Corpus del español del siglo XXI (CORPES)

Descripción del sistema de codificación

Libros y prensa

REAL ACADEMIA ESPAÑOLA

Corpus del español del siglo XXI *(CORPES)*
Descripción del sistema de codificación
Libros y prensa

Real Academia Española, 2013

Real Academia Española
Corpus del español del siglo XXI *(CORPES)*. Descripción del sistema de codificación. Libros y prensa [Recurso de Internet] / Madrid: Real Academia Española, 2013
Formato en PDF
Requisitos del sistema: Adobe Acrobat Reader

1. Real Academia Española-Publicaciones
2. Corpus Lingüísticos-Español
3. Corpus del español del siglo XXI (CORPES)-Crítica e interpretación

ISBN: 978-84-88292-13-1

Real Academia Española
Felipe IV, 4
28014 • Madrid

Académico responsable del proyecto *Corpus del español del siglo XXI (CORPES)*

Guillermo Rojo

Colaboradores

Mercedes Sánchez Sánchez (2006-), *coordinadora*

Julia Fernández Fernández (2015-)
Covadonga de Quintana Bermúdez la Puente (2008-)

Manuel José Santos Cobo (2007-2010)
Rufino Andrés Huertas (2008-2009)
Consuelo Mayor Andrés (2010-2013)
María Sancho Pascual (2011-2014)
Claret Ramos Saralegui (2007-2015)

INTRODUCCIÓN

En el XIII Congreso de la Asociación de Academias de la Lengua Española (ASALE), celebrado en Medellín en 2007, las academias acordaron encomendar a la Real Academia Española la construcción del *Corpus del español del siglo XXI (CORPES)*. Según el diseño aprobado, la primera fase del *CORPES* debería constar de un conjunto de 25 millones de formas para cada uno de los años comprendidos entre 2001 y 2012, procedentes de textos de los más diversos tipos y producidos en todos los países de habla hispana. Esos 25 millones de formas anuales, seleccionados según los parámetros del diseño del corpus, tienen una distribución geográfica que se reparte en un 70 % para textos producidos en América y un 30 % para textos producidos en España. Además, todas las formas incluidas deben recibir anotación morfosintáctica y lematización.

Un corpus textual es un conjunto muy amplio de textos de los más diversos tipos, representativos del estado de una lengua, en formato electrónico y codificados de modo que sea posible obtener de él la información que requiere la investigación lingüística en cualquiera de sus ramas. Del análisis del contenido de un corpus es posible, por ejemplo, extraer datos acerca de la frecuencia de una forma o un lema, los casos de una determinada expresión o los elementos que con mayor frecuencia se combinan con otro. Pero lo fundamental en un corpus de referencia, tal como se concibe en la actualidad, consiste en que cualquiera de esas extracciones pueda hacerse no solo para la totalidad del corpus, sino para cualquiera de los subconjuntos que se pueden configurar en su interior de forma dinámica. La recuperación selectiva de la información contenida en, por ejemplo, el *CORPES*, consiste en la posibilidad de obtener todos los casos de una determinada palabra o sus coapariciones en, por ejemplo, las noticias de prensa sobre economía y finanzas publicadas en la prensa guatemalteca entre 2004 y 2007. Lo obtenido en una consulta como esa puede ser luego comparado con lo que resulta de textos de otro tipo, otro tema, otro país u otro período.

La gran ventaja de un corpus como el *CORPES* consiste, pues, en la posibilidad de hacer recuperación selectiva de la información contenida en los numerosos subconjuntos que se pueden establecer en él. El modo de lograrlo es, naturalmente, la codificación adecuada de los textos, gracias a la cual se consigue que el sistema de búsqueda pueda determinar siempre a cuál de los subconjuntos posibles pertenece un texto y, claro está, todas las formas incluidas en él. Una codificación adecuada —todo lo pertinente y nada más que lo pertinente— es, pues, la base fundamental sobre la que se constituye todo lo demás. El presente documento expone con detalle el sistema que se ha seguido para la codificación del enorme conjunto de documentos que componen la primera fase del *CORPES*.

