

# Proyecto de Investigación Programmes: Procesos de Gramaticalización en la Historia del Español

## GUÍA DE ETIQUETADO DOCUMENTOS PROGRAMMES

AUTORA: NAYRA SÁNCHEZ VERA

DIRECCIÓN: DANIEL MOISÉS SÁEZ RIVERA



Septiembre de 2020

## TABLA DE CONTENIDO

<b>GUÍA DE ETIQUETADO TEI – PROYECTO PROGRAMES .....</b>	<b>3</b>
<b>1. INTRODUCCIÓN.....</b>	<b>3</b>
<b>2. CRITERIOS DE TRANSCRIPCIÓN .....</b>	<b>3</b>
<b>3. GUÍA DE ETIQUETADO PARA EL CORPUS PROGRAMES .....</b>	<b>4</b>
<b>4. NORMAS DE ETIQUETADO .....</b>	<b>4</b>
<b>5. ESTRUCTURA DEL DOCUMENTO TEI .....</b>	<b>5</b>
• Declaración de tipo de documento:.....	5
• Encabezado <code>&lt;teiHeader&gt;&lt;/teiHeader&gt;</code> .....	6
• Texto <code>&lt;text&gt;&lt;/text&gt;</code> .....	11
<b>6. ETIQUETADO DEL TEXTO.....</b>	<b>11</b>
<b>6.1. Etiquetas de formato.....</b>	<b>12</b>
<b>6.2. Etiquetas de fenómenos lingüísticos.....</b>	<b>18</b>
<b>6.3. Etiquetas de fenómenos lingüísticos solo empleadas en los textos transcritos entre 2003 y 2005.....</b>	<b>22</b>
<b>6.4. Problemas de anidamiento de etiquetas a tener en cuenta para el etiquetado. ....</b>	<b>23</b>
<b>BIBLIOGRAFÍA.....</b>	<b>25</b>

# **GUÍA DE ETIQUETADO TEI – PROYECTO PROGRAMES**

## **1. INTRODUCCIÓN**

La colección de textos que componen los documentos PROGRAMES<sup>1</sup> se ha ido elaborando y perfeccionando desde los comienzos del proyecto, en 2002. A través de la experiencia adquirida tanto en la transcripción de manuscritos e impresos antiguos como en el posterior etiquetado de los textos en base al esquema TEI, han ido surgiendo diferentes necesidades que finalmente han precisado de una unificación de criterios, tanto en materia de transcripción como de etiquetado.

Estas necesidades han dado lugar finalmente a una serie de criterios básicos de transcripción y a una primera guía de etiquetado, ambos adaptados a los objetivos de estudio de PROGRAMES.

Acerca de la historia, objetivos y posibilidades de investigación de los documentos PROGRAMES, *cf.* Sáez Rivera (2018).

La presente guía de etiquetado se encuentra disponible en la página web del proyecto PROGRAMES<sup>2</sup>.

## **2. CRITERIOS DE TRANSCRIPCIÓN**

Tal y como se detalla en la web de PROGRAMES, “se efectuará básicamente una transcripción paleográfica (grafías, acentuación y puntuación); en cuanto a las abreviaturas, se conservarán las referidas a formas y fórmulas de tratamiento, mientras que se desarrollarán entre corchetes el resto. Se acordó también no separar palabras por renglones, ni siquiera al final de la página, ya que el programa *WordSmith* no reconoce las palabras segmentadas y se trata de una información lingüística poco o nada relevante para nuestro estudio”.<sup>3</sup>

---

<sup>1</sup> <<https://www.ucm.es/procesosdegramaticalizacionenlahistoriadelespanol/documentos-programes>>

<sup>2</sup> Para acceder a la guía de etiquetado véase

<https://www.ucm.es/procesosdegramaticalizacionenlahistoriadelespanol/documentos-programes>

<sup>3</sup> <<https://www.ucm.es/data/cont/docs/821-2018-03-28-TRANSCRIPCI%C3%93N-ETIQUETADO%20PROGRAMES%20corr.pdf>>

### 3. GUÍA DE ETIQUETADO PARA EL CORPUS PROGRAMES

Para la presente guía de etiquetado se ha partido de un modelo previo elaborado en 2018 por una colaboradora del proyecto PROGRAMES, Elisa Caballero<sup>4</sup>. En él se explica la estructura básica de los documentos TEI adaptados al objetivo de estudio de PROGRAMES y aclara una serie de dudas frecuentes a la hora de realizar el etiquetado. En esta nueva guía de etiquetado se ha explicado igualmente la estructura de un documento TEI, aunque se ha detallado de forma más amplia y con ejemplos tanto las etiquetas de formato de texto como las de fenómenos lingüísticos, escogidas siempre en función del objetivo de estudio de la colección de textos del corpus de PROGRAMES. Sobre el etiquetado del corpus, cf. Fernández Martín (2012) y Sáez Rivera (2018).

### 4. NORMAS DE ETIQUETADO

**Estructura de la etiqueta:** Las etiquetas de los lenguajes de marcado tienen una estructura fija que es esencial respetar, ya que un etiquetado erróneo podría dar lugar a que las aplicaciones de análisis de textos no las reconozcan. El formato es el siguiente:

**<nombre de la etiqueta>contenido etiquetado</nombre de la etiqueta>**

Es también muy importante eliminar los espacios dentro de las etiquetas, para evitar posibles errores de reconocimiento. De igual modo, una etiqueta sin cerrar (</etiqueta>) podría tener como consecuencia que todo un documento quedase invalidado e imposible de analizar por los programas habilitados para ello, o que al menos esa etiqueta no sea reconocida por el programa de tratamiento de textos en cuestión, en el caso de PROGRAMES el programa *WordSmith Tools* de Mike Scott<sup>5</sup>.

