

Corpus OSTA-app

Manual de usuario

Francisco Gago-Jover

F. Javier Pueyo Mena



Versión 1.0

agosto, 2024

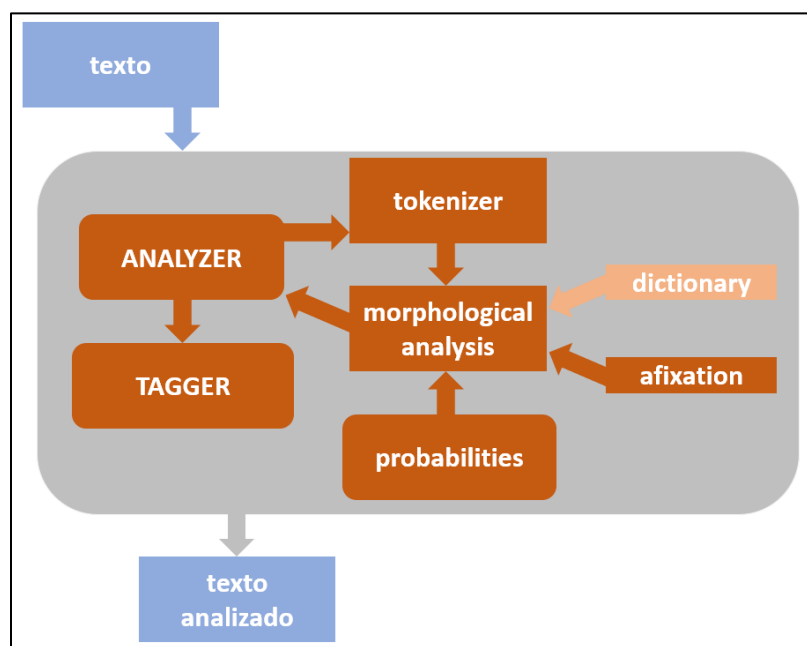
Tabla de contenidos

1. ¿QUÉ ES <i>Corpus OSTA-app</i> ?	4
2. INSTALACIÓN Y REQUISITOS	5
3. DESCRIPCIÓN DE LA INTERFAZ	6
4. TRABAJANDO CON <i>Corpus OSTA-app</i>	6
4.1 Corrección de las transcripciones paleográficas	6
4.2 Lematizado y etiquetado de textos	12
4.3 Edición del diccionario de formas del lematizador	14
a. forma simple	17
b. forma con otros análisis posibles	18
c. formas aglutinadas	18
d. forma verbo+clítico(s)	19
e. forma verbo+clítico(s)+auxiliar	19
f. forma adverbial en <i>mente</i>	20
g. forma adjetivo superlativo	20
5. ESTRUCTURA DEL DICCIONARIO	20
6. PROBLEMAS CONOCIDOS	21
 Apéndice A: Formato de los ficheros	 25
Apéndice B: Etiquetas EAGLES	32

1. ¿QUÉ ES *Corpus OSTA-app*?

Corpus OSTA-app es una herramienta de análisis textual desarrollada por F. Javier Pueyo Mena para el procesamiento de las transcripciones paleográficas utilizadas en la elaboración del *Old Spanish Textual Archive* (osta.oldspanishtextualarchive.org/). *Corpus OSTA-app* integra las librerías de FreeLing y lo combina con una serie de scripts para que, a partir del texto plano de las transcripciones paleográficas del HSMS, sea posible obtener un texto en formato XML con toda la información lingüística incorporada, manteniendo también todas las características estructurales de la obra para facilitar su lectura y su posterior presentación en los resultados de las consultas.

FreeLing 3.1,¹ una herramienta de Procesamiento del Lenguaje Natural diseñada para el análisis multilingüe, se sitúa en el centro de *Corpus OSTA-app*. FreeLing encadena diferentes procesos y programas (tokenizer, splitter, analizador morfológico, anotador lingüístico, desambiguador, etc.) y devuelve textos anotados lingüísticamente mediante el uso de una serie de recursos léxicos o diccionarios (de léxico común, de topónimos y antropónimos, de abreviaturas, de expresiones locutivas, etc.), de reglas (sean reglas de segmentación textual o sean reglas lingüísticas para la identificación de clíticos, análisis de morfemas derivativos, reconocimiento de terminaciones verbales, etc.) y de un modelo probabilístico, creado a partir de un corpus de entrenamiento corregido manualmente, que se aplica durante las etapas de desambiguación de homónimos y de análisis de las formas desconocidas.



¹ Carreras, Xavier, Isaac Chao, Lluís Padró y Muntsa Padró (2004): «FreeLing: An Open-Source Suite of Language Analyzers», *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*. [<http://nlp.lsi.upc.edu/publications/papers/carreras04.pdf>]

Padró, Lluís y Evgeny Stanilovsky (2012): «FreeLing 3.0: Towards Wider Multilinguality», *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)* ELRA. Istanbul, Turkey. May. [\[http://nlp.lsi.upc.edu/publications/papers/padro12.pdf\]](http://nlp.lsi.upc.edu/publications/papers/padro12.pdf)

Padró, Lluís (2011): «Analizadores Multilingües en FreeLing», *Linguamatica*, 3.2, pp. 13-20.
[<http://nlp.lsi.upc.edu/publications/papers/padro11.pdf>]

Una primera adaptación de FreeLing al español medieval fue realizada por Sánchez Marco, Boleda y Padró² (2011) quienes utilizaron parte de las transcripciones semi-paleográficas del HSMS para conformar su *golden corpus*, entrenar el programa, y crear y modificar los recursos y las reglas en los ámbitos arriba señalados. Por nuestra parte, hemos mejorado y ampliado considerablemente los recursos lingüísticos de FreeLing, en particular el diccionario (`hsms.scr`), que ahora solo contiene formas medievales (cerca de 260.000), y las reglas de afijación (`afixos.dat`), cuya sección de análisis de clíticos ha sido expandida.

2. INSTALACIÓN Y REQUISITOS

Corpus OSTA-app fue desarrollado para el tratamiento de las transcripciones paleográficas que forman la base del *Old Spanish Textual Archive*. Estas transcripciones siguen el método descrito en el manual de transcripción del Hispanic Seminary of Medieval Studies.³ Aunque también es posible procesar textos en formato plano, sin ningún tipo de codificación.

