



INSTITUTO DA LINGUA GALEGA (ILG)

**MANUAL DE TRANSCRICIÓN DOS TEXTOS DE
CORTEGAL, CORPUS DE TEXTOS GALEGOS ESCRITOS
*POR ESTUDANTES NO ÁMBITO ACADÉMICO***

Novembro de 2022



CONSELLERÍA DE CULTURA,
EDUCACIÓN, FORMACIÓN
PROFESIONAL E UNIVERSIDADES

PGC2018-096069-B-I00

Contidos

1. Divisións do texto	3
2. Texto borrado.....	4
2.1 Normas xerais.....	4
2.2 Supresión de elementos contiguos.....	5
2.3 Uso de parénteses ou corchetes para indicar supresión de texto	6
2.4 Caracteres truncados.....	6
2.5 Lecturas dubidosas do texto borrado	7
2.6 Supresion dentro dunha supresión	7
2.7 Cancelación de borrado.....	7
3. Texto engadido.....	7
4. Texto borrado e texto engadido	8
5. Espazos en branco	9
6. Lecturas dubidosas.....	10
7. Texto ilexible	10
8. Texto destacado	10
9. Notas	10
10. Outras cuestións.....	11

A transcripción ten por obxectivo ofrecer unha reprodución fiel do texto orixinal, o que implica que se deben respectar as **grafías, puntuación, acentuación, abreviaturas e unión e separación de palabras** da/do estudante.

Ao tempo, nesta transcripción sinalaranse engadidos, riscaduras, emendas, dificultades de lectura etc. que se aprecien no texto orixinal do estudante (non se transcribirán as anotacións da ou do docente). Con respecto a esta cuestión, o obxectivo non é tanto reflectir na transcripción as características formais do manuscrito (de aí que non identifiquemos, por exemplo, o lugar do texto en que se introducen os engadidos ou as diferentes formas de borrar unha palabra) como reflectir o proceso compositivo da/do estudante. O usuario poderá optar por ver o texto "final", sen riscados, ou por unha visualización que inclúe e identifica as formas suprimidas.

1. Divisións do texto

O texto propiamente dito quedará demarcado por un elemento <text> (<text> </text>). O texto estará dividido en parágrafos marcados polo elemento <p> (<p>, </p>). Tamén se marcarán os saltos de liña mediante <lb/> (line beginning). A etiqueta <lb/>marca o comezo dunha nova liña e, en consecuencia, debe aparecer ao comezo de cada unha delas.

Tamén se marcarán as páxinas mediante <pb/> (que se incluírá aínda que haxa unha soa páxina). Así pois, todos os textos irán encabezados por <text><p><pb/><lb/>:

<text><p><pb/><lb/>Na actualidade onde apenas existen barreiras xeo entre países debido ao <lb/>rápido desenvolvemento da tecnoloxía e dos medios de comunicación podese coñecer <lb/>a cultura gastronóomía dun país a golpe de click ou mesmo traer esa gastronomía ao propio país <lb/>abrindo un restaurante especializado neste tipo de comida, isto fai que creza <lb/>a súa popularidade aumente.</p>

<p>Esta expansión da cultura mundial é debida as ganas <lb/>de coñecemento e de probar cousas novas sen saír da propia cidade <lb/>Pode parecer que este cambio é pasaxeiro, xa que ao sobre <lb/>explotar unha determinada cultura gastronómica pode provocar o aborrecemento<lb/> da mesma, o que finalizaría co peche de moitos locais dese tipo de comida. Ademáis coa recuperación económica en camiño <lb/>e progresando día a día, permite que as persoas ao ter maiores ingresos, <lb/>poidan viaxar a determinados lugares e probar a comida de forma autóctona sen necesidade <lb/>de pagar máis por, nalgúns casos, menos calidade. <lb/>Por isto podese dicir que en determinada comida existiría un cambio cultural <lb/>duradeiro debido a distancia que hai que percorrer para que ese destino ou a preferencia desa comida en respecto a propia. <lb/>Ademáis este cambio podería afianzarse debido a entrada de emigrantes, que ao inportar e levar consigo a cultura e os saberes <lb/>da súa terra, os convirten nun negocio.</p>