---

<sup>4</sup> <<https://www.ucm.es/data/cont/docs/821-2018-05-11-Gu%C3%ADa%20para%20el%20etiquetado%20de%20documentos%20PROGRAMES.pdf>>

<sup>5</sup> <<https://www.lexically.net/wordsmith/>>

## 5. ESTRUCTURA DEL DOCUMENTO TEI

La estructura básica de un documento TEI (por Text Encoding Initiative, estándar de etiquetado de texto muy extendido, por no decir dominante, en las Humanidades Digitales<sup>6</sup>) se compone de dos secciones claramente diferenciadas, incluidas ambas dentro del elemento <TEI>. Por un lado, tenemos el encabezado <teiHeader>, donde se integran todos los metadatos descriptivos y, finalmente, el elemento <text> dentro del cuerpo del documento <body> que incluye el texto editado.

- **Declaración de tipo de documento:**

Antes de comenzar con la estructura del documento, tendremos que incluir la declaración SGML:

```
<!-- SGML declaration -->  
<!--Document Type Definition -->
```

SGML (*Standard Generalized Markup Language*) es un estándar de etiquetado de formato utilizado para los diferentes lenguajes de marcado, entre ellos el XML, en el que está basado el esquema TEI. Esta declaración define el tipo de documento.

A continuación, podemos comenzar con el elemento <TEI> que incluye la globalidad del documento. La descripción de este elemento incluye dos atributos, el primero de identificación del texto, así como la lengua principal del mismo.

```
<TEI xml:id=" " xml:lang="es"></TEI>
```

- ✓ **xml:id=""** – En el espacio situado entre comillas se incluye el identificador único del texto.

Los identificadores utilizados en PROGRAMES se crean según un patrón divididos en cinco partes, separadas por puntos:

PROGRAMES	4 PRIMERAS LETRAS DEL	AUTOR	AÑO	TEXTCLASS
-----------	--------------------------	-------	-----	-----------

---

<sup>6</sup> <<https://tei-c.org/>>

	TÍTULO			
--	--------	--	--	--

- ✓ **xml:Lang=""** – En el espacio situado entre comillas incluiremos el código de lengua principal del documento (“es” para español).

Aunque los textos incluidos en los documentos PROGRAMES están escritos en español, este atributo deberá modificarse por la lengua a la que correspondiese en cada caso. La lista de valores del atributo de lengua se puede consultar en el siguiente enlace:

[https://es.wikipedia.org/wiki/ISO\\_639-1](https://es.wikipedia.org/wiki/ISO_639-1)

Ejemplo:

<TEI xml:id="PROGRAMES.RELA.PAEZ.1615.RELACION" xml:lang="es">

- **Encabezado <teiHeader></teiHeader>.**

El encabezado, tal y como se ha mencionado anteriormente, engloba todos los metadatos descriptivos del documento.

El encabezado TEI estándar utilizado para las versiones de análisis de los textos de PROGRAMES es la siguiente:

```
<teiHeader type="text" status="new" date.created="dd/mm/aaaa">
  <fileDesc>
    <titleStmt>
      <title> </title>
      <funder> </funder>
      <respStmt>
        <resp> </resp>
        <name> </name>
      </respStmt>
      <respStmt>
        <resp> </resp>
        <name> </name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <publisher></publisher>
      <pubPlace></pubPlace>
      <distributor></distributor>
      <authority> </authority>
      <availability> </availability>
    </publicationStmt>
  <sourceDesc>
```

```

<biblStruct>
  <monogr>
    <respStmt>
      <resp></resp>
      <name> </name>
      <resp></resp>
      <name></name>
      <resp></resp>
      <name></name>
      <title></title>
      <idno type="origin"></idno>
    </respStmt>
  </monogr>
</biblStruct>
</sourceDesc>
<editionStmt>
  <edition>
    <date>dd/mm/aaaa</date>
  </edition>
</editionStmt>
<extent>
  <measure unit="bytes" quantity="000"></measure>
  <measure unit="words" quantity="0000"></measure>
</extent>
</fileDesc>
<profileDesc>
  <creation>
    <date>aaaa</date>
  </creation>
  <textClass></textClass>
</profileDesc>
<revisionDesc>
  <change> </change>
</revisionDesc>
</teiHeader>

```

En función de la complejidad y necesidades de cada tipo de textos, los encabezados pueden ampliar o simplificar su contenido.

Para los textos que forman parte del corpus de PROGRAMES, los principales elementos incluidos dentro del encabezado son los siguientes:

**- <fileDesc></fileDesc> -** Contiene los datos bibliográficos del texto. Este elemento contiene a su vez las siguientes etiquetas:

- ✓ **<titleStmt></titleStmt>** : Incluiremos en este elemento la información relativa al título de la obra, aunque también a la edición de la misma, como el responsable de la transcripción y codificación y del revisor (etiqueta <resp>).

En cuanto a la etiqueta **<funder>**, PROGRAMES tiene una tabla de códigos propia que asocia el año de etiquetado y transcripción del texto a la fase del proyecto en que se

realiza la misma. Esta tabla también incluye la información que encontraremos en el siguiente elemento, **<publicationStmt>**, para la etiqueta **<authority>**:

<b>AÑOS</b>	<b>PROYECTO (&lt;funder&gt;)</b>	<b>IP PROYECTO (&lt;authority&gt;)</b>
2001 – 2004	PROGRAMES1 REF: BFF2001-1340	José Luis Girón Alconchel
2004 – 2007	PROGRAMES2 REF: HUM04-3610	José Luis Girón Alconchel
2008 – 2011	PROGRAMES3 REF: FFI02828/FILO	José Luis Girón Alconchel
2012 – 2016	PROGRAMES4 REF: FFI2012-31427	José Luis Girón Alconchel
2016 – 2019	PROGRAMES5 REF: FFI2015-64080-P	Francisco Javier Herrero Ruiz de Loizaga

Ejemplo:

```

<titleStmt>
  <title>Relación breve de las fiestas</title>
  <funder>PROGRAMES5 REF: FFI2015-64080-P </funder>
  <respStmt>
    <resp>Transcriptor y codificador</resp>
    <name>Nayra Sánchez Vera</name>
  </respStmt>
  <respStmt>
    <resp>1º revisor</resp>
    <name>Daniel M. Sáez Rivera</name>
  </respStmt>
</titleStmt>

```

- ✓ **<publicationStmt></publicationStmt>** - Contiene información relativa a la publicación y distribución del texto.