Corpus OSTA-app es un programa para Windows 10 o superior.⁴ Para instalarlo hay que extraer los contenidos del fichero `HSMS-app.zip` en la carpeta que vayamos a utilizar para nuestro trabajo. Los textos deben estar situados en su propia carpeta, dentro o fuera de la carpeta en la que hayamos instalado *Corpus OSTA-app*. Para este manual asumimos la siguiente estructura:

`C:\HSMS-tools` [carpeta general]

`C:\HSMS-tools\analizador` [carpeta con *Corpus OSTA-app*]

`C:\HSMS-tools\textos` [carpeta con los textos que queremos procesar; en esta misma carpeta aparecen los textos procesados en formato HTML]⁵

Para el correcto funcionamiento de *Corpus OSTA-app* los ficheros de texto que van a ser analizados deben estar codificados en formato UTF-8 y tener la siguiente estructura en su nombre: `TEXT.XXX.txt`, donde XXX corresponde al nombre del fichero o código del texto: `TEXT.PFG.txt`.


Para iniciar *Corpus OSTA-app* basta con hacer clic en `corpusOSTA-app`, dentro de `C:\HSMS-tools\analizador`.⁶

² Sánchez-Marco, Cristina, Gemma Boleda, y Lluís Padró (2011): «Extending the tool, or how to annotate historical language varieties», *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 1-9, Portland, OR, USA, 24 June 2011. [<http://nlp.lsi.upc.edu/papers/sanchezmarco11.pdf>]

³ Mackenzie, David (1997). *A Manual of Manuscript Transcription for the Dictionary of the Old Spanish Language*. Fifth Edition Revised and Expanded by Ray Harris-Northall. Madison: Hispanic Seminary of Medieval Studies. [www.hispanicseminary.org/docs/HSMS-manual.pdf]

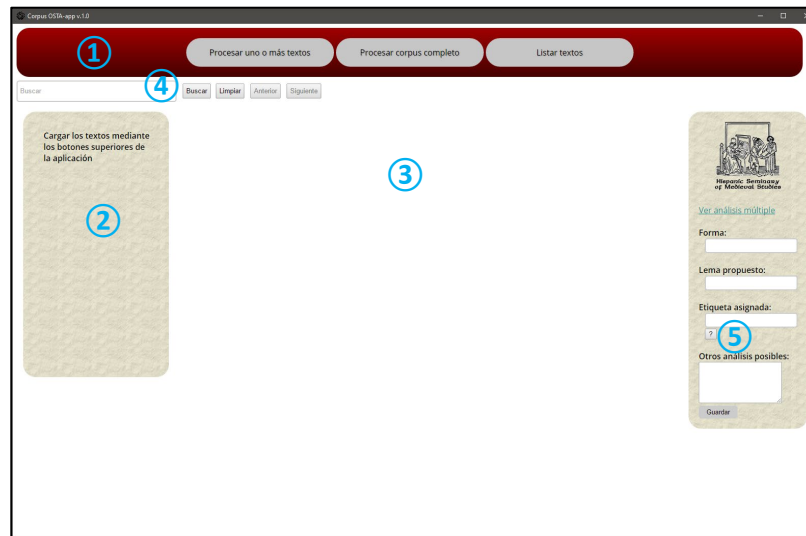
⁴ Es posible ejecutar *Corpus OSTA-app* en un ordenador Apple. Para ello debemos instalar una máquina virtual para poder ejecutar Windows 10 u 11. Los ficheros con los que vamos a trabajar pueden colocarse en una carpeta compartida, accessible para los dos sistemas operativos.

⁵ Los textos pueden colocarse en cualquier carpeta del disco duro e incluso en la nube (se ha comprobado el correcto funcionamiento de *Corpus OSTA-app* en Dropbox y Google Drive).

⁶ Si este método no funciona, debemos abrir la terminal de comandos  + `cmd`, y luego escribir `"C:\HSMS-tools\analizador\node-webkit\nw.exe" "C:\HSMS-tools\analizador\hsms-app"`

3. DESCRIPCIÓN DE LA INTERFAZ

La interfaz de trabajo de *Corpus OSTA-app* consta de los siguientes elementos: **barra de selección de procesos** ①, **barra de textos** ②, **ventana de texto** ③, **caja de búsquedas** ④ y **analizador** ⑤.



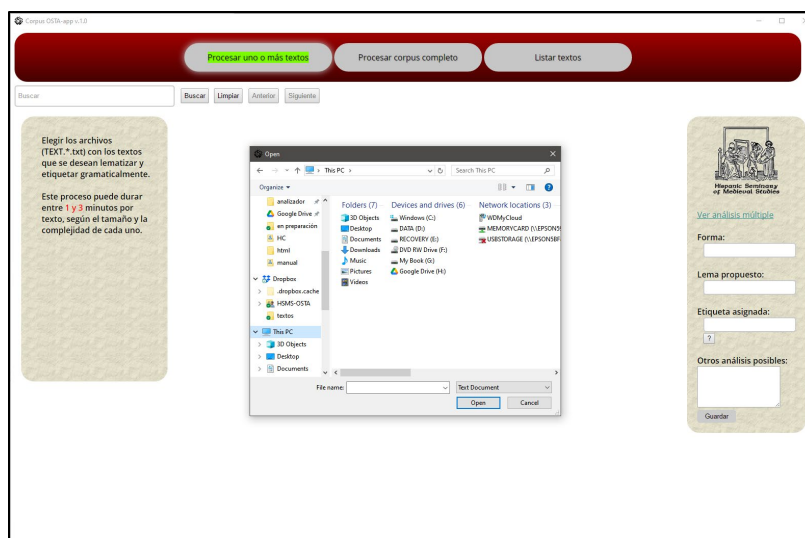
4. TRABAJANDO CON *Corpus OSTA-app*

Corpus OSTA-app es una herramienta diseñada para trabajar en diferentes fases del proceso de elaboración del *Old Spanish Textual Archive*: 1) corrección de las transcripciones paleográficas, 2) lematizado y etiquetado morfológico de los textos, y 3) edición del diccionario de formas utilizado por el lematizador.

