



Corpus de Referencia do Galego Actual  
(CORGA)

# **Relación de etiquetas empregadas na codificación do corpus**

**Versión 4.0    2022**

A continuación descríbense brevemente os diferentes valores que poden presentarse relacionados coas palabras que aparecen sobre fondo amarelo ou vermello. Colocando o punteiro do rato sobre a forma destacada en cor emerxe unha ventá aclaratoria coa información de a que marca de codificación corresponde.

## Etiquetas comúns aos distintos tipos de texto

As etiquetas que se empregan no corpus responden aos obxectivos seguintes:

- A) Impedir que texto non pertencente ao galego actual produza ruído no cómputo de formas que constitúen o corpus ou na recuperación da información.
- B) Proporcionar información sobre as características do segmento marcado que rodean a súa representación gráfica.
- C) Indicar que no lugar onde se atopa a marca hai texto que non é representable tendo en conta as posibilidades de estruturación que definimos.

### A) Impiden a consulta e indexación

Co fin de evitar que texto non pertencente ao galego actual produza ruído, o texto caracterizado coas etiquetas seguintes non pode ser obxecto de busca, aínda que si pode verse contextualmente destacado no sistema de consultas co fondo en cor amarela. Esas formas non se indexan, co que tampouco formarán parte das frecuencias que se ofrezan.

**Descoñecido** Corresponde a unha lingua inventada (falar de cans, corvos etc.) ou que non encaixa en ningún dos outros valores dispoñibles para *distinto*.

Abratt dar ob afg-te Idrawggin ith gorm na agulmat ob nob diremmag bad it shamash or dre , alom bathum , a gotdamde ish na bathum babg de teer na nhog shamash . [CORGA: Manuel Lourenzo González, *Arqueofaxia*, Edicións Xerais, 1995]

**Non normativo** Caracteriza texto reintegracionista ou que segue unha normativa de mínimos.

" En Fernán Velho achamos umha paisagem e umha voz , quer dizer , un reino onde o poeta reina solitário , e onde ao sumo , com emanada da própria luz , da própria sombra do dia e da noite , do abrente e do solpor desse mundo de formosura e enlevo , se forma umha figura que ten a graça da curva feminina , e ven a completar , a corroborar , a reafirmar o domínio absoluto do poeta como humanizaçom ou dominaçom , como mitificaçom e transposiçom da natureza sonhada , realizada e habitada polo modulador do canto " . [CORGA: Manuel María, *O Correo Galego* / CG2000-11-02/62, Editorial Compostela, 2000]

**Outra lingua** Indica que o texto que engloba pertence a unha lingua diferente do galego. Márcanse sempre con esta etiqueta os parágrafos e oracións que responden á característica anterior. Agora ben, cando se trata de palabras soltas non se

codifican a non ser que apareza a continuación a tradución ou explicitamente se indique que é unha unidade pertencente a outra lingua.

" The journalistic photographer can have no other than a personal approach , and it is impossible for him to be completely objective . [CORGA: Vítor Vaqueiro, *Sermos Galiza / SG2012-09-21/27*, Sermos Galiza S.A., 2012]

Cógelo Cándido , van para allá , márchate rápido . [CORGA: Berta Dávila, *A arte do fracaso / O último minuto de Cándido Vargas*, Edicións Barbantesa, 2010]

En Cataluña era coñecida polo nome de cap de mort de carbasson . [CORGA: *O Xornal de Galicia / XG2010-11-04/72*, Xornal de Galicia S.A., 2010]

En lugares como Somiedu, o oso ou bardión (que é o nome asturiano da especie) está a ser aceptado pola poboación ó ser unha boa fonte de ingresos (un turismo rural en aumento). [CORGA: Xermán García Romai, *Natureza Dixital. Retallos de natureza en formato dixital / Os últimos osos*, 2011]

## Outro período

Indica que o segmento destacado pertence a un estadio anterior da nosa lingua (anterior a 1955).