Ejemplo:

```

<publicationStmt>
  <publisher>PROGRAMES</publisher>
  <pubPlace>Madrid</pubPlace>
  <distributor>UCM</distributor>
  <authority>Francisco Javier Herrero Ruiz de Loizaga</authority>
  <availability>Creative Commons</availability>
</publicationStmt>

```

- ✓ **<sourceDesc></sourceDesc>** incluye los datos relativos a la fuente de origen del mismo, en este caso el manuscrito o impreso original.

Ejemplo:

```
<sourceDesc>
  <biblStruct>
    <monogr>
      <respStmt>
        <resp>Autor</resp>
        <name>Iuan Paez de Valençuela y Castillejo</name>
        <resp>Impresor</resp>
        <name>Viuda de Andres Barrera</name>
        <resp>Paradero</resp>
        <name>Madrid, BNE, 3/39118</name>
        <title>Relacion [...]</title>
        <idno type="origin">E</idno>
      </respStmt>
    </monogr>
  </biblStruct>
</sourceDesc>
```

- ✓ **<editionStmt></editionStmt>**: Contiene los datos relativos a la fecha de la primera versión de la transcripción y el etiquetado del texto.

Ejemplo:

```
<editionStmt>
  <edition>
    <date>15/10/2019</date>
  </edition>
</editionStmt>
```

- ✓ **<extent></extent>**: Incluye información relativa al tamaño del texto, cuya información está expresada en los atributos de medida “bytes” y “words”.

Ejemplo:

```
<extent>
  <measure unit="bytes" quantity="49925">49925 bytes</measure>
  <measure unit="words" quantity="9089">9089 palabras</measure>
```

</extent>

- <profileDesc></profileDesc> Dentro de esta etiqueta encontramos los elementos contextuales del texto original, como puede ser la fecha de creación, la clase de texto o su temática.

Ejemplo:

```
<profileDesc>
  <creation>
    <date>1615</date>
  </creation>
  <textClass>Relación</textClass>
</profileDesc>
```

En cuanto a las clases de texto (<textClass></textClass>), en PROGRAMES se ha elaborado una propuesta de posibles denominaciones de *textClass* para cada tipo de documento en función de la forma y de la temática, tal y como figura en la tabla a continuación:

<i>Textclass</i>	<b>Ejemplos</b> <b>(Documentos PROGRAMES)</b>
<b>Biografía</b>	<b>Rial, A. Santiago:</b> <i>Compendio de la vida del cardenal Alberoni</i> (post 1752).
<b>Carta</b>	<b>Cotera, Gil de la:</b> <i>Carta del señor Gil de la Cotera, al Sargento Palomino, sobre la Censura Theologica</i> (1699)
<b>Diálogo escolar</b>	<b>Fernández, M.:</b> <i>Conversations ou dialogues en quatre langues, française, espagnole, italienne, anglaise</i> (1838)
<b>Manual de cartas</b>	Mañanares, Gerónimo Paulo de: <i>Estilo y formulario de cartas familiares segun el gobierno de Prelados y Señores temporales</i> (1607)
<b>Narración histórica</b>	<b>Castellví:</b> <i>Narraciones históricas desde el año 1700 hasta el año 1725</i> (1725)
<b>Prosa Jurídico – administrativa</b>	<b>Salazar y Castro, Luis:</b> <i>Papel de Orden del Excelentísimo Señor Duque de Arcos.</i> (1701)
<b>Prosa Periodística</b>	<b>Nifo y Cagigal, Francisco Mariano:</b> <i>Caxon de sastre, o monton de muchas cosas, buenas, mejores, y medianas</i> (1761)
<b>Prosa Técnica</b>	<b>Puertas, Carlos Antonio:</b> <i>Gobierno moral y medico para conservar la salud y buenas costumbres.</i> (1694)
<b>Relación de auto de fe</b>	<b>Ossorio y Basurto, Diego:</b> <i>Relacion verdadera de la grandiosidad con que se celebró el Auto de la Fe en Valladolid [...]</i> (1623)

<b>Relación de sucesos</b>	<i>Breve noticia de los alegres jubilos con que la Ciudad de Matarò solemnizò la Proclamacion de la Magestad del Rey nuestro Señor Don Lvis Primero (1724)</i>
<b>Sermón</b>	<b>Barzia y Zambrana, Joseph:</b> <i>Compendio de los cinco tomos del despertador cristiano (1691)</i>
<b>Tratado</b>	<b>Peña, Felio de la:</b> <i>Político discurso (1681)</i>
<b>Vida de santos</b>	<b>Paez de Valenzuela y Castillejo, Juan:</b> <i>Vida del Siervo de Dios Francisco de Sancta Anna hermano Mayor de los hermitaños del Monte de lAlbayda de la Ciudad de Cordova (1621)</i>

- **<revisionDesc></revisionDesc>** en este elemento se incluirán los datos relativos a la última revisión o cambios en la edición del texto.

Ejemplo:

```
<revisionDesc>
  <change>15/05/2020: Última revisión</change>
</revisionDesc>
```

- **Texto <text></text>**

Elemento que engloba el texto transcrito y etiquetado, incluido a su vez en el elemento **<body></body>**.

Ejemplo:

```
<text>
<body>
<pag[1]><port><p>RELACION BREBE DE LAS FIESTAS, QVE EN LA CIVDAD
DE Cordoua se celebraron à la Beatificacion de la gloriosa Patriarcha Santa Theresa de
IESVS, fundadora de la reformation de Descalços y Descalças Carmelitas. Con la justa
Literaria, que en ella vuo</p><p>< curs>Y SERMON QVE PREDICO EL DOCTOR
Alvaro Piçaño de Palacios Canonigo de Escritura en la Sancta Iglesia de CORDOVA.
</ curs></p><p>POR EL LICENCIADO IVAN PAEZ DE Valençuela Clerigo
Presbytero.</p><p>< curs>DIRIGIDA A. < ft>D.</ft> FRANCISCA DE GVZMAN
< ft>Marquesa del Carpio, señora de las Villas de Adamuz y
Perabad</ft>.</ curs></p><p>[grab. xil : escudo del
impresor]</p></port></pag[1]>[...]
```

## 6. ETIQUETADO DEL TEXTO

El etiquetado de los textos incluidos en los documentos PROGRAMES se ha realizado, como hemos comentado anteriormente, conforme a unos criterios unificados, teniendo

en cuenta los objetivos de investigación del proyecto. Para ello, se han definido una serie de etiquetas clasificadas como etiquetas de formato, aquellas que designan el formato del texto (párrafos, páginas, etcétera) y las etiquetas de fenómenos lingüísticos, que han variado a lo largo de las distintas fases del proyecto, permaneciendo únicamente aquellas que son objeto de investigación estable, y según otras razones que se detallan después.