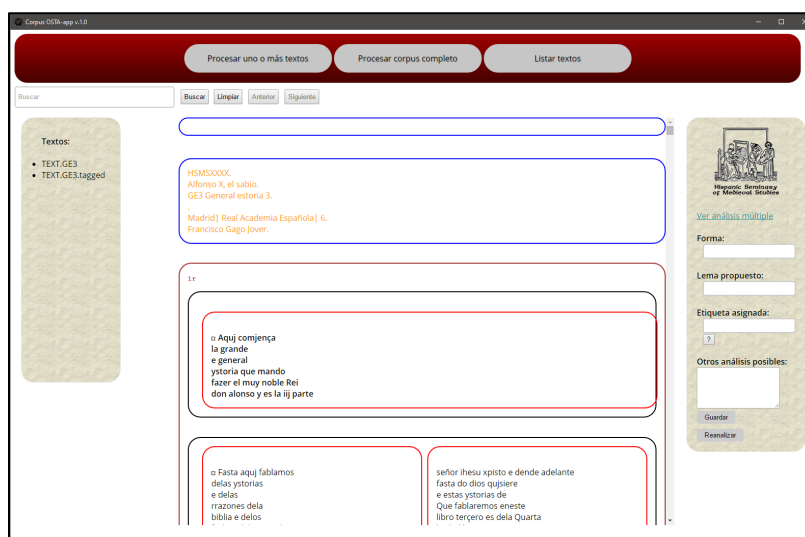
4.1 Corrección de las transcripciones paleográficas

Una vez hecha la transcripción paleográfica siguiendo las normas del HSMS, podemos utilizar *Corpus OSTA-app* para su corrección.

Para ello hacemos clic en **Procesar uno o más textos** en la **barra de selección de procesos**, y seleccionamos en la ventana que aparece en nuestra pantalla la transcripción que vamos a corregir.



Una vez seleccionado el texto comienza su procesamiento, que puede durar entre 1 y 3 minutos, según su tamaño y complejidad, permaneciendo la **ventana de texto** en blanco. El texto se muestra en ella una vez procesado.



El texto que aparece en la **ventana de texto** resuelve abreviaturas, añadidos y supresiones, mantiene la disposición del texto en el código (foliación, columnas, cabeceras, signatures y reclamos), e incluye el etiquetado morfológico (visible al pasar el cursor por encima de una palabra). Las palabras marcadas en **azul** corresponden a aquellas palabras identificadas como formas verbales con clítico. Las palabras marcadas en **rojo** corresponden a aquellas que no han podido ser lematizadas por no encontrarse en el diccionario de formas del lematizador, pudiéndose tratar de palabras transcritas correctamente, pero aún no lematizadas y analizadas, o de errores de transcripción.

% dy non temes los bramjdos dela entrada temerosa njn dela selua espantosa los sus canes & ladridos [^b] temer se deuen las cosas q<ue> han poder de nozir o mal fazer otras non son pauorosas [^f] % ya las terresçio theseo & dubdolas el alc'ides duques espertos en lides (^dizes quando proserpjna) (^fue Roba) & temjolas periteo [^b] dizes q<u><<a>>(ua)[n]do proserpjna fue Robada non gozo desa vegada la congregac'io<n> maligna [^f] % delos dioses c'elestiales las estigias son temjdas non temes las eumenjdas nj<n> los mostruos jnfernales njn los ojos jnflamados de caron [^b] non njn toda la Region do se penan los culpados	¶ dy non temes los bramjdos dela entrada temerosa njn dela selua espantosa los sus canes & ladridos b temer se deuen las cosas que han poder de nozir o mal fazer otras non son pauorosas f ¶ ya las terresçio theseo & dubdolas el alc'ides duques espertos en lides dizes quando proserpjna fue Roba & temjolas periteo b dizes quando proserpjna fue Robada non gozo desa vegada la congregaçion maligna f ¶ delos dioses çelestiales las estigias son temjdas non temes las eumenjdas njn los mostruos jnfernales njn los ojos jnflamados de caron b non njn toda la Region do se penan los culpados
<i>texto original</i>	<i>texto procesado</i>

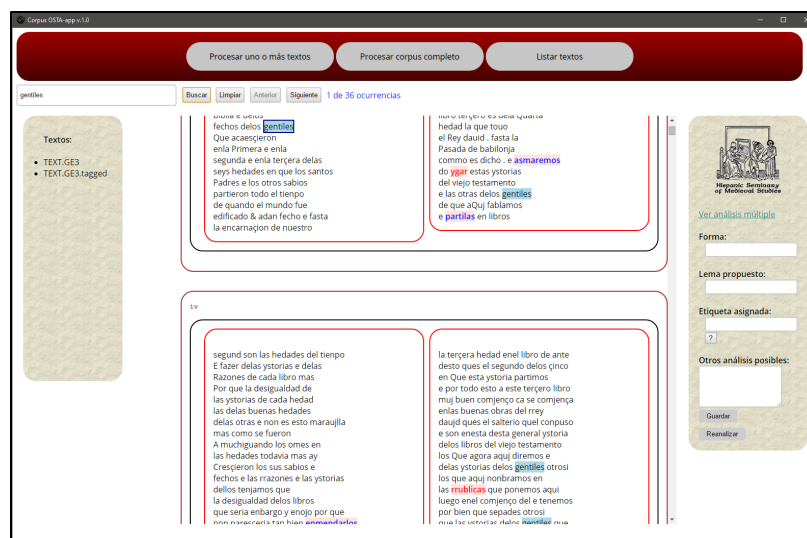
La corrección de la transcripción paleográfica no se realiza directamente desde *Corpus OSTA-app*, sino a través de un editor de texto plano como BBEEdit⁷ o Notepad++⁸, en el que trabajaremos directamente con el fichero que contiene la transcripción (TEXT.xxx.txt).

Para navegar por el texto podemos utilizar la barra de desplazamiento vertical o la **caja de búsquedas**, que permite ir a un folio concreto introduciendo el número en la ventana, p. ej.: 12r.

La **caja de búsquedas** también permite hacer consultas dentro de la transcripción paleográfica. Para ello introducimos el término [palabra completa, fragmento o secuencia de caracteres] en la caja de consultas y luego hacemos clic en **buscar**. Los resultados de la búsqueda aparecen resaltados en azul en la **ventana de texto**. Podemos desplazarnos por los resultados utilizando los botones **anterior** o **siguiente**, cambiando el resultado actual a azul. El botón **limpiar** borra los resultados en la **caja de búsquedas** y la **ventana de texto**.