I. DISEÑO DEL *CORPUS DEL ESPAÑOL DEL SIGLO XXI*

El *CORPES* pretende ser un corpus de referencia que se mueva en los parámetros utilizados actualmente en esta línea de trabajo: 25 millones de formas por año y una distribución general del 70 % para textos americanos y el 30 % para textos españoles. Se concibe como un corpus semiabierto, es decir, un corpus que se irá incrementando en los próximos años con las cantidades previstas. Los textos que integran el *CORPES* han sido seleccionados de acuerdo con los siguientes parámetros:

Medio

El 90 % de los textos corresponde a la lengua escrita y el 10 % a la lengua oral.

Soporte (textos escritos)

Los materiales escritos proceden de libros (40 %), publicaciones periódicas (40 %), material de Internet (7,5 %) y miscelánea (2,5 %).

Geográfico

La distribución general del *CORPES* asigna un 30 % del total a formas procedentes de España y un 70 % a formas procedentes de América.

El material producido en América se clasifica, a su vez, en las zonas lingüísticas habituales: andina, Antillas (caribeña), Caribe continental, chilena, Estados Unidos, México y Centroamérica y Río de la Plata.

La distribución correspondiente a las grandes áreas lingüísticas del mundo hispánico se establece mediante el cruce de criterios diferentes, entre los que figuran la población, el volumen de publicaciones, la cantidad de ediciones digitales de periódicos y revistas, etc.

Es importante señalar, como una de las novedades en los corpus académicos, que el *CORPES* incluye también textos correspondientes a Guinea Ecuatorial y a Filipinas.

España		
América	Andina	Bolivia, Ecuador, Perú
	Antillas (caribeña)	Cuba, Puerto Rico, República Dominicana
	Caribe continental	Colombia, Venezuela
	Chilena	Chile
	México y Centroamérica	Costa Rica, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá
	Río de la Plata	Argentina, Paraguay, Uruguay
	Estados Unidos	
Filipinas		
Guinea Ecuatorial		

Temático

Todos los textos siguen una clasificación temática común, lo que posibilita la búsqueda por materias de una forma, independientemente de su realización oral o escrita.

La clasificación temática distribuye los textos escritos en dos grandes bloques, ficción y no ficción, que, a su vez, se escinden en distintas áreas temáticas:

Bloque	Tema
No ficción (Libros y prensa)	Actualidad, ocio y vida cotidiana
	Artes, cultura y espectáculos
	Ciencias sociales, creencias y pensamiento
	Ciencias y tecnología
	Política, economía y justicia
	Salud
Ficción (Libros)	Guion
	Novela
	Relato
	Teatro

Tipos de texto

Además de las asignaciones procedentes del medio y el soporte, los textos reciben una caracterización por tipo o género textual: novela, relato, teatro o guion para los textos de ficción; noticias, reportajes, opinión, crónica, etc., para periódicos y revistas; prosa académica y no académica; entrevistas, conversaciones, etc., para orales; texto escrito para ser leído (noticias de radio o televisión), etc.

El cruce del género con el medio, el soporte y el área temática produce una riquísima tipología textual que permite a los investigadores afinar considerablemente la recuperación selectiva de la información desde la aplicación de consulta.

Bloque	Soporte	Géneros (Ficción) Temas (No ficción)	Tipo de texto
Ficción	Libro	Guion Novela Relato Teatro	Ficción
No ficción	Libro	Actualidad, ocio y vida cotidiana Artes, cultura y espectáculos Ciencias sociales, creencias y pensamiento Ciencias y tecnología Política, economía y justicia Salud	Académico Biografía, memoria Divulgación Jurídico-administrativo Libro de texto
	Prensa	Actualidad, ocio y vida cotidiana Artes, cultura y espectáculos Ciencias sociales, creencias y pensamiento Ciencias y tecnología Política, economía y justicia Salud	Académico Carta al director Crítica Crónica Divulgación Editorial Entrevista Noticia Opinión Reportaje Varios

II. CODIFICACIÓN DEL *CORPUS DEL ESPAÑOL DEL SIGLO XXI*

El *CORPES* es un corpus codificado o etiquetado, formado por textos sometidos a un proceso de marcación que tiene como fin prepararlo para una fase posterior de análisis.

Sistema de codificación

El sistema de codificación, en XML y basado en la *Text Encoding Initiative*, ha sido desarrollado íntegramente en la Real Academia Española. Se recoge en la DTD utilizada para la validación de los textos.