La clasificación, por tanto, sería la siguiente:

### **6.1. Etiquetas de formato.**

#### ✓ `<body></body>` **Cuerpo del texto**

Va incluida dentro del elemento `<text>` y contiene la totalidad de la transcripción del manuscrito o impreso.

Ejemplo:

```
<text> <body> <folr[I]> <port> <p> LA FEE TRIVNFANTE EN QVATRO AVTOS  
CELEBRADOS EN MALLORCA POR EL SANTO OFICIO DE LA INQVISICION  
EN QUE AN SALIDO OCHENTA, I OCHO REOS, I DE TREINTA, I SIETE  
RELAIADOS SOLO VVO TRES PERTINACES.</p> <p> <curs>EXPRESSADA  
POR EL <ft>R.</ft> <ft>P.</ft>FRANCISCO GARAV de la Compañia de Iesus,  
Calificador del <ft>Santo</ft>Oficio, Examinador Sinodal de este Obispado,  
Excatedratico de Prima de Teologia en el Colegio de Barcelona, i oi Retor del Colegio  
de Montesion de la misma Compañia.</curs> <p> <p>I LA DEDICA A LOS <ft>MVI  
ILLVSTRES MAGnificos Señores</ft>Iurados de la Ciudad, i Reino de Mallorca.</p>  
<p>Con licencia en Mallorca en la Emprinta de la Viuda Guasp. Año 1691.</p>  
</port> </folr[I]> [...] </body></text>
```

#### ✓ `<col#>, </col#>` **Columna**

Señala que en el manuscrito o impreso original el texto está dispuesto en columnas. Se enumerará cada una sustituyendo el símbolo [#] por el número de columna que corresponda, numeración cuyo crecimiento indica disposición de izquierda a derecha, siendo por ejemplo `<col1>` la primera columna, más a la izquierda y `<col2>` la segunda columna, a la derecha de la primera.

Ejemplo:

```
[...]<col1>Don Iuan de Guzman.</col1><col2>Diego Martinez Escriuano Publico.</col2>[...]
```

✓ `<colf>`, `</colf>` **Colofón**

Usaremos esta etiqueta para señalar la anotación final del libro que indica alguna información relativa a la impresión del texto o su transcripción.

Ejemplo:

`<colf><p>` CON LICENCIA: En Madrid, en la Imprenta de `<ft>`D. `</ft>` Gabriel Ramirez, Calle de Atocha. `</p></colf>`

✓ `<curs>`, `</curs>` **Cursiva**

Señala aquellas partes del manuscrito impresas en cursiva.

Ejemplo:

[...] `<curs>`Lo que hay de mas y de ménos en España, para que sea lo que debe ser, y no lo que es; `</curs>`[...]

En cuanto a la foliación, dependiendo del tipo de texto marcaremos los folios (si solo hay número en el recto de la hoja) o las páginas (si hay numeración en el recto y el vuelto de la página) con las siguientes etiquetas, donde “fol” corresponde a “folio”, “r” a “recto”, “v” a “vuelto” y “pag” a “página”:

- ✓ `<folr#>`, `</folr#>` **Folio recto**
- ✓ `<folv#>`, `</folv#>` **Folio vuelto**
- ✓ `<pag#>`, `</pag#>` **Página**

Ejemplos:

- [...] `<folv224>` no debo, ni puedo resistirme `<reg>` a todo lo que se considere remedio eficaz de los daños `</reg>`. `</p>` `<p>` Deseando el Consejo, como s[iem]pre cumplir el R[ea]l precep[t]o de `<ft>` V. M. `</ft>` encuentra `<cd>` tantas dificultades en satisfacer `<ci>` a el `</ci>` plenam[en]te `</cd>` q[ue] se vee

obligado **<reg>** a confesar por imposible el asunto **</reg>**, pues aunque no lo sea dar en abstracto penas correspondientes, a los delitos, como se practica en la promulgacion de las leyes, como estas llevan insitas ensimismas **<cd>** las disposic[i]ones de ôtras leyes **</cd>**, **<rel>** que **</rel>** **<cd>** las **</cd>** declaran, limitan, ò moderan, según las personas, tiempos, hedades, sexos, calidades, y otras, innumerables circunstancias, **<rel>** que **</rel>** en el càso concreto para hacer Juicio caval se consideran no ay peligro en la **</folv224>**[...]

- [...] **<pag[12]>****<p>**y ordenadamente dispuestas luminarias, y globos de fuego, que coronauan las paredes del nueuo edificio, y los texados y campanario del viejo, que oy se habita. Poblose tambien dellas vn lienço de la muralla, que mira al Conuento; y vna torre del se adereço con algunas luminarias artificiosas de papel de diferentes colores, que lucian por su eminencia, y hermosura. Diose principio à la fiesta desde la hora de Visperas, con el repique de campanas, atabales, y trompetas de la Ciudad, y los ministriles de la Iglesia mayor, que respondiendose vnos á otros de diferentes sitios, y juntandose á vezes, hiziero[n] la tarde muy entretenida y de recreacion: haziendoles à ratos salua muchos tiros, que se dispararon. Media hora antes de anochecer se encendieron los fuegos, y auiuò la musica siguie[n]dolos toda la Ciudad, **<marc>**con que**</marc>** [...] **</p>****</pag[12]>** [...]