<pb/><p>Este aumento e entrada de novas gastronomías que amplían a oferta á hora <lb/>de comer xeran maior variedades e opcións, pero non hai que esquecer a propia gastronomía.</p></text>

No caso de que un salto de liña coincida cun corte de palabra, a etiqueta <lb/>colocarase no medio da palabra. O único que hai que ter en conta é que non pode haber espazos nin antes nin despois de <lb/> para que, posteriormente o tokenizador identifique a unidade como unha única palabra.

meti<lb/>culoso

2. Texto borrado

2.1 Normas xerais

Os fragmentos que no texto orixinal aparezan borrados mediante calquera sistema (riscado, borrancho etc.) marcaranse mediante un elemento . Escollendo o modo de visualización adecuado, estes fragmentos veranse riscados.

No caso de que o texto borrado sexa lexible, incorporase o dito texto tras o elemento coma nos seguintes exemplos, en que as formas suprimidas foron "que" e "consuman" respectivamente:

ao consumo que e ó diñeiro

dos produtos que consuman merquen.

A supresión tamén pode ter lugar dentro dunha palabra ou unidade gráfica, coma no seguinte exemplo:

mercados e pasaranse a

Nestes caso, é moi importante que non haxa espazos antes de (nin despois de , se a palabra continúa), para non romper a unidade da palabra.

Se, pola contra, o texto riscado é ilexible (sexa unha palabra, un carácter, unha secuencia de palabras ou de caracteres), empregárase un elemento e no seu interior un elemento baleiro <gap/> (o que na visualización dá lugar a [...]), con indicación, se é posible, do número de caracteres ("character") ou palabras ("word") riscadas (mediante os atributos @quantity e @unit):

<gap quantity="4" unit="character"/>

<gap quantity="3" unit="word"/>

Se resulta moi difícil precisar a cantidade, pode usarse o atributo @extent, de carácter máis descritivo:

<gap extent="several words" unit="word"/>

Este é o mesmo sistema que se empregará cando se percibe claramente que se borrou con típlex unha ou varias palabras ou un ou varios caracteres dentro dunha palabra e se deixou o espazo en branco.

Cando nunha secuencia borrada se entremesturan palabras ou caracteres lexibles con outros ilexibles usaremos tamén o elemento e no seu interior o elemento <gap/>

 casa de <gap quantity="1" unit="word"/>

2.2 Supresión de elementos contiguos

Cando hai varios elementos contiguos borrados pola/polo estudante haberá un único elemento se a riscadura parece afectar conxuntamente a toda a secuencia (independentemente de que o alumno, formalmente, incluíse unha única riscadura para toda a secuencia, como no primeiro exemplo que se ofrece, ou varias, unha para cada palabra, como no segundo). No primeiro caso bórrase o complemento da preposición *para* para substituílo por un complemento distinto e no segundo bórranse dous elementos pertencentes a unha perífrase, o que é claro indicio de que en ambos os casos estamos ante un único borrado conxunto.