⁷ <https://www.barebones.com/products/bbedit/>

⁸ <https://notepad-plus-plus.org/>

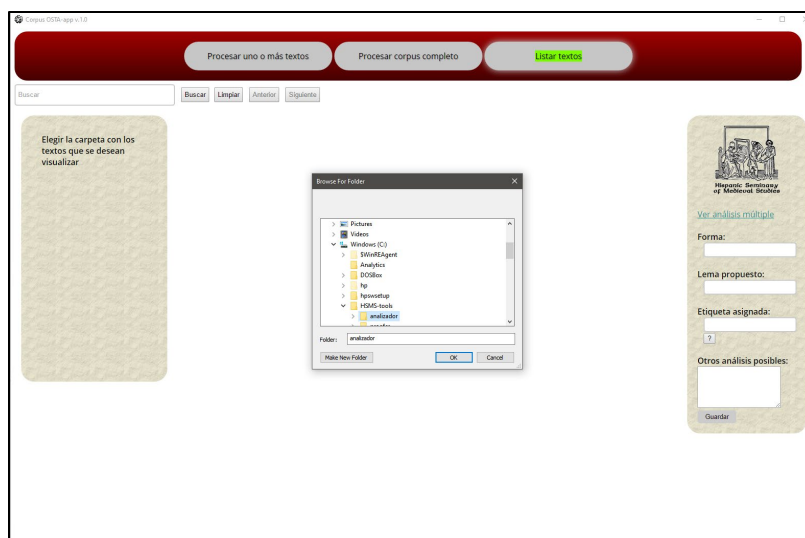


La **caja de búsquedas** permite utilizar las siguientes expresiones regulares:

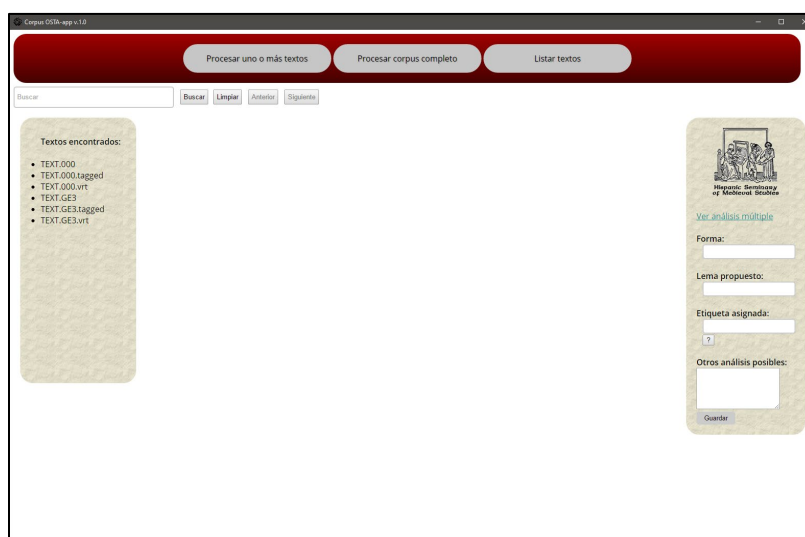
- equivale a un único carácter: `n.n`, para encontrar *nin*, *njn* y *nyn*.
- ? el carácter que le precede puede aparecer cero veces o una vez: `arciprestes?`, para encontrar *arcipreste* y *arciprestes*.
- * el carácter que le precede puede aparecer cero o más veces: `cas*as`, para encontrar *casas* y *cassas*.
- + el carácter que le precede puede aparecer una o más veces: `ab+ades`, para encontrar *abades* y *abbades*.
- [] cualquiera de los caracteres dentro de los corchetes puede aparecer como mucho una vez: `ar[cç][ei]`, para encontrar *arcediano*, *arcedianos*, *arçediano*, *arçedianos*, *arcipreste*, *arciprestes*, *arçipreste* y *arçiprestes*.
- \b para marcar el comienzo o final de una palabra: `adgo\b`, para encontrar todas las palabras que terminan en *adgo*.

De esta forma es posible saber si cierta combinación de caracteres o la forma que estamos corrigiendo aparece más veces en el texto.

Es posible seleccionar una transcripción ya procesada para comenzar o continuar el proceso de corrección. Para ello debemos hacer clic en **Listar textos**, y seleccionar en la ventana *abades* la carpeta que contiene los textos procesados con los que queremos trabajar.



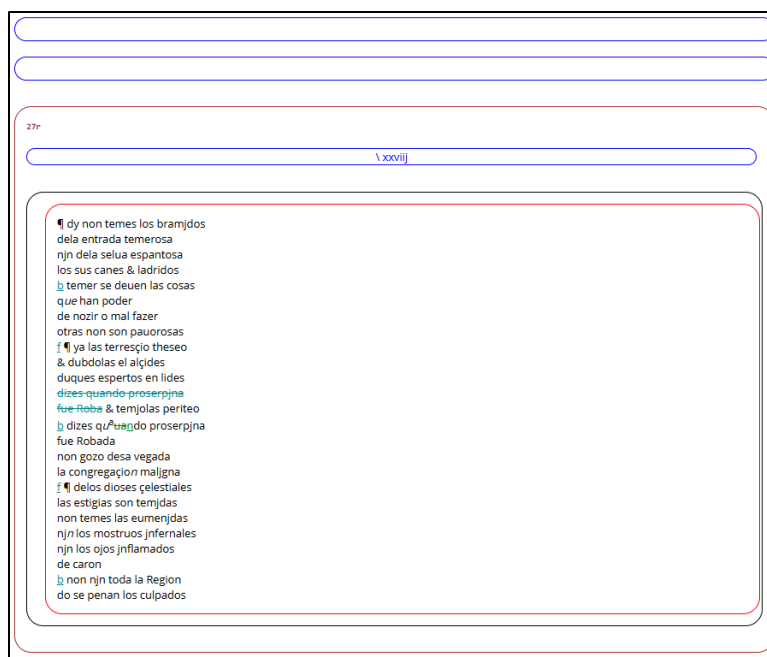
mostrándose en la **barra de textos** el nombre de todos los ficheros disponibles.



Cada uno de los textos está disponible en tres formatos diferentes⁹:

- **TEXT.XXX**: se resuelven abreviaturas, añadidos y supresiones; se mantiene la disposición del texto en el código (foliación, columnas, cabeceras, signaturas y reclamos); el texto no ofrece ningún tipo de etiquetado morfológico. Este fichero, con extensión HTML, se encuentra en **C:\HSMS-tools\textos**, existe también una versión en formato XML.