En 1493 , o Concello de Pontevedra , tendo en conta que " as mulleres de partido que a esta villa ven bebyr e abytar non ten casas lycitas nin honestas para seu oficio ... e por hevytar escándalos e males que por razón de sobre dito se podería reclecer , e porque as ditas mulleres del partido morasen apartadas por estremadas das mulleres casadas " ... acomoda deste xeito as chamadas mulleres da vida e pretende acabar coa prostitución libre e aberta intramuros, para convertela nunha actividade pechada e marxinal . [CORGA: Sabino Torres, *As tres columnas. Crónica sentimental da Moureira das putas*, Editorial Galaxia, 2008]

## B) Achegan información adicional

### Acoutación

Reservada unicamente para as obras de teatro e guións de series televisivas, esta etiqueta remite a secuencia á parte non dialogada do texto dramático onde se atopan as indicacións escénicas, ben orientadas á lectura, ben orientadas á posta en escena. A identificación dunha secuencia como texto integrado nunha acoutación só é visible dende o contexto, nunca dende as concordancias.

**Acoutación:** ( O Busto de Beethoven volve arrastrándose coa bandeixa de café .

**Acoutación:** O equilibrio é difícil pero conségueo .

**Acoutación:** Empeza a servir . )

[CORGA: Xosé Luís Martínez Pereiro, *Carne de estatua*, Bahía Edicións, 2001]

## Alongamento

Indica que no interior da palabra destacada había unha ou varias consoantes ou vogais repetidas (por exemplo, *veeeennnnn*). Tamén pode indicar a repetición de sílabas (por exemplo, *fafafafacer* ou *mi-milagre*) coa que se manifesta habitualmente a tartamudez na escrita. No texto que se incorpora ao corpus modifícase a forma para a súa grafía normal, suprimindo os caracteres ou sílabas reduplicados e trasladando a indicación de alongamento á etiqueta. Isto facilita o seu recoñecemento automático.

\_ **Non** ! \_ replicou a moza recuperando a fregona \_ . [CORGA: Francisco Galván, *As crendas de San Porfirio. Diario terapéutico dun esquizofrénico*, Edicións Morgante, 2012]

O Tatego: **Que non me poderei facer con el** ! [CORGA: Roberto Vidal Bolaño, *Touporroutou da Lúa e do Sol (Farsada choqueira para actores e bonecos, ou viceversa)*, Asociación Socio-Pedagóxica Galega, 1996]

## Interlocutor

Caracteriza os personaxes das obras de teatro ou guións televisivos, así como os intervenientes nunha entrevista, de xeito que os nomes desaparecen do texto en si, manténdose como metainformación do parágrafo e impedindo así a aparición, poñamos por caso, de 200 *Avelino*, 200 *Pousa* e 200 *Antelo*. A identificación dunha secuencia como texto emitido por un interlocutor concreto só é visible dende o contexto, nunca dende as concordancias. O nome do interlocutor aparece destacado en vermello e se se sitúa o cursor sobre el emerxe o valor *Interlocutor*.

**Álex:** En **cinco minutos** sae o coche de liña .

Interlocutor

**Raquel:** Cando saias da facultade ven pra a casa directamente .

**Raquel:** Voute buscar eu se queres .

[CORGA: Xosé Morais e Alberto Guntín, *Serramoura. Tempada 1, episodio 1: Homicidio*, 2014]

**Avelino Pousa Antelo:** Estiven **a pique de entrar** pouco antes no PSG .

Interlocutor

**Avelino Pousa Antelo:** O PGSD fomos aliados co PPG .

**Avelino Pousa Antelo:** Foi un fracaso .

[CORGA: Suso de Toro, *Camilo Nogueira e outras voces*, Edicións Xerais, 1991]

Nas transcricións, cando non se está seguro de que a intervención corresponde ao falante que se lle atribúe, márcase o interlocutor como *Asignación dubidosa*:

- ▶ **Xosé Luís Barreiro:** pero é o único <pausa/> que <pausa/> o único que están facendo <pausa/> non che parece a ti por aí <pausa/> Benxamín ?
- ▶ **Luis Villamor:** xa
- ▶ **Benxamín Casal:** si <pausa/> bueno <pausa/> é o **único que se permite** <pausa/> de momento <pausa/> eh
- ▶ **Xosé Luís Barreiro:** exacto
- ▶ **Benxamín Casal:** Interlocutor | Asignación dubidosa enin <pausa/> recorda <pausa/> a resposta que vos deu <pausa/> antonte <pausa/> Portomeñe <pausa/> pola noite

## Superíndice

Sinala que as letras e números caracterizados con esta etiqueta son índices que se colocan na parte superior dereita dun signo.