para sacarlle maior proveito garantir

tendo que

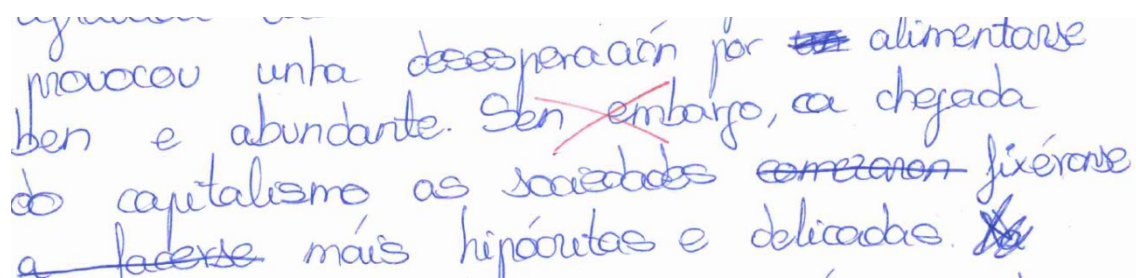
Cando a riscadura parece ter un carácter máis individual utilizaremos varios elementos . É o que sucede no seguinte exemplo, onde os elementos non forman parte dunha unidade sintáctica nin léxica:

teoría, porque é unha ~~teoría~~ de vida, ~~que~~ ~~está~~ ~~en~~ ~~contra~~ da ~~religión~~ ~~que~~ ~~está~~ ~~en~~ ~~contra~~'"/>

trata expón

A única excepción a este tratamento atopámolo naqueles casos en que o respecto á orde orixinal nos impide borrar conxuntamente unha unidade. Por exemplo, no seguinte caso, parece claro que "comezaron a facerse" debería borrarse conxuntamente. Con todo a/o

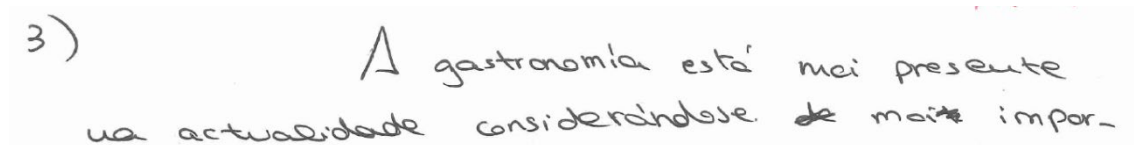
estudiante engadiu no espazo entre *comezaron* e *a* a palabra *fixéronse*, de modo que se queremos ser respectuosos co texto orixinal debemos separar os membros da unidade:



mostrou unha desesperación por ~~se~~ alimentarse
ben e abundante. Sen ~~embargo~~, co chegada
do capitalismo as sociedades ~~comezaron~~ *fixéronse*
~~a facerse~~ máis hipócritas e delicados.

comezaron <add>fixéronse</add> <lb/>a facerse

Por suposto cando se suprime unha palabra e un carácter doutra palabra contigua, é necesario introducir dous elementos :



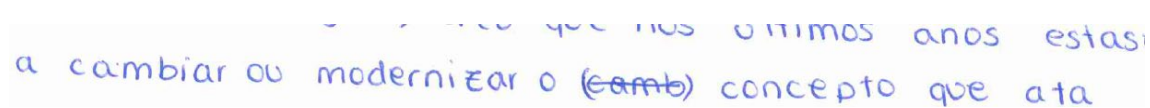
3) A gastronomía está moi presente
na actualidade considerándose ~~de~~ *moi* ~~t~~ impor-

de *moi* t

Os casos dubidos serán anotados para discusión.

2.3 Uso de parénteses ou corchetes para indicar supresión de texto

Algúns/algunhas estudantes empregan parénteses ou corchetes para indicar a eliminación dunha secuencia, frecuentemente en combinación co riscado. Cando as/os estudantes empregan as parénteses con esa funcionalidade, non deben ser incluídas na secuencia borrada, da mesma maneira que non se inclúe o riscado.



que nos últimos anos estase
a cambiar ou modernizar o ~~(camb)~~ *o* concepto que ata

o camb concepto

Nalgunhas ocasións é posible que haxa dúbidas sobre o valor das parénteses (como marca de supresión do texto ou como marca dun inciso ou aclaración, que a/o estudante decide borrar). Estes casos serán anotados para discusión.

