas cun número superescrito.

9. O final dos artigos e das notas deben aparecer recollidas todas as referencias bibliográficas, co seguinte formato:

Referencias bibliográficas

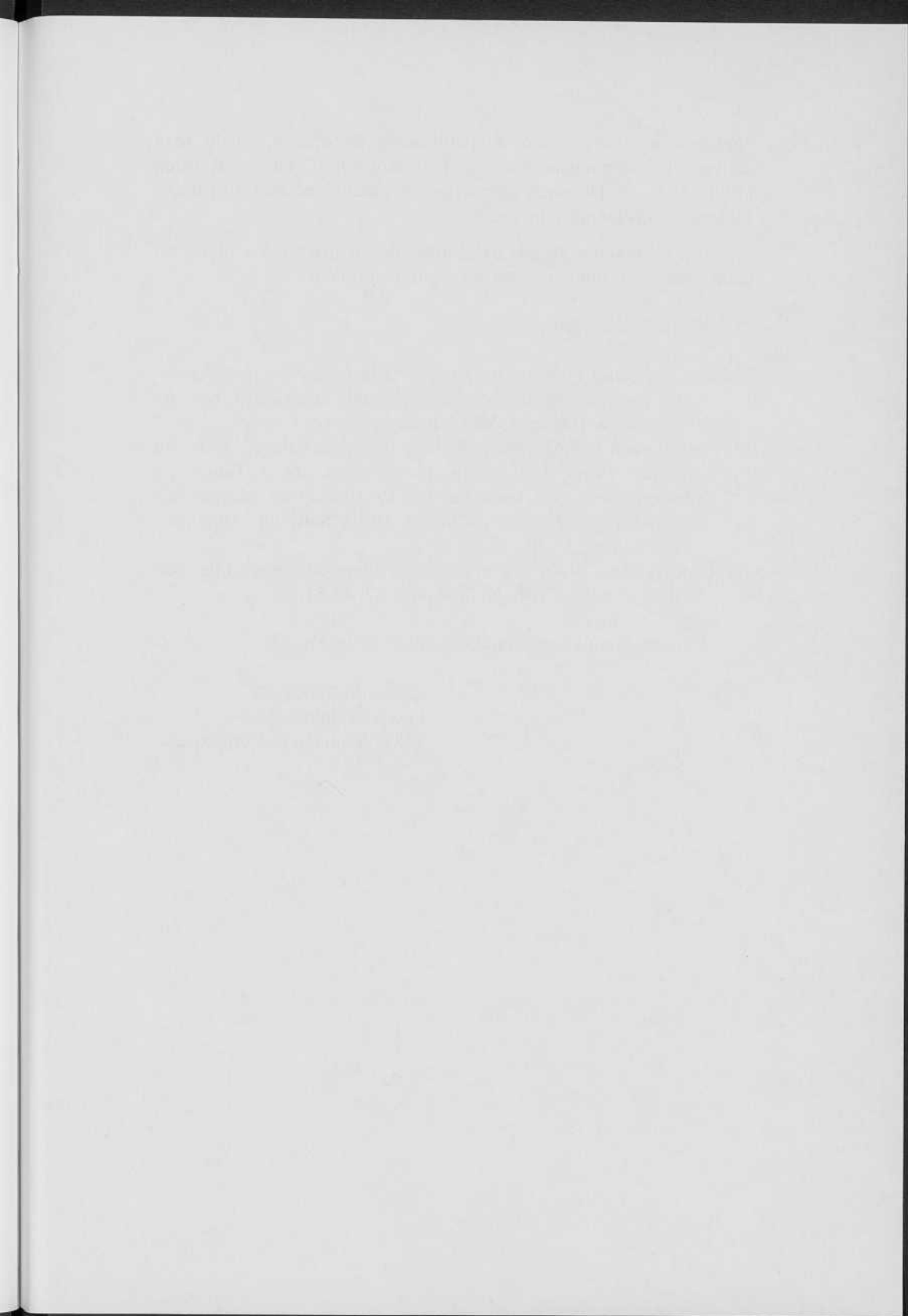
Kabatek, Johannes (1996): *Die Sprecher als Linguisten. Interferenz- und Sprachwandelphänomene dargestellt am Galicischen der Gegenwart*. Tübingen: Max Niemeyer.

Lorenzo, Ramón (1996): "Breve historia da lingua galega", in Jacyra Mota / Vera Rollemberg (eds.): *Atas do I Congresso Internacional da Associação Brasileira de Linguística (Universidade Federal da Bahia, 1994)*. Salvador: ABRALIN / FINEP / UFBA, 33-48.