Os 315 municipios repártense entre as catro provincias cunha extensión media de 93,9 km<superíndice>2</superíndice> . [CORGA: Juan Luís Dalda Escudero, *A parroquia en Galicia: pasado, presente e futuro / O papel da organización parroquial na análise e na ordenación do modelo de asentamento poboacional*, Xunta de Galicia, 2009]

## Subíndice

Sinala que as letras e números caracterizados con esta etiqueta son índices que se colocan na parte inferior dereita dun signo.

Ben deberíamos recordar que un litro de auga embotellada xera 600 veces máis CO<subíndice>2</subíndice> que un litro da billa . [CORGA: Ramsés Pérez, *Cerna. Revista galega de ecoloxía e medio ambiente / CE2010-01/41*, ADEGA, 2012]

## Texto en táboa

Sinala que as palabras ortográficas destacadas aparecen no texto orixinal no interior dunha táboa, motivo que pode explicar a peculiar estrutura sintáctica na que se atopan.

**Pleuronectiformes ( planos ) : linguado , solla , rodaballo** [CORGA: Xoán Carlos Rodríguez Sánchez e Xesús León Guillén, *Pesca submarina en Galicia*, Edicións Lea, 1997]

### Idiófonos

Percutidos directamente : castañolas , claves , crócalos , látego , pratos

Percutidos con outro elemento : xilófono , marimba , vibráfono , glockenspiel , celesta , campás tubulares , triángulo , chocallos , gong e tam-tam , caixas chinesas , templeblocks

Sacudidos : ferreñas , sistro , axóuxeres , maracas , cabaza , chimes

Láminas de metal

Raspados e fretados : güiro , matracas , serrucho , harmónica de cristal

Punteados : guimbarda , caixa de música

[CORGA: Mónica Groba Lorenzo, *Introdución á linguaxe musical*, Editorial Galaxia, 2006]

## Texto en verso

Identifica o texto/fragmento marcado como pertencente a un poema.

**Panadeiriñas de Cea , aprendede a traballar , que non vos vai durar sempre o voso panadear.** [CORGA: Ánxel Sevillano, *Historias do canizo / O pan e as panadeiras*, Editorial Galaxia, 1979]

## Texto riscado

Sinala as palabras ortográficas que no texto orixinal aparecen grafadas cun risco por riba e que o autor quere manter polo motivo que sexa. A xeito de

relación paradigmática, ríscase unha palabra ou varias para indicar que o texto que segue substitúe o riscado. No CORGA esa información trasládase a unha etiqueta para facilitar o recoñecemento automático.

Como é natural , ao rapaz dálle un ataque de risa e , logo de **pasarme** pasarlle un envurullo de plástico negro que , ao despregalo , se convertería na bolsa máis fermosa que **vin** o papá viu na **miña** súa vida ( vale , o papá era eu e a nena Violeta ! , pero xuro que nese momento escoitei campás celestiais e mesmo cantar aos anxos ) sóltame : [CORGA: opaipai, *Se o vello Sinbad tivese fillos. Terapia de grupo sobre todas esas pequenas cousas de pequenos que fan que pais e nais nunca deixemos de ser bolseiros / Papi, piiiiiiiiiii!!!*, 2014]

### Texto unido

Indica que as formas caracterizadas con esta marca que se visualizan nunha grafía convencional aparecían no texto orixinal unidas con guión ou se soletreaban.