2.4 Caracteres truncados

Nalgunhas ocasións poderemos encontrar caracteres truncados (por exemplo, un t ao que lle falta a raíña), que o estudante non acabou de construír porque decidiu que a palabra que estaba escribindo non era a máis adecuada e optou por riscala. En casos coma estes, "completaremos" o carácter se temos evidencias (formais e semánticas) sobre cal pode ser ese carácter. En caso contrario reproduciremos os caracteres completos da forma riscada e empregaremos <gap/> para o carácter que non logramos identificar.

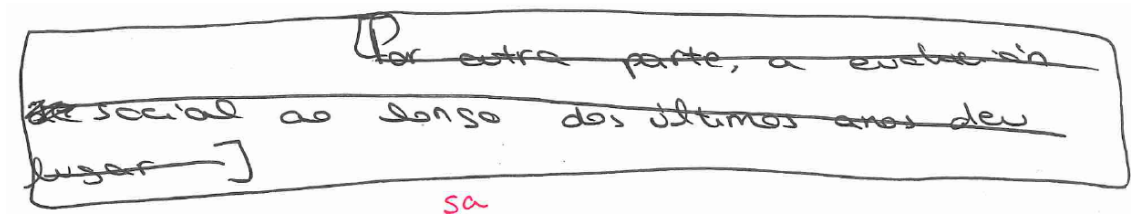
2.5 Lecturas dúbidas do texto borrado

Se a lectura dunha palabra, carácter... borrado é insegura, irá demarcada por un elemento <unclear> dentro de :

<unclear>casa</unclear>

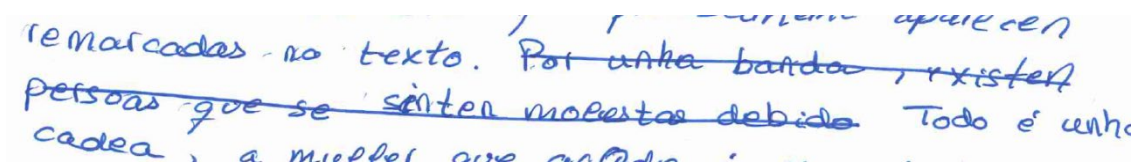
2.6 Supresión dentro dunha supresión

Se se borra un elemento dentro doutro, utilizaremos varios elementos , como no seguinte caso:



<p><lb/>Por outra parte, a evolución <lb/> de social ao longo dos últimos anos deu <lb/> lugar</p>

Ou nestoutro, onde ademais debemos incorporar unha adición (vid. §3):



no texto. Por unha banda, existen persoas que se se<add>i</add>nten molestas debido Todo é

2.7 Cancelación de borrado

É posible que unha/un estudante risque unha palabra e despois decida cancelar ese riscado (por exemplo, indicando "vale" cunha frechiña. Nese caso empregamos o elemento <restore> en combinación con , como no seguinte exemplo. Caso de utilizarse esa etiqueta debe anotarse para informar ao respecto, pois actualmente non se visualiza de ningún xeito:

<restore> casa</restore>

3. Texto engadido

O texto claramente engadido a posteriori (por exemplo, porque figura na interliña ou porque se engade mediante unha nota) indicárase simplemente mediante un elemento <add>, coma no seguinte exemplo, con independencia do sistema que a/o estudante empregase para introducir a adición:



a <add>comprar</add>

Escollendo o modo de visualización adecuado, o texto engadido quedará destacado en vermello.

Cando a autora / o autor inclúa un engadido nun texto xa de por si engadido, empregaremos varios elementos <add>. Supoñamos que alguén escribe "cando os grandes restaurantes" e engade na interliña "como Maruja Limón" e sobre este engadido engade a maiores por riba do texto "ou o Solla". O texto debería transcribirse así:

candos os grandes restaurantes <add>como Maruja Limón<add> ou o Solla </add></add>
traballan

Non se incluírán as marcas que a/o estudante emprega para remitir a un engadido (por exemplo un asterisco). Non deben confundirse os engadidos, que se integran na redacción do texto, coas notas a pé (vid. §9). No caso de que resulte difícil distinguir entre un engadido e unha nota a pé, debe consultarse.