Por la experiencia acumulada en la codificación de otros corpus de la Academia *(CREA* y *CORDE),* se ha concluido que la etiquetación debe concebirse como un medio para recuperar la información y nunca como un fin en sí misma; ha de ser, además, objetiva y no interpretativa. Ese es el propósito que orienta el desarrollo de este sistema de codificación.

Cada texto integrado en el *CORPES* está, pues, etiquetado y estructurado en dos niveles: la cabecera (con datos sobre el archivo fuente y el texto electrónico) y el texto propiamente dicho.

El sistema de etiquetado se centra principalmente en la codificación de la cabecera. Cada texto incorporado al *CORPES* debe estar perfectamente documentado (país, año, zona geográfica, tema, tipo de texto, etc.) a través de los elementos, atributos y valores contenidos en la cabecera, donde se incorporan los datos bibliográficos habituales (autor, título, editorial, fecha, etc.) y aquellos que lo sitúan en cada uno de los subconjuntos en que puede ser analizado el corpus y que coinciden con los parámetros de su diseño: medio, zona, país, soporte, área temática, tipo de texto y año. A través de la aplicación de consulta, cada uno de estos parámetros de selección puede ser combinado con todos los demás, de modo que es posible obtener los casos de una palabra o expresión procedentes de, por ejemplo, noticias de prensa publicadas en Panamá en 2010 y que traten de economía.

En el cuerpo del texto se ha reproducido, en XML, la marcación tipográfica original relevante para las consultas: párrafos y marcas de formato. En ocasiones se prescinde de aquellos fragmentos que no introducen texto analizable —tablas, cuadros o ilustraciones— y se marca adecuadamente su eliminación.

Al proceso de marcación le acompaña siempre una lectura atenta del texto fuente destinada a revisar de manera exhaustiva erratas tipográficas o errores de lectura de escáner. El texto electrónico resultante debe reproducir fielmente las formas originales del texto fuente.

Existen en el mercado diversas herramientas para la codificación de corpus. El *CORPES* utiliza el editor *Oxygen XML* para etiquetar y validar sus textos de acuerdo con la DTD, que asegura la correcta introducción de todas las etiquetas y, en procesos posteriores, la adecuada manipulación de las marcas para conseguir la recuperación selectiva de la información introducida.