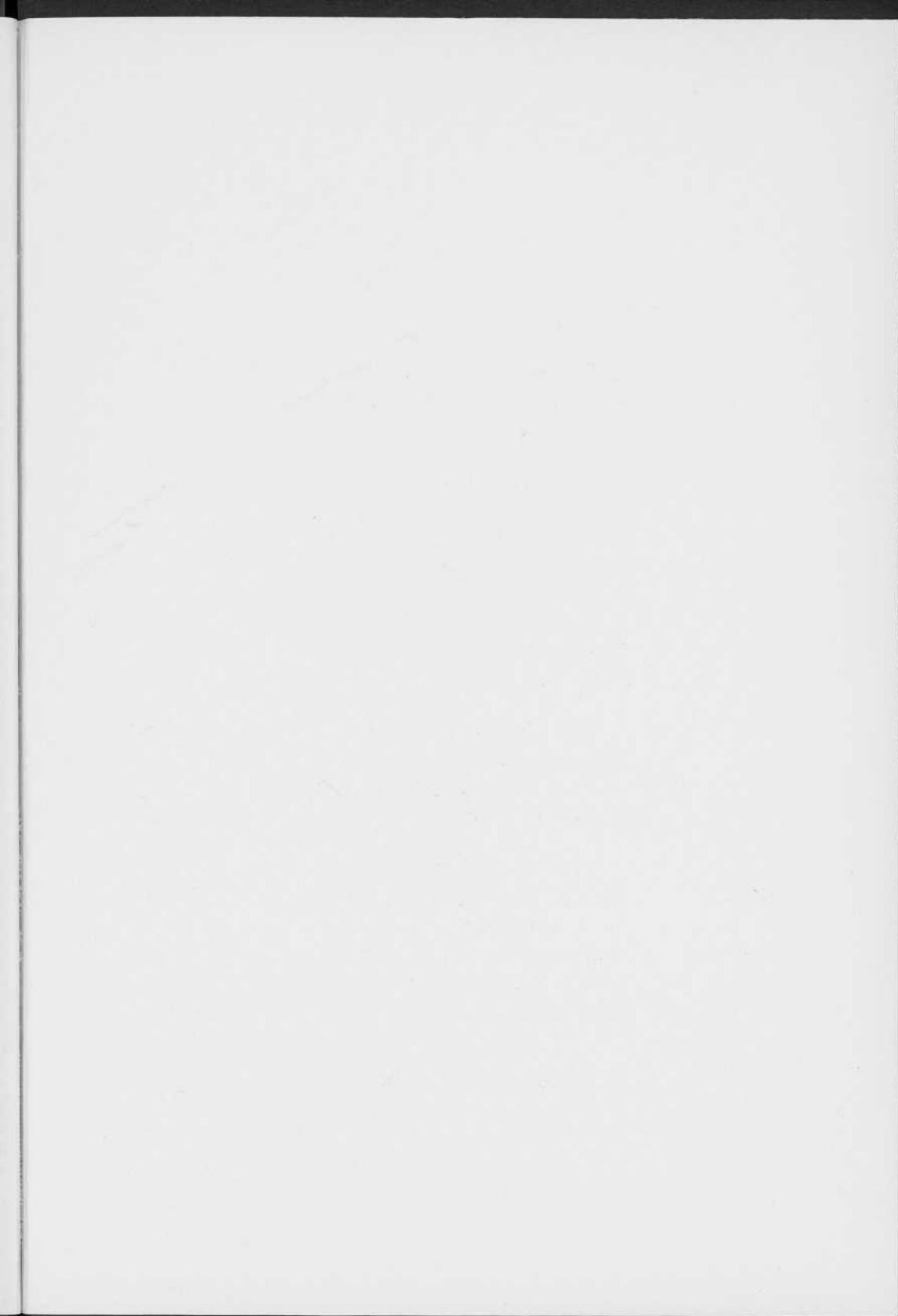
Asaka, Takekazu (1994): "O emprego do futuro do pretérito de indicativo en galego", *Studia Romanica* 27, 47-51.

10. Os orixinais para publicación enviaranse a:

XOSÉ LUÍS REGUEIRA
Paseo da Quinta, 23
15897 Santiago de Compostela









CONTIDO

- 5 X. H. Costas González,
A deturpación da toponimia galega
do Eo-Navia
- 27 S. Domínguez Portela,
Olla e mira, dous marcadores
discursivos en tres linguas:
portugués, galego e español
- 75 A. B. Escourido Pernas,
A prosodia de Camelle:
Descrición acústica dunha fala
- 123 A. Rodríguez Guerra,
A oración en galego medieval
(AOGM): O banco de datos
sintácticos medievais do ILG
- 151 E. Domínguez Noya,
Fco. M. Barcala Rodríguez,
M. A. Molinero,
Avaliación dun etiquetador
automático estatístico para o
galego actual: Xiada

CADERNOS de Lingua

30/31