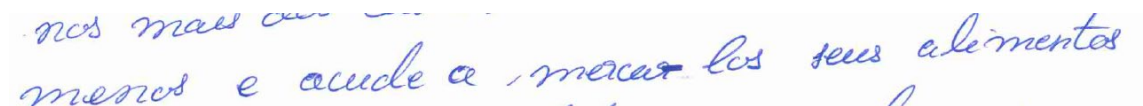
4. Texto borrado e texto engadido

En certas ocasións será necesario combinar os elementos e <add>. Faremos isto cando hai constancia evidente de que se produce eliminación e adición a posteriori, como no seguinte caso, en que se borra *ben* e se engade *adecuadamente* na interliña.



manexan ben <add>adecuadamente</add>

Empregarase este mesmo sistema de anotación cando se substitúen un ou máis caracteres por outros no texto.



e acude a mercar<add>-</add> los seus alimentos



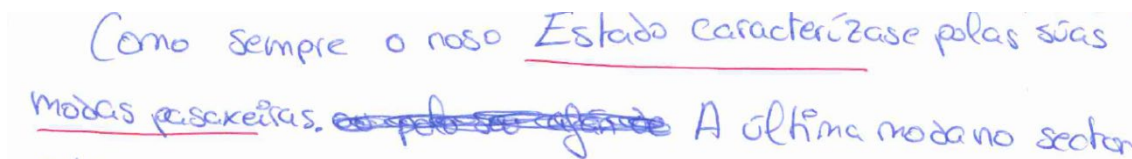
e a productiv<add>tiv</add>idade foron



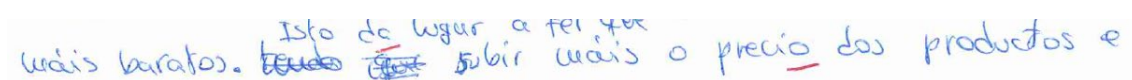
abandonar o<add>a</add>s úa<add>eu</add>s costumes

Cando os elementos ou <add> se introducen no medio da palabra, como nos casos anteriores, non pode haber espazos nin antes e despois das etiquetas.

Como se observa, en todos estes casos primeiro colocamos o elemento e despois o elemento <add>. Constitúen excepción os casos en que o elemento engadido aparece no texto antes do elemento borrado, como estes en que se introduce un signo de puntuación:



pasaxeiras<add>.</add> ou polo seu afán de



baratos <add>.</add> tendo que <add>Isto da lugar a ter que</add>
subir

Nestes casos a orde dos elementos virá determinada pola orde en que aparecen no texto o signo de puntuación engadido e a palabra ou secuencia borrada. No caso de que no primeiro destes fragmentos o signo de puntuación figurase despois do texto eliminado, o elemento debería preceder ao elemento <add>.

No caso de que a/o autora/autor do texto engada un texto (por exemplo, na entreliña) e despois o borre, usaremos combinadas os elementos e <add>, nesta orde, como no seguinte exemplo:

<add>na miña e na súa opinión</add>

Cando a/o estudante borra algo nun engadido, o elemento incluírase dentro de <add>. Supoñamos que engade na entreliña "na miña e na súa opinión" e despois borra "na súa". A etiquetaxe debe ser esta:

<add>na miña e na súa opinión</add>

5. Espazos en branco

Se o texto presenta un espazo en branco claramente perceptible (como por exemplo o deixado adrede para ser cuberto posteriormente cunha palabra que se descoñece no momento da redacción), empregárase o elemento <space/>, cos atributos @quantity e @unit:

o consumo <space quantity="4" unit="character"/> e

No caso de que apareza un espazo no texto, é conveniente acudir ao manuscrito orixinal para detectar se a/o estudante empregou típex. Neste caso, tal e como se indicou no apartado *Texto borrado*, empregárase o elemento cun elemento baleiro <gap/> no seu interior.