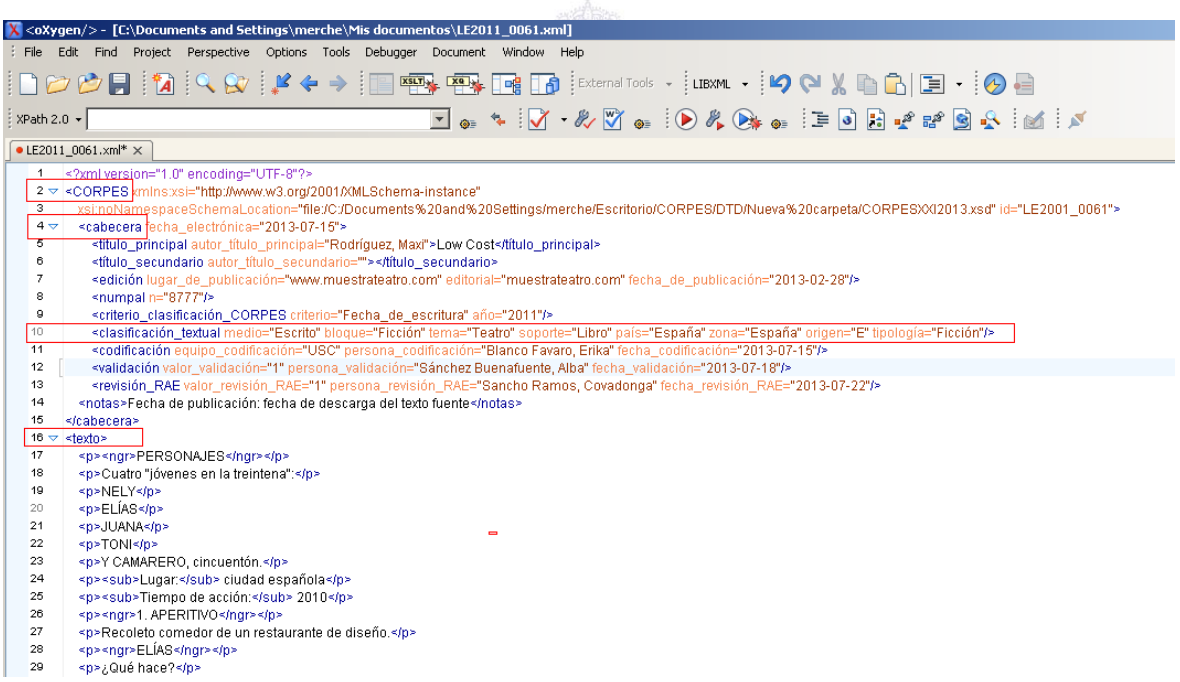
Estructura de los textos codificados

En el *CORPES* se incorporan dos clases de fuentes de información (textos): unitarios y anidados. Los unitarios (una monografía o una novela, por ejemplo) constituyen en sí mismos una fuente principal y los anidados son parte de una fuente principal (las noticias o los relatos son textos anidados en una unidad superior: un periódico o una recopilación de relatos).

Los textos se estructuran en dos niveles: cabecera y texto. A su vez, la cabecera contiene tres tipos de datos:

- Datos bibliográficos (referencia bibliográfica del texto).
- Datos de clasificación del texto en el corpus (en los parámetros de su diseño).
- Datos sobre el archivo electrónico (número de palabras, equipo de codificación, nombres de los revisores, etc.).

Por su parte, el texto se estructura en párrafos, y en los párrafos se codifican los formatos tipográficos habituales (cursiva, negrita, subrayado y versalita). También se da cuenta de algún otro fenómeno relacionado con elementos no representables o dudosos.



Pantalla de trabajo

Descripción de los elementos de codificación

Cada elemento o etiqueta puede llevar uno o varios atributos que, a su vez, recogen diferentes valores. Todos ellos se definen en el esquema de codificación.

La sintaxis general de los elementos, atributos y valores puede representarse de este modo:

```
<elemento atributo1=”valor1_del_atributo1”
atributo2=”valor3_del_atributo2”>Texto etiquetado por esos elementos, atributos
y valores</elemento>
```

Los elementos del *CORPES* y sus correspondientes atributos y posibles valores se enumeran y describen a continuación:

```
<CORPES id="">…</CORPES>
```

Descripción: delimita cada una de las unidades textuales o fuentes de información que componen el corpus.

<CORPUS DEL ESPAÑOL DEL SIGLO XXI>
    <CORPES [primer texto]>
    …
    </CORPES>
    <CORPES [segundo texto]>
    …
    </CORPES>
    Etc.
</CORPUS DEL ESPAÑOL DEL SIGLO XXI>

Atributos:

[ID], su valor coincide con el nombre del archivo y se construye con datos referentes a los criterios de diseño: soporte, origen y año. Se añade un número correlativo para cada archivo con relación a la numeración que se sigue en el *CORPES*.

Sintaxis:

```
<CORPES id="PA2001_0012_012"> [Artículo de prensa procedente de América y
clasificado en 2001]
<CORPES id="LE2002_0001"> [Libro procedente de España y clasificado en
2002]
```

`<cabecera fecha_electrónica="">…</cabecera>`

Descripción: recoge los datos bibliográficos, los de clasificación y los del archivo electrónico.

Atributos:

[`fecha_electrónica`], cuyo valor recoge la fecha de creación del archivo .XML, en formato «AAAA-MM-DD».

Sintaxis:

```
<cabecera fecha_electrónica="2008-01-14">…</cabecera>
```

`<título_principal autor_título_principal="">…</título_principal>`

Descripción: título del texto principal. En el caso de la prensa, se refiere al nombre del periódico. La fuente de información para el título y para la mención de responsabilidad será la portada del libro[1].

Mención de responsabilidad: nombre de las personas o entidades responsables directa o indirectamente del contenido del documento.