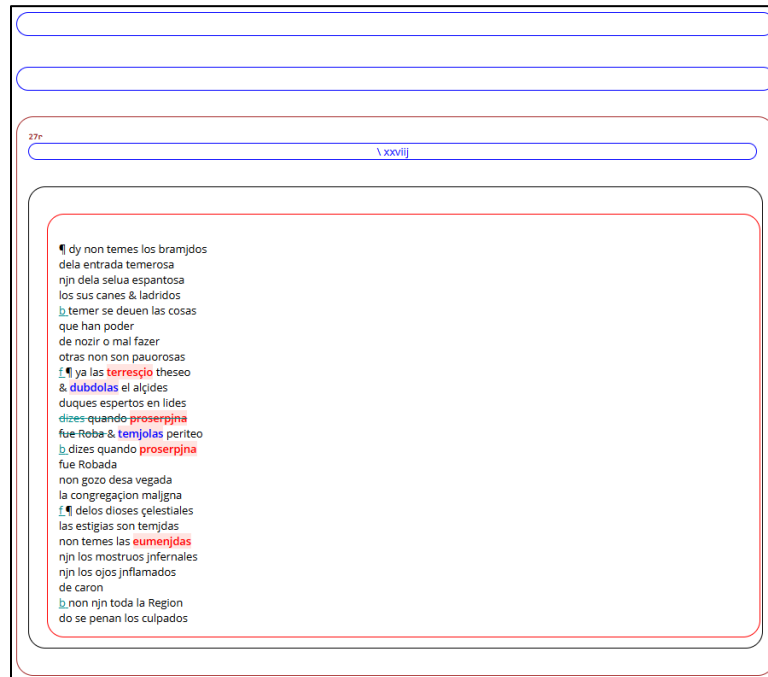
⁹ El **Apéndice A** ofrece muestras del texto original y de los tres diferentes formatos.



- **TEXT.XXX.vrt**: se resuelven abreviaturas, añadidos y supresiones; se mantiene la disposición del texto en el código (foliación, columnas, cabeceras, signatures y reclamos); el texto incluye el etiquetado morfológico (visible al pasar el cursor por encima de una palabra) y se añade el número de línea.



- **TEXT.XXX.tagged**: se resuelven abreviaturas, añadidos y supresiones; se mantiene la disposición del texto en el código (foliación, columnas, cabeceras, signatures y reclamos); el texto incluye el etiquetado morfológico (visible al pasar el cursor por encima de una palabra); las palabras marcadas en **rojo** corresponden a aquellas que no han podido ser lematizadas; las palabras marcadas en **azul** corresponden a las formas verbales con clítico; en algunas ocasiones el análisis no es correcto pues no existe la entrada correspondiente en el diccionario, por lo que es siempre conveniente revisar estas formas.¹⁰ Este fichero, con extensión HTML, se encuentra en **C:\HSMS-tools\textos**, existe también una versión en formato XML.



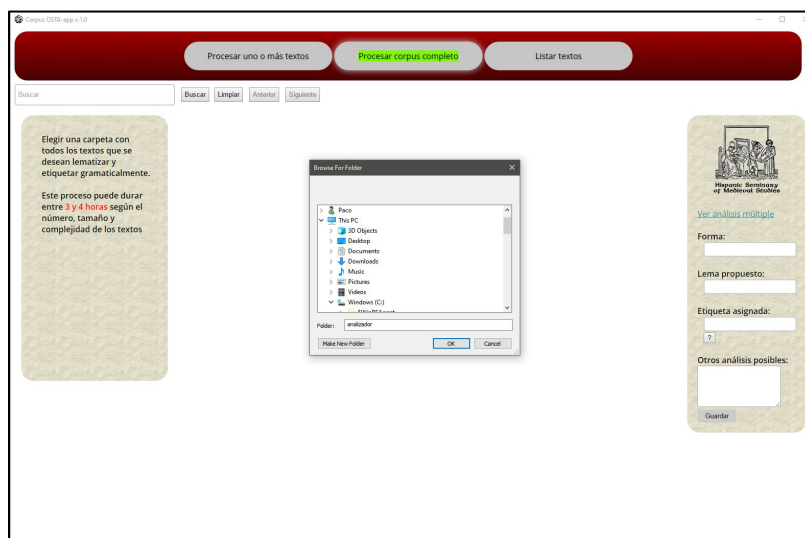
Para trabajar en la corrección de la transcripción paleográfica siempre debemos seleccionar este último formato (**TEXT.XXX.tagged**).

4.2 Lematizado y etiquetado de textos

Otras de las funcionalidades de *Corpus OSTA-app* es la lematización y etiquetado morfológico de las transcripciones paleográficas. Una vez corregidas podemos procesarlas de forma individual haciendo clic en **Procesar uno o más textos**, o bien procesar todo un corpus de texto, para ello hacemos clic **Procesar un corpus completo**, donde podemos seleccionar la carpeta que contiene el corpus de textos que

¹⁰ Supongamos que la palabra **ljdiase**, que debería aparecer etiquetada como imperfecto de subjuntivo, **ljdiase lidiar VMSI3S0**, aparece analizada en *Corpus OSTA-app* como forma verbal+clítico, **ljdiase lidiar+se VMIP3S0+PC3CN000**). Esto se debe a que en el diccionario no aparece la entrada **ljdiase lidiar VMSI3S0**, pero sí la entrada **ljdia lidia NCFS000 lidio NCFS000 Lidia NP000G0 lidio AQ0FS0 lidiar VMIP3S0 lidiar VMM02S0**, y el módulo de afijación de FreeLing analiza **ljdiase** a partir de la información que le proporciona el diccionario.

queremos procesar. Una vez seleccionado el corpus de texto comienza el procesado, que puede durar entre 1 y 4 horas, según el número, tamaño y complejidad de los textos.