Caso de empregar o elemento <space/> debe deixarse constancia nun documento porque actualmente non se visualiza.

6. Lecturas dubidasas

As palabras cunha lectura dubidosa indicaranse mediante o elemento <unclear>. O usuario visualizaraas cun fondo azulado.

a xente <unclear>pensa</unclear>

7. Texto ilexible

As palabras ou secuencias ilexibles márcanse co elemento baleiro <gap/> seguido dos atributos @quantity e @unit:

<gap quantity="3" unit="character"/>

8. Texto destacado

É posible que a/o estudante destaque unha palabra ou secuencia de palabras. Para indicar este feito empregaremos o elemento <hi>, que irá acompañado do atributo @rend, co valor correspondente, que normalmente será "underline". Se utiliza outros sistemas para destacar, como cursiva, debe consultarse o valor de @rend. Unha secuencia como "el quere ir" representaríase así

el <hi rend="underline">quere</hi>

No caso de que se use o elemento <hi> debe anotarse, porque actualmente os valores de <hi> non se visualizan.

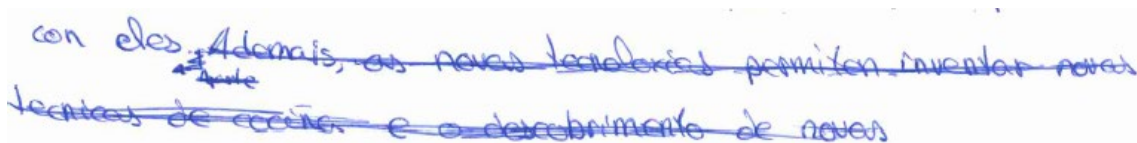
9. Notas

Se algunha/algún estudante incluíse unha nota (do tipo de nota a pé de páxina), esta será introducida co elemento <note>. A nota colocarse no lugar en que se remite a ela (mediante asterisco, número en superíndice etc.). O atributo n indica o número de referencia da nota:

en Masterchef <note n="1">Programa da TVE</note>

Como se indicou, en §3, non deben confundirse as notas a pé cos engadidos marcados mediante sistemas empregados habitualmente nas primeiras (un número ou un asterisco). En caso de dúbida, debe consultarse.

Tamén se presentarán como notas as indicacións formais que o estudante inclúe no texto, como neste caso, onde inicialmente se indicaba que o parágrafo debe ir á parte:

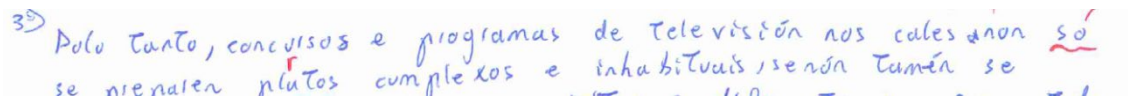


con eles. ^{Ademais} ~~as novas tecnoloxías permiten inventar novas~~
~~lextos de acción e o descubrimento de novas~~

con eles. <note>Aparte</note> Ademais, as novas tecnoloxías permiten inventar novas <lb/>

10. Outras cuestións

Non se transcribirá o número que serve para identificar a pregunta nin aqueles elementos que a/o estudante emprega para marcar que o texto continúa na páxina seguinte, que vén da páxina anterior, ou para contar o número de palabras que leva escritas... Por exemplo, neste caso, en que a/o estudante repite o número da pregunta na segunda páxina, este elemento non se transcribirá,



3) polo tanto, concursos e programas de televisión nos cales non só se crean novos complexos e inabituais senón tamén se

nin tampouco o símbolo ">" empregado ao final da última liña da páxina:



Non había aínda de consumismo na moda >