Deben tenerse en cuenta las siguientes particularidades:

- Los nombres de hasta tres autores se recogen en formato «apellidos, nombre» y se separan por «;»; las entidades se separan por «,».
- Si hay más de tres autores, se toma el nombre del primero seguido del signo de omisión (…) y de la abreviatura [et al.].
- No se consignan términos que expresen dignidades o cargos.
- Si la responsabilidad es la de editor, compilador, coordinador o director, se hace constar entre corchetes: [ed.], [coord.].

En el sistema de codificación del *CORPES* la mención de responsabilidad se recoge en el atributo «`autor`» (principal o secundario), del elemento del título correspondiente.

Atributos:

[`autor_título_principal`], que recoge el autor o responsable principal. En el caso de la prensa, queda vacío.

[1] Para más información, véase <http://www.mcu.es/publicaciones/docs/MC/Reglas_Catalogacion/descp_bibliografica_1.pdf.> [Consulta: 07/11/2013]

Sintaxis:

```
<título_principal autor_título_principal="">El País</título_principal> [Prensa]
<título_principal autor_título_principal="Freire, Espido">Cuando comer es un
martirio</título_principal> [Libro]
```

`<título_secundario autor_título_secundario="">...</título_secundario>`

Descripción: título del texto anidado. En prensa se trata del titular de la noticia[2]. En el caso de libros, puede tratarse de

- una obra colectiva o una compilación; el título secundario es el del capítulo;
- un libro de relatos; el título secundario es el del relato que se codifica;
- un relato que aparece en una revista literaria; el título secundario es el del relato y el principal, el de la revista.

En los textos unitarios el elemento no aparece.

Atributos:

[autor_título_secundario], que recoge el autor del texto anidado. En el caso de la prensa, es el lugar previsto para el autor del artículo, si se conoce; en caso contrario, el atributo queda vacío[3].

Sintaxis:

```
<título_secundario autor_título_secundario="Benítez, Wilfrido">Media docena de
delincuentes asalta una tienda</título_secundario> [Prensa]
<título_secundario autor_título_secundario="Muñiz, Carlos ... [et al.]">El
proyecto geopiratería: el caso del Ecuador</título_secundario> [Libro]
```

[2] Se respetan la acentuación y ortografía originales de los titulares, salvo errores evidentes.
[3] Si únicamente aparecen las iniciales del autor, el atributo queda vacío.

```
<edición lugar_de_publicación="" editorial=""
fecha_de_publicación=""/>
```

Descripción: elemento vacío. Reproduce en XML la cita bibliográfica y recoge los datos referentes a la edición del texto fuente.

Atributos:

[lugar_de_publicación], que recoge la localidad donde tenga su sede el editor. Si no aparece el dato, se utiliza «s. l.».

[editorial], que recoge la editorial responsable de la publicación. Si no aparece el dato, se utiliza «s. n.». Si se trata de un texto procedente de Internet, se consigna la dirección electrónica donde aparece el texto fuente sin los datos «http://www».

[fecha_de_publicación], su valor es el de la fecha de la edición utilizada. Si se trata de un libro, el formato es «AAAA». Si se trata de prensa, el formato es «AAAA-MM-DD». Si se trata de textos de ficción descargados de la red sin datos editoriales, se consigna como fecha de publicación la de adquisición del texto fuente. Además, se deja constancia de esta circunstancia en el elemento <notas>.

Sintaxis:

```
<edición lugar_de_publicación="Osorno" editorial="australosorno.cl"
fecha_de_publicación="2006-03-14"/>
```

```
<numpal n=""/>
```

Descripción: elemento vacío. Marca el número de formas de cada texto.

Atributos:

[n], cuyo valor es el número de formas del texto.

Sintaxis:

```
<numpal n="375"/>
```

```
<criterio_clasificación_CORPES criterio="" año=""/>
```

Descripción: elemento vacío. Recoge el año (AAAA) de clasificación del texto en el *CORPES* y el criterio seguido para esta clasificación.

Atributos:

[criterio], cuyos posibles valores son:

Criterio	Soporte
[Primera_edición]	Libro, prensa
[Fecha_de_escritura]	Libro (obra de teatro, guion)
[Fecha_de_estreno]	Libro (obra de teatro, guion)
[Ver_nota][4]	Libro, prensa…

[año], da cuenta del año, en formato «AAAA», que se corresponde con el criterio de selección cronológico[5].