Los textos lematizados y etiquetados morfológicamente (`TEXT.XXX.vrt.html`) aparecen en la misma carpeta donde se encuentran los textos originales. Los ficheros `TEXT.XXX.vrt.html` son ficheros en texto plano codificados en formato UTF-8 y con etiquetas XML para cada uno de sus elementos.

```
<FOL>
<folio>27r</folio>
<HD>
<w data1='\'•PUNCT•PUNCT'>\</w>
<w data1='xxviiij•28•Z'>xxviiij</w>
</HD>
<CB1>
<CB>
<LN id='1'><LNC>1</LNC></LN>
<w data1='¶•¶•Fz'>¶</w>
<w data1='dy•dar•VMIS1S0'>dy</w>
<w data1='non•no•RN'>non</w>
<w data1='temes•temer•VMIP2S0'>temes</w>
<w data1='los•el•DA0MP0'>los</w>
<w data1='bramjdos•bramido•NCMP000'>bramjdos</w>
<BR>
```

Cada línea del original se identifica con la etiqueta `<LN id='1'><LNC>1</LNC></LN>` y cada una de las palabras en una línea del original aparece etiquetada con `<w data1='dy•dar•VMIS1S0'>dy</w>`, siendo el símbolo • el utilizado para separar los tres componentes del lematizado y análisis morfológico¹¹:

¹¹ El conjunto de etiquetas morfológicas utilizadas en OSTA se basa en las etiquetas propuestas por el grupo *Expert Advisory Group on Language Engineering Standards* (EAGLES) <http://www.ilc.cnr.it/EAGLES/browse.html>. En el **Apéndice B** se ofrece la lista completa de las etiquetas utilizadas.

forma original • lema • análisis morfológico

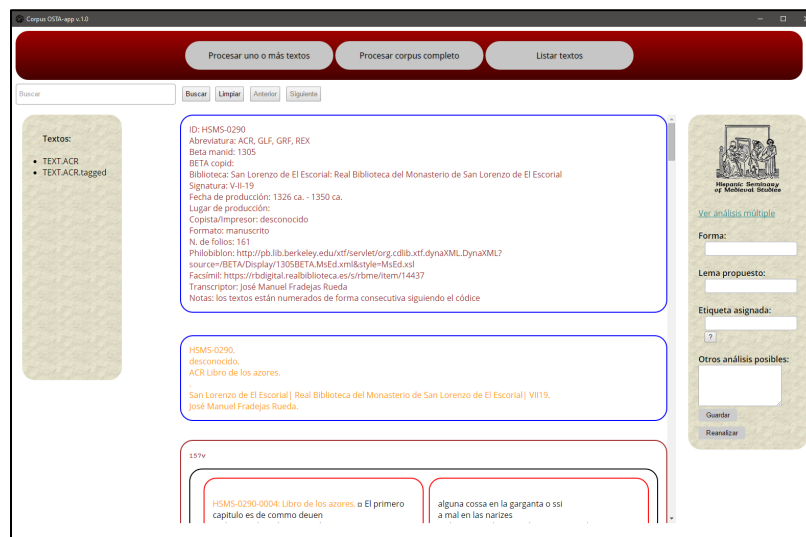
Los ficheros `TEXT.XXX.vrt.html` pueden ser procesados para su exportación a otros formatos.

4.3 Edición del diccionario de formas del lematizador

La última funcionalidad de *Corpus OSTA-app* es la edición del diccionario de formas utilizado por el lematizador. Este diccionario, `hsms.scr`, se encuentra en la carpeta `hsms-es` y es un fichero de texto plano en formato UTF-8 con una entrada por línea:¹²

```
descumunjon excomuni3n NCFS000
descuntamentos descontamiento NCMP000
descuntar descontar VMN0000
descureremos discurrir VMIF1P0
descuridad de+escuridad SP+NC
descuriendo discurrir VMG0000
descurrece discurrecer VMIP3S0 discurrecer VMM02S0
descurrir discurrir VMN0000
descuvjerta descubrir VMP00SF descubierto AQ0FS0 descubierta NCFS000
descuyda descuidar VMIP3S0 descuidar VMM02S0
```

Para seleccionar el texto con el que vamos a trabajar hacemos clic en **listar textos** en la **barra de selecci3n de procesos** y luego seleccionamos en la **barra de textos** el fichero `TEXT.XXX.tagged` correspondiente.



Los dos cuadros que aparecen antes del texto constituyen la cabecera del texto. El cuadro superior muestra los metadatos de los textos ya indexados en el *Old Spanish Textual Archive*, estar3 vaci3 en los

¹² Véase la secci3n 5. **ESTRUCTURA DEL DICCIONARIO**, para una explicaci3n m3s detallada del formato y estructura del diccionario.

{RMK: *código alfanumérico del código.*}¹³

{RMK: *autor.*}

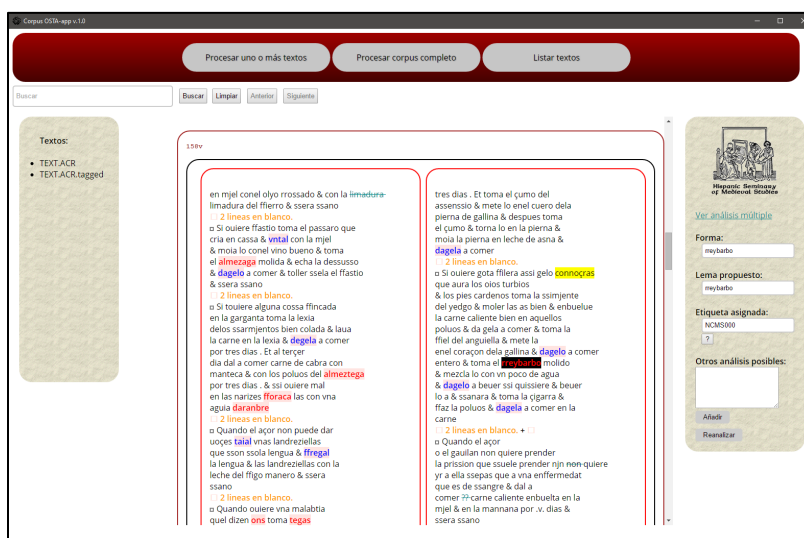
{RMK: [*código alfanumérico de la transcripción*] *título.*}¹⁴