Se sustituirá siempre el símbolo [#] por el número de folio o página que corresponda. El número se sitúa al final del resto de la etiqueta para que se pueda producir su reconocimiento por *WordSmith*, por eso “r” y “v” aparecen antes de # y después de “fol”, en lugar de lo acostumbrado en una transcripción, en la que el orden habitual sería “fol. 1r”, por ejemplo.

Para hojas sin foliar, se introduce número romano entre corchetes.

- |                                                        |                                                       |
|--------------------------------------------------------|-------------------------------------------------------|
| ✓ <b>&lt;folr[I]&gt;</b> , <b>&lt;/folr[I]&gt;</b>     | <b>Folio I no numerado de preliminares (recto)</b>    |
| ✓ <b>&lt;folr[II]&gt;</b> , <b>&lt;/folr[II]&gt;</b>   | <b>Folio II no numerado de preliminares (recto)</b>   |
| ✓ <b>&lt;folr[III]&gt;</b> , <b>&lt;/folr[III]&gt;</b> | <b>Folio III no numerado de preliminares (recto)</b>  |
| ✓ <b>&lt;folr[IV]&gt;</b> , <b>&lt;/folr[IV]&gt;</b>   | <b>Folio IV no numerado de preliminares (recto)</b>   |
| ✓ <b>&lt;folv[I]&gt;</b> , <b>&lt;/folv[I]&gt;</b>     | <b>Folio I no numerado de preliminares (vuelto)</b>   |
| ✓ <b>&lt;folv[II]&gt;</b> , <b>&lt;/folv[II]&gt;</b>   | <b>Folio II no numerado de preliminares (vuelto)</b>  |
| ✓ <b>&lt;folv[III]&gt;</b> , <b>&lt;/folv[III]&gt;</b> | <b>Folio III no numerado de preliminares (vuelto)</b> |
| ✓ <b>&lt;folv[IV]&gt;</b> , <b>&lt;/folv[IV]&gt;</b>   | <b>Folio IV no numerado de preliminares (vuelto)</b>  |

Ejemplos:

[...]**<folv[VII]>** **<p>**siempre las han de bolver al puesto de donde salen: **< curs>**  
**<foreign>**Ibant, et revertebantur.**</foreign>** **</ curs>**Ezech. 1. 13. I paraque nada falte á  
la idea, tampoco faltó allí la mano de vn ombre que tomava la pluma, ó la regia: **< curs>**  
**<foreign>**Et manus hominis sub pennis ejus.**</foreign>** **</ curs>**ibi. Bien sé lo mucho  
que va de lo vivo á lo pintado, i mas quando está tan muerto el pinzel, como mi pluma;  
pero avn por esso pido á **<ft>**V. S.**</ft>**aquella Palma que siempre en sus Escudos, i  
Armas invencible se levanta; paraque assi deva la Feé á este Catolicissimo Reino, como  
dentro sus muros los triunfos, fuera de sus confines, la Palma. [...]**</p>** **</folv[VII]>**  
[...]

Se podrán numerar tantos folios como fuese necesario según el texto.

✓ **<foreign>**, **</foreign>** **Fragmento o palabra en lengua extranjera**

Se usa esta etiqueta para enmarcar aquellos fragmentos escritos en un idioma distinto del español, siendo los más habituales en los documentos PROGRAMES el latín y el catalán. Si se diese el caso de etiquetar un texto cuyo idioma principal no fuese el español, esta etiqueta se utilizaría para marcar fragmentos de texto distintos de esa lengua.

Ejemplo:

[...]**<foreign>**dispens. & praecept**</foreign>**[...]

✓ **<lim>**, **</lim>** **Liminares**

El uso de esta etiqueta está destinado para el marcado de índices, etcétera, al final del texto.

[...]**<lim>** **<pag212>** **<p>**SOBREESCRITOS DE CARTAS A VIRREYES, y Capitanes  
generales.**</p>** **< marg>**Al Virrey de Napoles.**</ marg>** **<p>** **< curs>**Al Duque,  
Marques, o Conde, N. Virrey, y Capitan general del Reyno de Napoles.**</ curs>** **</p>**  
**< marg>**Al de Milan.**</ marg>** **<p>** **< curs>**Al Duque Marques o Conde. N. Governador  
del Estado de Milan, y Capitan general de Italia.**</ curs>** **</p>****</ pag212>** **</ lim>**[...]

✓ **< marg>**, **</ marg>** **Nota al margen**

Se emplea esta etiqueta para marcar las posibles notas al margen en el texto.

Ejemplo:

[...]**< marg>**Mart. lib. 7. epig 7.**</ marg>**[...]

✓ `<ms>`, `</ms>` **Notas manuscritas en un impreso.**

Se usará esta etiqueta para marcar cualquier nota manuscrita a un texto impreso.

Ejemplo:

[...] `<ms>`Deste Consuo. de S. Diego de Odon `</ms>`[...]

✓ `<p>`, `</p>` **Párrafo.**

Es una de las etiquetas más comunes, con la que se marcan todos los párrafos diferenciados en el manuscrito. Cabe destacar la importancia de esta etiqueta, ya que en los documentos en formato txt con los que trabaja *WordSmith Tools* se eliminan los saltos de línea.

Ejemplo:

[...] `<p>`Platica de dos amigos que se topan en la calle el vno llamado Pablo, y el otro Garcia, y va[n] los dos en casa d[e] vn mercader a comprar paño. `</p>`[...]

✓ `<pie>`, `</pie>` **Nota al pie de página.**

Con esta etiqueta se marcarán las posibles notas al pie de página del texto, normalmente impresos.

Ejemplo:

[...] `<pie>`(1) Los certificados se entregarán dentro de la oficina de Correos al empleado correspondiente `</pie>`[...]

✓ `<portada>`, `</portada>` **Portada.**

Dentro de esta etiqueta se incluirá todo el contenido de la portada del texto, habitualmente en el caso de impresos de cierta longitud (no tanto breves como las relaciones de sucesos incluidas en los documentos PROGRAMES).