Sintaxis:

```
<criterio_clasificación_CORPES criterio="Primera_edición" año="2006"/>
```

[4] Se utiliza el valor [Ver_nota] cuando se proporciona información adicional no prevista sobre el texto seleccionado: nacionalidad del autor, etc.
[5] El criterio cronológico general de incorporación al *CORPES* es el de primera edición, si bien, en el caso de las obras de teatro o los guiones, en las que se pueden tener en cuenta la fecha de escritura, la de estreno y la de primera edición, se considera la más antigua para su incorporación al corpus.

```
<clasificación_textual medio="" bloque="" tema="" soporte="" país=""
zona="" origen="" tipología=""/>
```

Descripción: elemento vacío. Recoge la clasificación del texto en los parámetros de selección del *CORPES*.

Atributos y posibles valores:

[medio]: Escrito | Oral

[bloque]: Ficción | No_ficción

[tema]: Actualidad_ocio_y_vida_cotidiana | Artes_cultura_y_espectáculos | Ciencias_sociales_creencias_y_pensamiento | Ciencias_y_tecnología | Política_economía_y_justicia | Salud
Guion | Novela | Relato | Teatro

[soporte]: Libro | Miscelánea | Prensa |Internet

[país]: Argentina | Bolivia | Chile | Colombia | Costa_Rica | Cuba | Ecuador | El_Salvador | España | Estados_Unidos | Filipinas | Guatemala | Guinea_Ecuatorial | Honduras | México | Nicaragua | Panamá | Paraguay | Perú | Puerto_Rico | República_Dominicana | Uruguay | Venezuela

[zona]: Andina | Antillas | Caribe_continental | Chilena | España | Estados_Unidos | Filipinas | Guinea_Ecuatorial | México_y_Centroamérica | Río_de_la_Plata

[origen][6]: A | E | G | F

[tipología]: Académico | Biografía_memoria | Carta_al_director | Crítica | Crónica | Divulgación | Editorial | Entrevista | Ficción | Jurídico_administrativo | Libro_de_texto | Noticia | Opinión | Reportaje | Varios

Sintaxis:

```
<clasificación_textual medio="Escrito" bloque="No_ficción"
tema="Ciencias_sociales_creencias_y_pensamiento" soporte="Prensa"
país="Argentina" zona="Río_de_la_Plata" origen="A" tipología="Noticia"/>
```

```
<codificación equipo_codificación="" persona_codificación=""
fecha_codificación=""/>
```

Descripción: elemento vacío. Recoge los datos del codificador.

Atributos:

[equipo_codificación], cuyo valor es el nombre del equipo responsable de la codificación del texto.

[6] A: América
E: España
G: Guinea Ecuatorial
F: Filipinas

[persona_codificación], recoge el nombre de la persona que codifica el texto, dentro del equipo.

[fecha_codificación], toma el año, mes y día en el que se realiza la tarea. Se considera la fecha de codificación del texto, no de la creación de la cabecera, que lleva su propia fecha. Puede coincidir o no con esta última.

Sintaxis:

```
<codificación equipo_codificación="UAB" persona_codificación="Andrés Ramos,
José" fecha_codificación="2008-02-05"/>
```

```
<validación valor_validación="" persona_validación=""
fecha_validación=""/>
```

Descripción: elemento vacío. Recoge los datos del revisor del equipo que realiza la codificación.

Atributos:

[valor_validación], recoge con los valores 1, 2 o 3, los distintos momentos de la validación, si los hubiera.

[persona_validación], recoge el nombre del responsable final de la validación.

[fecha_validación], toma el año, mes y día de la validación final.

Sintaxis:

```
<validación valor_validación="1" persona_validación="Ramos Santos, Mercedes"
fecha_validación="2008-02-04"/>
```

```
<revisión_RAE valor_revisión_RAE="" persona_revisión_RAE=""
fecha_revisión_RAE=""/>
```

Descripción: elemento vacío. Recoge los datos de la revisión final en la Real Academia Española.

Atributos:

[valor_revisión_RAE], recoge diferentes momentos de la revisión.

[persona_revisión_RAE], recoge el nombre de la persona responsable de la revisión.

[fecha_revisión_RAE], lleva el valor de la fecha en la que se realiza esta tarea.