{RMK: *ciudad | impresor | fecha de impresión.*}¹⁵

{RMK: *ciudad | biblioteca | signatura.*}

{RMK: *nombre del transcriptor.*}

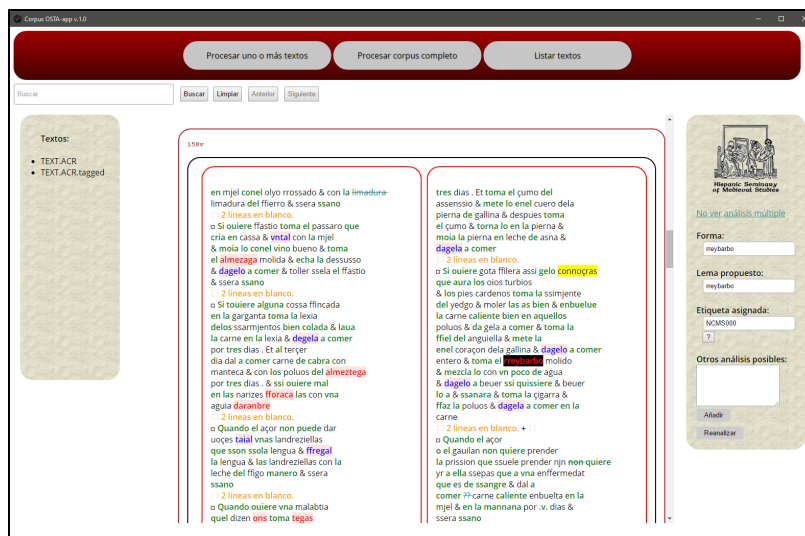
Al hacer clic en cualquier palabra, esta cambia de color: **palabra**, si se trata de una ya incluida en el diccionario, o **palabra**, si no está incluida.



- **Forma:** muestra la palabra según aparece en la transcripción
- **Lema propuesto:** muestra el lema asignado a la forma. En las palabras que aparecen definidas en el diccionario (hsms.src), se asigna el lema correspondiente. En las palabras que **no** aparecen definidas el diccionario, todas aquellas en **rojo**, el lema propuesto suele corresponder con la forma.

¹⁵ Este campo aparece en blanco {RMK:.} en el caso de los manuscritos.

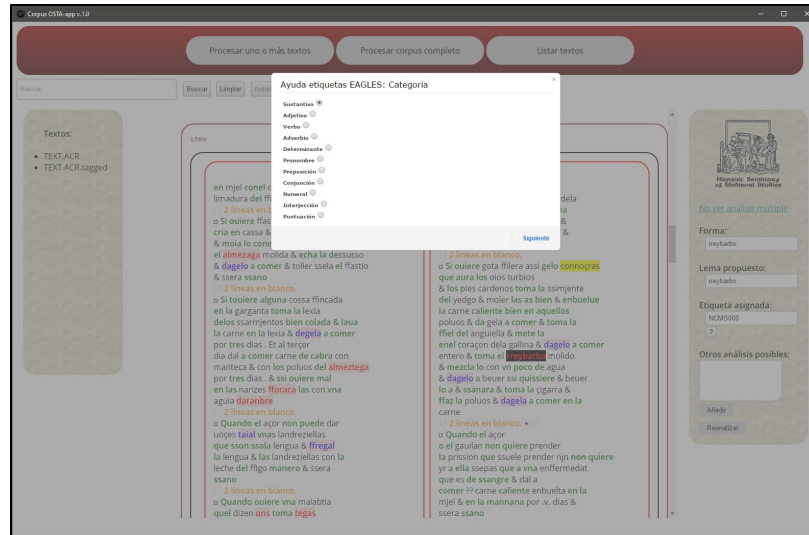
- **Etiqueta asignada:** muestra la etiqueta morfológica¹⁶ asignada por el lematizador.
- **Otros análisis posibles:** muestra los diferentes análisis posibles de cualquier palabra. Para saber qué palabras tienen múltiples análisis podemos hacer clic en `Ver análisis múltiple` para mostrarlas en un **color** diferente



Para editar el diccionario debemos seguir los siguientes pasos:

1. Seleccionamos una de las palabras en **rojo** o en **amarillo**.
2. Editamos manualmente el lema y la etiqueta morfológica asignada. Para este segundo proceso, si no conocemos la etiqueta morfológica correspondiente, podemos utilizar el botón **?** que muestra una ventana de ayuda para el etiquetado donde podemos hacer clic en la categoría correspondiente para seleccionarla e indicar en las siguientes pantallas todos los elementos necesarios para definir la etiqueta. A continuación, hacemos clic en **finalizar** para que la etiqueta aparezca en la caja de **Etiqueta asignada**.

¹⁶ Véase el **Apéndice B**.



3. Añadimos la nueva entrada al diccionario haciendo clic en **Añadir** (en el caso de tratarse de una **palabra** no lematizada o analizada en el diccionario) o **Modificar** (en el caso de tratarse de una **palabra** cuya lematización y análisis queremos editar). En unos segundos aparecerá una ventana confirmando que la entrada ha sido modificada; debemos hacer clic en **OK** para cerrarla.
4. Podemos comprobar el resultado haciendo clic en **Reanalizar**, que procesa el texto con el nuevo diccionario. Las entradas editadas que antes aparecían en **rojo**, aparecen ahora en negro.

A continuación, ofrecemos una serie de ejemplos para mostrar el proceso de edición/modificación de las entradas del diccionario. La columna de la izquierda corresponde a las diferentes cajas de texto del **analizador** y la de la derecha al texto que introducimos en cada una de ellas.

a. forma simple

Forma:	prophanados
Lema propuesto:	profanar
Etiqueta asignada:	VMP00PM
Otros análisis posibles:	

Forma:	tituliuiio
Lema propuesto:	Tito_Livio ¹⁷
Etiqueta asignada:	NP000P0
Otros análisis posibles:	

¹⁷ El guion bajo, **_**, sirve para separar los dos elementos que componen el lema.

b. forma con otros análisis posibles

Forma:	amare
Lema propuesto:	amar
Etiqueta asignada:	VMSF1S0
Otros análisis posibles:	amar@VMSF3S0 ¹⁸ amar@VMIF1S0

Forma:	noziente
Lema propuesto:	nuciente
Etiqueta asignada:	AQ0CS0
Otros análisis posibles:	nocir@VMPP0SC

Forma:	coxos
Lema propuesto:	cojo
Etiqueta asignada:	AQ0MP0
Otros análisis posibles:	cojo@NCMP000

c. formas aglutinadas: Las formas aglutinadas deben ser definidas mediante sus componentes básicos, tal y como aparecen en el texto.

Forma:	quadavno
Lema propuesto	quada+vno
Etiqueta asignada:	DI+PI
Otros análisis posibles:	

Forma:	larçebispe
Lema propuesto	l+arçebispe
Etiqueta asignada:	DA+NC
Otros análisis posibles:	

Forma:	dindia
Lema propuesto	d+india
Etiqueta asignada:	SP+NP
Otros análisis posibles:	

Forma:	sesfuerça
Lema propuesto	s+esfuerça
Etiqueta asignada:	PP+VM
Otros análisis posibles:	

De forma alternativa, es posible utilizar la correspondiente forma moderna, no apocopada, de artículos, pronombres y preposiciones:

¹⁸ La arroba, @, sirve para separar el lema y la etiqueta morfológica.