Ejemplo:

`<port>` `<p>` ESTILO Y FORMULARIO DE CARTAS FAMILIARES segun el gobierno de Prelados y Señores temporales `</p>`. `<p>`DO SE PONEN OTRAS CARTAS con sus respuestas, y algunas de oficios de Republica `</p>`. `<p>`DIRIGIDO

AL ILUSTRMO SEÑOR DON Bernardo de Rojas y Sandoval, Arçobispo de Toledo, Cardenal de la Sa[n]ta Iglesia de Roma, del Consejo de Estado del rey N. S.</p><p>POR EL MAESTRO GERONIMO PAVlo de Mançanares, Arzipreste de Uzeda</p>. <p> <ms>Deste Consuo. de S. Diego de Odon</ms> </p>. <p>Año 1607</p> <p> < curs>Con privilegio</ curs>, en Madrid, < curs>Por Alonso Martin</ curs>. A costa de Alonso Perez, Mercader de libros</p> </port>

✓ <prelim>, </prelim>      **Preliminares.**

Incluiremos dentro de esta etiqueta los prólogos, páginas preliminares, etcétera.

Ejemplo:

[...]<prelim> <pag[III]> <p> Prólogo. </p> <p>Se han publicado hasta el presente varios estilos de cartas, pero se puede asegurar que en ninguno de ellos han llenado sus autores el objeto que se propusieron: todos son tan anticuados, que aun las gentes menos cultas no se atreverian á seguir su correspondencia por su estilo poco usado entre las personas bien educadas.</p> <p>Estas consideraciones nos han movido á dar á la prensa El nuevo secretario español, arreglado al estilo moderno, que contiene la correspondencia para todos los casos que suelen ocurrir en la sociedad; su estilo es claro, sencillo y noble, y en las advertencias que se</p> </pag[III]>[...].</p> </pag[IV]> </prelim>[...]

✓ <sic>, </sic>      **Errores.**

Etiquetaremos solo los errores o erratas relevantes del texto.

Ejemplos:

- [...]<sic>estima-icion</sic>[...]
- [...]<sic> ciudad a nos </sic>[...]

✓ <tach>, </tach>      **Texto tachado.**

Identificaremos con esta etiqueta las partes tachadas del manuscrito.

Ejemplo:

[...]<tach>Serdeña</tach> <superp>Cerdeña </superp>, Mallorca, Sicilia, Principado de Cathaluña [h tachada], y sus Confines; que han sido el centro dela sangrienta Guerra que empezó en este Siglo. <tach>[...]

✓ `<superp>`, `</superp>`      **Texto superpuesto.**

Utilizaremos esta etiqueta para marcar, si las hubiese, las partes del texto superpuestas, normalmente manuscritas.

Ejemplo:

[...] `<superp>`Sucedidos.`</superp>``<p>``<p>`En los Reÿnos de `<superp>`Castilla`</superp>`[...]

## **6.2. Etiquetas de fen3menos lingüísticos.**

✓ `<cdp>`, `</cdp>`      **Complemento directo preposicional.**

Son objeto de estudio los complementos directos construidos con la preposición “a”. Identificaremos los que forzosamente exigen la preposición. En el *Diccionario Panhispánico de Dudas* (RAE 2005) se explica de manera clara, accesible y rápida (aunque completa) el funcionamiento de este tipo de complementos directos (ver <https://www.rae.es/dpd/a>)

Ejemplo:

[...] `<p>`Brio; con este eximiò `<cdp>`à Betulia`</cdp>`de la opresion fatal de vn Olofernes; [...]

✓ `<conda>`, `</conda>`      **Condicional analítico.**

✓ `<futa>`. `</futa>`      **Futuro analítico.**

Tal y como explica María Jesús Torrens (Torrens, 2007: 106), “Las formas de futuro y condicional del español proceden de las perífrasis de obligación constituidas por infinitivo + presente del verbo *haber* en el caso del futuro (*comer é*), y por infinitivo + imperfecto de haber para el condicional (*comer ía*)”. Aunque estas formas estaban presentes en textos de la Edad Media y permitían intercalar pronombres entre el verbo y el auxiliar, aunque con el tiempo fueron sintetizándose (Torrens, 2007: 106).

Actualmente no se etiquetan estas formas en el marcado TEI de los textos del corpus de PROGRAMES, simplemente porque ya no aparecen en los textos de los siglos XVII y XVIII en los que se centra el proyecto, pero sí encontramos algunos ejemplos en los textos etiquetados en fases anteriores del proyecto.

Ejemplos de etiquetado de condicional analítico:

- [...]<conda>holgarmeia</conda>[...]
- [...]<conda>darleia</conda>[...]

Ejemplos de etiquetado de futuro analítico:

- [...]<futa>quedar se ha</futa>[...]
- [...]<futa>informar se ha</futa>[...]
- 

✓ <dcl>, </dcl>                      **Duplicación clítica.**

Se usará esta etiqueta para marcar las duplicaciones clíticas, producidas cuando “los pronombres átonos aparecen dentro de la misma oración junto con el complemento tónico al que se refieren”. El funcionamiento completo de este fenómeno lingüístico se puede consultar en el *Diccionario Panhispánico de Dudas* (RAE, 2005) (ver <https://www.rae.es/dpd/pronombres%20personales%20C3%A1tonos>).

Ejemplo:

[...]<dcl>le dio passo al respetoso buelo de la pluma</dcl>[...]

✓ <ft>, </ft>                      **Forma/fórmula de tratamiento.**

Para el estudio del corpus, y teniendo en cuenta la naturaleza general de los textos, las fórmulas de tratamiento etiquetadas son formas nominales para apelar a desconocidos, títulos profesionales o sociales, títulos honoríficos o tratamientos que denotan militancia política o de índole religiosa (Carricaburro, 1997:59). Se incluyen en la etiqueta <ft> también los pronombres de tratamiento como “vos”.

Ejemplos:

- [...]<ft>Siluio</ft>[...]
- [...]<ft>su Majestad</ft>[...]
- [...]<p>POR quanto por parte de <ft>vos</ft>[...]