El formato de los atributos es el mismo que el de las etiquetas anteriores.

Sintaxis:

```
<revisión_RAE valor_revisión_RAE="1" persona_revisión_RAE="Pérez Serrano, Ana"
fecha_revisión_RAE="2008-03-12"/>
```

`<notas>…</notas>`

Descripción: reproduce las particularidades que el codificador desee incluir, bien relativas al texto fuente, bien a la codificación.

Sintaxis:

```
<notas>Finalista del Premio Casa de América - Festival de Escena Contemporánea
de Dramaturgia Innovadora 2005. Fecha de publicación: fecha de descarga del
texto fuente</notas>
```

`<texto>…</texto>`

Descripción: contiene las formas del texto. Se estructura en párrafos.

`<p>…</p>`

Descripción: marca los párrafos del texto.

Sintaxis:

```
<p>Marca los párrafos del texto</p>
<p>Puede incluir marcas de formato</p>
```

Puede incluir marcas de formato y marcas relativas a fenómenos concretos:

`<sub>…</sub>`

Descripción: marca el texto subrayado.

Sintaxis:

```
<sub>Texto subrayado</sub>
```

```
<csv>…</csv>
```

Descripción: marca el texto en cursiva.

Sintaxis:

```
<csv>Texto en cursiva</csv>
```

```
<ngr>…</ngr>
```

Descripción: marca el texto en negrita.

Sintaxis:

```
<ngr>Texto en negrita</ngr>
```

```
<vrs>…</vrs>
```

Descripción: marca el texto en versalitas[7].

Sintaxis:

```
<vrs>MARCA VERSALITAS</vrs>
```

```
<csvngr>…</csvngr>
```

Descripción: marca el texto en cursiva y negrita.

Sintaxis:

```
<csvngr>Fragmentos en cursiva y negrita a la vez</csvngr>
```

[7] Los números romanos se codifican en mayúsculas; en este caso, se prescinde de las versalitas para evitar un exceso de marcación que no aporta información adicional.

```
<sic>…</sic>
```

Descripción: marca fragmentos confusos[8].

Sintaxis:

```
<p>El menú ofrecido consistirá de una comida limitada, completamente en
ocasiones con una serie de botanas calientes, cada una de las cuales tendrá un
precio diferente. <sic>ral ningún indicio que permita supone</sic></p>
```

```
<nrp/>
```

Descripción: elemento vacío. Marca fragmentos no representables: listas, tablas, fórmulas… Es un elemento sin contenido.

Sintaxis:

```
<p>Podemos verlo en la tabla que sigue a continuación<nrp/></p>
```

```
<rsi>…</rsi>
```

Descripción: resalte sin identificar. Marca fragmentos en los que la intención de resalte del autor resulta clara. Su uso está restringido a aquellos casos en los que no se pueda utilizar ninguna otra etiqueta de las expuestas más arriba.

Sintaxis:

```
<p><rsi>ME
        TRO
        PO
        LIS</rsi></p>
```

[8] Las erratas evidentes se corrigen sin dejar ningún tipo de huella.

Resumen del esquema de codificación

```
<CORPES id="">
  <cabecera fecha_electrónica="">
    <título_principal autor_título_principal="">...</título_principal>
    <título_secundario autor_título_secundario="">...</título_secundario>
    <edición lugar_de_publicación="" editorial="" fecha_de_publicación=""/>
    <numpal n=""/>
    <criterio_clasificación_CORPES criterio="" año=""/>
    <clasificación_textual medio="" bloque="" tema="" soporte="" país="" zona="" origen="" tipología=""/>
    <codificación equipo_codificación="" persona_codificación="" fecha_codificación=""/>
    <validación valor_validación="" persona_validación="" fecha_validación=""/>
    <revisión_RAE valor_revisión_RAE="" persona_revisión_RAE="" fecha_revisión_RAE=""/>
    <notas>...</notas>
  </cabecera>
  <texto>
    <p>
      <ngr>...</ngr>
      <sub>...</sub>
      <csv>...</csv>
      <csvngr>...</csvngr>
      <vrs>...</vrs>
      <rsi>...</rsi>
      <sic>...</sic>
      <nrp/>
    </p>
  </texto>
</CORPES>
```

Índice

Con el patrocinio de

Santander