Unhas pirámides por aquí , unhas puntillas de Camariñas por alá , ben de Giorgi Line , un bigotiño pubertoso e xa tedes un estupendo Prince , ou " Oartistanatescoñecidocomoprince " , ou " TAFKAP " , ou simplemente " Elartista " , ou **como queira que se chame agora este home** . [CORGA: opaipai, *Se o vello Sinbad tivese fillos. Terapia de grupo sobre todas esas pequenas cousas de pequenos que fan que pais e nais nunca deixemos de ser bolseiros / Entroido 2015: DIY & BBB*, 2015]

Valentino: **Quero que veñas xa !** [CORGA: Roberto Salgueiro, *Valentino Rufini e Ákil Pillabán, de viaxe a Milán (e van sen un can). Obra de relixión, sexo e violencia para ver en familia*, Baía Edicións, 2009]

Ricardo: E non así polas boas , non , que houbo que formalizalo como el quixo , **documentadamente** , cedéndolle os dereitos da educación do neno . [CORGA: Manoel Riveiro Loureiro, *O coronel ten a quen lle escribir*, Edicións Xerais, 1989]

Naturalmente, é posible combinar as etiquetas anteriores, e en consecuencia tamén as formas destacadas en amarelo sinalan a información que corresponde a ambas as marcas. No caso seguinte a información que emerxe ao situarse enriba das palabras destacadas é *texto riscado-texto unido*:

Tiñades que verlle a cara de felicidade cando chegou á casa un paquete da avoa ferrolá que contiña dentro todo aquilo co que ela podía soñar ! ( algo así como flor corazón rosa con tarta , que é a expresión que usa ultimamente para dicir que algo **é a puta hostia** lle encanta ) . [CORGA: opaipai, *Se o vello Sinbad tivese fillos. Terapia de grupo sobre todas esas pequenas cousas de pequenos que fan que pais e nais nunca deixemos de ser bolseiros / As nenas xa non queren ser princesas*, 2015]

## C) Substitúen algo no texto

### Fórmula

Esta etiqueta emprégase cando existe unha fórmula para cuxa representación habería que botar man de caracteres especiais. Pode aparecer tamén substituíndo grafías gregas ou caracteres cirílicos.

A superficie desa esfera imaxinaria é **<fórmula/>** mentres que a superficie da xanela circular pola que nós miramos as estrelas é de **<fórmula/>** ou sexa : **<fórmula/>** [CORGA:

Ramón Vilalta López, *Astronomía práctica e outras cousas*, Toxosoutos, 1999]

## Táboa

Sinala que alí onde aparece esta etiqueta no texto orixinal, maiormente ensaios ou noticias de revistas, hai unha táboa referenciada no texto que non é representable coas posibilidades de estruturación que se definiron.

Segundo o Libro de montaría conservado no Arquivo Histórico Provincial de Lugo , entre xuño de 1586 e abril de 1591 foi abatido case un millar de lobos en terras lucenses (29).</oración>

<táboa/>

O cadro anterior reflicte un claro predominio da captura de lobos novos collidos nos niños onde as lobas pariran as súas crías , distribución que coincide coa observada noutras zonas da península \_ Ojeda , 2005 : 204 \_ . [CORGA: Hortensio Sobrado Correa, *Os animais domésticos na historia de Galicia / Aproximación á caza defensiva na idade moderna: "correr e bater os lobos, outras feras e animais nocivos" na Galicia dos séculos XVI-XIX*, tresCtres\_Editores, 2008]

Téñase en conta que o emprego na escrita das etiquetas *Palabra cortada*, *Alongamento*, *Texto riscado* e *Texto unido* límitase polo momento aos textos de máis recente incorporación. A introdución no corpus de transcricións e de blogs provocou a inclusión destas novas marcas para dar conta de fenómenos que usualmente non se recollen nos corpus, pero que pode ser interesante telos delimitados; non obstante, non se revisaron aínda todos os documentos xa incluídos en versións anteriores para aplicar as novas etiquetas.