✓ <marc>, </marc>                      **Marcador del discurso**

La definición del Centro Virtual Cervantes para los marcadores de discurso es la siguiente: “Los marcadores del discurso son unidades lingüísticas invariables cuya función es señalar («marcar») la relación que se establece entre dos segmentos textuales. Estas unidades no ejercen función sintáctica alguna, sino que constituyen enlaces supraoracionales que facilitan la cohesión textual y la interpretación de los enunciados”. Esta información se encuentra de manera más ampliada y detallada en el artículo completo disponible en el Centro Virtual de Cervantes (ver [https://cvc.cervantes.es/ensenanza/biblioteca\\_ele/diccio\\_ele/diccionario/marcadoresdiscurso.htm](https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/marcadoresdiscurso.htm)).

Únicamente se etiquetarán los siguientes marcadores de discurso: *con que/conque, a saber, a pesar de (que), no obstante, de hecho, últimamente* (en el sentido de ‘por último’).

Ejemplos:

- [...]<cd> el valor, y el entendimiento de cada uno </cd> </reg> <marc> con que </marc> se hace preciso en la ley natural perdonar <cdp>[...]
- [...]<futa>Hareisselo, señor, notificar</futa>publicamente en Audie[n]cia, estando juntos, de suerte que lo entiendan todos, y executareis luego nuestra sentencia, y lo en ella proveydo, <marc>no obstante</marc>qualquiera apelacion que se interponga por ellos</p>[...]

✓ <lasm>, </lasm>                      **Laísmo**

Uno de los fenómenos lingüísticos más relevantes objeto de investigación de PROGRAMES es el laísmo. Este consiste en el “uso impropio de *la(s)* en función de complemento indirecto femenino, en lugar de *le(s)*, que es la forma a la que corresponde etimológicamente ejercer esa función”. Toda la información detallada sobre este

fenómeno lingüístico está disponible en el *Diccionario Panhispánico de Dudas* (RAE, 2005, ver artículo completo <https://www.rae.es/dpd/la%C3%ADsmo>). Se marcarán los laísmos usando esta etiqueta.

Ejemplos:

- [...]Y co[n]siderando esta por una parte, y por otra dessea[n]do <las m>darla</las m>gusto en lo q[ue] nos ha pedido[...]
- [...]<ft>V.m.</ft>esta, he querido escribir esta carta, y <las m>suplicarla</las m>por ella se anime, y espere en nuestro Señor <las m>la</las m>darà salud[...]

✓ <les m>, </les m>                      Leísmo

Se utilizará esta etiqueta para marcar los leísmos, también definidos por el *Diccionario Panhispánico de Dudas* (RAE, 2005) como el “uso impropio de *le(s)* en función de complemento directo, en lugar de *lo* (para el masculino singular o neutro), *los* (para el masculino plural) y *la(s)* (para el femenino), que son las formas a las que corresponde etimológicamente ejercer esa función” (ver artículo completo <https://www.rae.es/dpd/le%C3%ADsmo>).

Ejemplos:

- [...]<cur s>Yo, Señores, he considerado la audacia de <les m>averle</les m>metido el <ft>P.</ft>Sandoval con vnos hombrones como los que de presente componen el Difinitorio de Castilla;[...]
- [...]<cur s>(y mas si el Alcalde me <les m>le traia probado</les m> con instrumento)[...]

✓ `<losm>`, `</losm>`                      **Loísmo**

De igual modo, se usará esta etiqueta para los posibles loísmos que se encontraran en los manuscritos. El loísmo es el “ uso impropio de *lo(s)* en función de complemento indirecto masculino (de persona o de cosa) o neutro (cuando el antecedente es un pronombre neutro o toda una oración), en lugar de *le(s)*, que es la forma a la que corresponde etimológicamente ejercer esa función. Se encuentra igualmente disponible la información detallada sobre loísmo en el artículo del *Diccionario Panhispánico de Dudas* (2005) (ver <https://www.rae.es/dpd/lo%C3%ADsmo>).

[...]pero los Franceses llevan en esto gran ventaja á los Ingleses; pues como se aplican á convertirlos á la Religion christiana con dulzura, `<losm>`los`</losm>`tienen mas afectos; y como no pretenden dominio sobre ellos sino su amistad, subministrándoles todo lo necesario para su caza, pesca y `</pag32>` [...]

### **6.3. Etiquetas de fenómenos lingüísticos solo empleadas en los textos transcritos entre 2003 y 2005.**

- ✓ `<cd>`, `</cd>`                      **Complemento directo**  
Ejemplo: [...]`<p>` Para que conozcan `<cd>`vna persona`</cd>`[...]
- ✓ `<ci>`, `</ci>`                      **Complemento indirecto**  
Ejemplo: [...]escruiieren `<ci>`a todo genero de personas`</ci>`[...]
- ✓ `<ppas>`, `</ppas>`                      **Participio de pasado.**  
Ejemplo: [...]`<ppas>`Hecho y recopilado`</ppas>`[...]
- ✓ `<ra>`, `</ra>`                      **Forma verbal en -ra**  
Ejemplo: [...]`<ra>`holgara`</ra>`[...]
- ✓ `<reg>`, `</reg>`                      **Régimen preposicional de los verbos**  
[...] `<reg>`confio de`</reg>` `<rel>`quien`</rel>` `<ft>`V. Señoria Illustrissima`</ft>`[...]

- ✓ `<rel>`, `</rel>`      **Relativo**  
Ejemplo: [...] `<reg>`confio de`</reg>`  
`<rel>`quien`</rel>` `<ft>`V. Señoria  
Illustrissima`</ft>`[...]
- ✓ `<relart>`, `</relart>`      **Artículo + relativo.**  
Ejemplo: [...]se puede encarecer `<cd>`  
`<relart>`lo que`</relart>`[...]
- ✓ `<relcomp>`, `</relcomp>`      **Relativo Compuesto**  
Ejemplo: [...]por las cosas naturales, en  
especial `<relcomp>`por las  
q[ue]`</relcomp>`[...]

Aunque en la fase inicial de proyecto PROGRAMES se procedía a etiquetar los fenómenos lingüísticos detallados en esta lista, finalmente se ha descartado dicho etiquetado.

Etiquetar tanto los complementos tanto directo como indirecto resultaba muy costoso, sobre todo en cuanto a tiempo, ya que exigían una revisión exhaustiva (al ser numerosos los errores de etiquetado) que no podía asumirse ni costearse.

Por otro lado, los participios de pasado pueden localizarse a través de las herramientas de análisis como *WordSmith*, introduciendo los parámetros “\*ado” ó “\*ido”, con el comodín de cadena de caracteres de \*. Este mismo procedimiento se podía repetir con las formas verbales terminadas en *-ra*,

En cuanto al régimen preposicional, el relativo compuesto y el artículo + relativo, su etiquetado resultaba igualmente muy laborioso y presentaba numerosos errores que también exigían una minuciosa revisión posterior.

#### **6.4. Problemas de anidamiento de etiquetas a tener en cuenta para el etiquetado.**

A partir del trabajo de etiquetado realizado en los documentos incluidos en PROGRAMES, se han observado varias incidencias en cuanto al anidamiento de etiquetas que han llevado a unificar algunos de criterios de etiquetado como solución, con la finalidad de preparar el corpus para su análisis por programas distintos a *WordSmith* (que al reconocer etiquetas *ad hoc* todas al mismo nivel jerárquico no tiene problemas de ese tipo)..

En cuanto a los párrafos (<p></p>), conviene finalizarlos antinaturalmente al final de cada página, aunque dicho párrafo no finalice de manera real en el texto. *Wordsmith* recoge el final del párrafo sin problema, al tener todas las etiquetas el mismo rango, pero puede no ocurrir lo mismo con otros programas de análisis.

Ejemplo:

[...]de los hombres de suerte, que aun</p></pag[10]><pag[11]><p>sombras de emulacion, [...]

Otro posible caso de anidamiento podría darse al tener lugar varios fenómenos lingüísticos en una misma frase o fragmento del texto, de modo que unos queden marcados en el interior de otra etiqueta. Esto se debe a que para etiquetar un fenómeno lingüístico se debe marcar una cantidad mínima de texto que permita identificar claramente cada caso, pero este fragmento frecuentemente puede incluir a su vez otros fenómenos lingüísticos que deben etiquetarse igualmente, aunque necesariamente tendrán que quedar marcados dentro de esta otra etiqueta. Un caso frecuente es el de la duplicación clítica (<dcl></dcl>), que puede incluir a su vez leísmos (<lesm></lesm>), laísmos (<lasm></lasm>) o complementos directos preposicionales (<cdp></cdp>), así como otros fenómenos como fórmulas de tratamiento (<ft></ft>).

Ejemplos:

- [...]<dcl>A <ft>Don</ft> Ruy Gonzalez pesó le de esto</dcl>,[...]
- [...]<dcl> <lesm>le an de sepultar</lesm>en el Mar airado, <cdp>al complice del delito.</cdp> </dcl>[...]
- [...]<cdp>a <ft>su Magestad</ft>[...]

Para minimizar estos posibles conflictos a la hora de etiquetar, se empieza por la etiqueta más englobadora, que es <dcl>, duplicación clítica de objeto, construcción que incluye un complemento, un pronombre átono y un verbo, y luego los componentes de esa construcción, que puede ser un <cdp> (aunque la mayoría de duplicaciones en la lengua actual y en el corpus es de CI), y además también incurrir en leísmo <lesm> o laísmo <lasm> (verbo + pronombre átono).

Del mismo modo, las fórmulas de tratamiento <ft> se situán dentro del ámbito de las etiquetas correspondientes a la duplicación clítica <dcl> o complemento directo preposicional <cdp>.

Hay que prestar especial atención en incluir el texto mínimo necesario para que se distinga claramente de qué fenómeno lingüístico se trata. Como ya hemos comentado en este apartado, *Wordsmith* no tiene rangos de etiquetas, por lo que no registra ningún conflicto en este aspecto. Sin embargo, podría ocurrir en otros programas utilizado para el análisis de textos.

## BIBLIOGRAFÍA

Carricaburro, Norma (1997): *Las fórmulas de tratamiento en el español actual*. Madrid: Arco Libros, S.L.

Consortium, The TEI. *The TEI Guidelines*. p. 1958. - <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (Accedido 9 de julio de 2020).

Fernández Martín, Patricia (2012): “El modo de trabajo: Criterios de transcripción y etiquetado”, *Filología y lingüística, métodos, corpus y nuevas tecnologías – Propuestas de adaptación de las humanidades a las nuevas formas de hacer ciencia* (pp. 13 – 23). Saarbrücken: Editorial Académica Española.

PROGRAMES – *Transcripción de Etiquetado*.

<https://www.ucm.es/data/cont/docs/821-2018-03-28-TRANSCRIPCI%C3%93N-ETIQUETADO%20PROGRAMES%20corr.pdf>

(Accedido 9 de julio de 2020).

PROGRAMES - *Guía para el etiquetado de documentos PROGRAMES* -

<https://www.ucm.es/data/cont/docs/821-2018-05-11-Gu%C3%ADa%20para%20el%20etiquetado%20de%20documentos%20PROGRAMES.pdf> (Accedido 9 de julio de 2020).

Sáez Rivera, Daniel M. (2018): “Algunas posibilidades de investigación del proyecto *Procesos de gramaticalización en la historia del*

*español* (Documentos PROGRAMES)”, en María Dolores Romero / Manuel Salamanca (eds.), *Entornos digitales: Humanidades y Ciencias sociales en la Universidad Complutense de Madrid* (pp. 141-157). México D. F.: Red de Humanidades Digitales de México.

Torrens Álvarez, M.<sup>a</sup> Jesús (2007): “Morfosintaxis (II). El verbo y los elementos de relación”, *Evolución e historia de la lengua española* (pp 101-119). Madrid: Arco Libros, S.L.

UCM-Proyecto de Investigación Programes: Procesos de Gramaticalización en la Historia del Español.

<https://www.ucm.es/procesosdegramaticalizacionenlahistoriadelespanol/documentos-programes>. (Accedido 9 de julio de 2020).