## Etiquetas específicas para o oral:

### Alongamento

Indica que no interior da palabra destacada ten lugar a pronuncia prolongada dun ou varios sons, sen especificar a parte da palabra na que se produce. A raíz da súa introdución nas transcricións, aplícase esta marca cando cómpre tamén nos documentos máis recentes do CORGA cuxa orixe é a escrita.

ti crees **que** <pausa/> así polos agasallos <pausa/> que se coñece á persoa ?  
[CORGA: *Pensando en ti*. 18/06/1997]

### Inintelixible

Alí onde aparece sinala que o transcriptor escoita algo, mais non é quen de identificar que di o falante.

como decía aquel **<inintelixible/>** <pausa/> aquel señor <pausa/> se fai falta levala pola canción <pausa/> hai que levala pola canción [CORGA: *Diario Cultural*. 10/01/1997]

### Palabra cortada

Indica que o segmento marcado é un fragmento dunha palabra. As formas gráficas así marcadas non poden ser obxecto de consulta no sistema de recuperación de información, non se etiquetan e, por último, tampouco

aparecerán nas listaxes de frecuencias. A raíz da súa introdución nas transcricións aplícase esta marca cando cómpre tamén aos documentos do CORGA cuxa orixe é a escrita.

bueno <pausa/> vós **imaxinada** <pausa/> imaxinábadesvos <pausa/> algo <pausa/> hum <pausa/> semellante ó que ocorreu ? [CORGA: *Camiño de Volta*. 04/03/1996]

### Rindo

Indica que as palabras que abrangue o segmento destacado se emitiron entre risas do falante. Esta circunstancia pode ser importante para entender o texto sen ter que ir ó audio.

**si** <pausa/> **xa o sabemos** [CORGA: *Diario Cultural*. 10/01/1997]

### Risa

Esta marca substitúe as distintas interxeccións coas que se representa a risa dun falante cando esta non afecta a dicción.

**<risa/>** e xogas con el <pausa/> si ? [CORGA: *Xamón Xamón*. 11/01/1997]

### Ruído bucal

Reflicte os distintos ruídos bucais que pode realizar un falante.

de feito <pausa/> **<ruído bucal/>** o título Lobo nos outeiros fai un pouco referencia a esto <pausa/> non ? [CORGA: *Diario Cultural*. 30/12/1996]

### Sic

Aclara que a forma gráfica así marcada corresponde a un erro do falante na pronuncia dunha palabra e non a un erro do transcritor.

a infanta dona **Elela** de Borbón <pausa/> renovou esta mañá en representación do rei de España <pausa/> a ofrenda nacional ó Apóstolo Santiago [CORGA: *Informativo* 14:00. 25/07/1997]

### Texto solapado

Sinala que os segmentos textuais así marcados se emiten simultaneamente con outros. Na representación gráfica, para facilitar o recoñecemento automático, o solapamento incide sempre sobre a palabra completa, de xeito que esta nunca se corta cunha etiqueta.

**é case sempre posterior** <pausa/> **si** [CORGA: *Diario Cultural*. 10/01/1997]

### Transcrición dubidosa

Sinala que o transcritor non está seguro de que a palabra que transcribe sexa a que se pronuncia en realidade, malia que é a que cre escoitar.

para **gastiños** [CORGA: *Xamón Xamón*. 04/01/1997]

### Transcrición non documentada

Indica que se trata dunha transcrición cuxa representación gráfica non foi posible documentar, mais é a que corresponde ao son.

a partir das dez e media da noite <pausa/> o ballet moldavo **Koderanka** <pausa/> i a Orquesta Sinfónica Nacional de Moldavia <pausa/> representarán na Praza da



Igrexa de Lalín <pausa/> o espectáculo **Carmelu** [CORGA: *Informativo 14:00. 24/07/1997*]

[...]

Marca a omisión de texto transcrito nas intervencións sistemáticas que nunha tertulia ou entrevista se producen nunha lingua distinta do galego.

[...] sin lugar a dudas <pausa/> la mejor antología <pausa/> de la <pausa/> poesía gallego-portuguesa [CORGA: *Diario Cultural 10/01/1997*]

Pola contra, se aparece en laranxa [...], non ten nada que ver coa codificación dos textos, senón cunha marca que emprega a aplicación para indicar que por cuestións de espazo non se está amosando todo o texto que contén a secuencia, o cal si pode verse realizando outras buscas.

Por último, a xeito de recordatorio, cómpre remarcar que as unidades caracterizadas coas marcas *Descoñecido*, *Non normativo*, *Outra lingua*, *Outro período*, *Palabra cortada* e *Sic* non poden ser obxecto de busca nin aparecen nas diversas listaxes de frecuencias.