Forma:	larçebispe
Lema propuesto	el+arçebispe
Etiqueta asignada:	DA+NC
Otros análisis posibles:	

Forma:	dindia
Lema propuesto	de+india
Etiqueta asignada:	SP+NP
Otros análisis posibles:	

Forma:	sesfuerça
Lema propuesto	se+esfuerça
Etiqueta asignada:	PP+VM
Otros análisis posibles:	

En ambos análisis es fundamental comprobar que cada uno de los componentes de la forma aglutinada aparece correctamente definido en el diccionario `hsms.src`, de lo contrario el lematizador genera un error al procesar los textos. Así, en la forma *larçebispe*, habría que confirmar la existencia de las entradas

1 1 NCFS000 50 Z el DA0MS0 (que define el artículo)
arçebispe arzobispo NCMS000 (que define el sustantivo)

Si este no fuera el caso, habría que añadirlas al diccionario.

- d. **forma verbo+clítico(s):** Este tipo de entradas requiere un tratamiento diferente. En el caso de `vngendolo`, debemos simplemente definir la forma verbal `vngendo`,

Forma:	vngendo
Lema propuesto	ungir
Etiqueta asignada:	VMG0000
Otros análisis posibles:	

haciendo posible que el módulo de afijación del lematizador se encargue de definir todas las combinaciones de `vngendo`, con clíticos: `vngendote`, `vngendoles`, etc.

- e. **forma verbo+clítico(s)+auxiliar:** Este tipo de entradas no son procesadas por el módulo de afijación y por lo tanto requieren que se defina cada uno de sus componentes.

Forma:	tomarleas
Lema propuesto	tomar+le+as
Etiqueta asignada:	VM+PP+VA

Otros análisis posibles:	
---------------------------------	--

Forma:	cubrirseloeys
Lema propuesto	cubrir+se+lo+eys
Etiqueta asignada:	VM+PP+PP+VA
Otros análisis posibles:	

- f. **forma adverbial en *mente*:** Este tipo de entradas requiere un tratamiento diferente, según la forma básica del adjetivo. En el caso de *afamadamente*, donde la forma adjetival se corresponde con la forma moderna del lema (en femenino) debemos simplemente definir la forma *afamada*,

Forma:	afamada
Lema propuesto	afamado
Etiqueta asignada:	AQ0FS0
Otros análisis posibles:	afamar@VMP00SF

para que el módulo de afijación del lematizador se encargue de etiquetarla automáticamente.

En el caso de *senyalladament*, donde la forma adjetival no se corresponde con la forma moderna del lema (en femenino), debemos simplemente definir la forma de la siguiente manera

Forma:	senyalladament
Lema propuesto	señaladamente
Etiqueta asignada:	RM
Otros análisis posibles:	

- g. **forma adjetivo superlativo:** En el caso de *suauissimas* debemos definir la forma de la siguiente manera:

Forma:	suauissimas
Lema propuesto	suaue
Etiqueta asignada:	AQ0FP0
Otros análisis posibles:	

5. ESTRUCTURA DEL DICCIONARIO

El diccionario que FreeLing utiliza para el procesamiento de los textos, *hsms.src*, es un fichero en formato UTF-8 localizado en la carpeta *C:\HSMS-tools\analizador\hsms-es*.

Cada entrada ocupa una única línea y tiene la siguiente estructura: *forma lema1 PoS1*

```
contygo contigo PP2CS000
guirlanda guirnalda NCFS000
reuuscassen rebuscar VMSI3P0
thiriaca triaca NCFS000
```

Sin embargo, algunas entradas contienen más de un lema y análisis morfológico ya que se trata de formas homógrafas:

```
casa casa NCFS000 casar VMIP3S0 casar VMM02S0
esto estar VAIP1S0 este PD0NS000
furto hurtar VMIP1S0 hurtar VMIS3S0 hurto NCMS000
manleuase manlevar VMSI1S0 manlevar VMSI3S0
vinno venir VMIS3S0 vino NCMS000
```

Las entradas correspondientes a formas aglutinadas, cuyos componentes van a ser analizados individualmente, tienen un formato similar: `forma forma1+forma2 PoS1+PoS2`, donde `forma1`, `forma2` son las formas (no los lemas) de las palabras aglutinadas. Por ejemplo:

```
desmeralda d+esmeralda SP+NC
```

Esta línea expresa que siempre que se encuentra la forma `desmeralda`, esta es reemplazada por sus componentes: `d` y `esmeralda`, asignándose a cada uno de ellos la etiqueta, `PoS1` y `PoS2` definida en `esmeralda` su entrada correspondiente.

```
d d NCFS000 de SPS00 500 Z
esmeralda esmeralda NCFS000
```

De este modo, `d` será analizado como `SP` y `esmeralda` como `NCFS000`.

Un número reducido de entradas de formas aglutinadas tienen la siguiente estructura: `forma forma1+forma2 PoS1+PoS2/PoS2`

```
avna a+una SP+PI/DI
poral pora+el SP+DA/PP
```

En estos casos la barra `/` separa dos posibles análisis, que FreeLing determina utilizando el módulo de probabilidades.

Cuando editamos el diccionario con *Corpus OSTA-app* las nuevas entradas aparecen entre las etiquetas `<Entries>` `</Entries>` al final del fichero `hsms.src`. Estas nuevas entradas también aparecen en el fichero `additions-hsms.src`, en la carpeta `C:\HSMS-tools\analizador\hsms-es\`.

6. PROBLEMAS CONOCIDOS

a. copias de seguridad

Para poder recuperar una versión de `hsms.src` que funcione correctamente es conveniente trabajar con dos ficheros `master-hsms.src` y `hsms.src`, inicialmente idénticos, pero que tras nuestra sesión de trabajo cambian ya que *Corpus OSTA-app* solo efectúa cambios en `hsms.src`. También es posible guardar una copia de `hsms.src` en otra carpeta.

b. modificación del fichero additions-hsms.src

Si modificamos los contenidos del fichero additions-hsms.src y reanalizamos un texto con *Corpus OSTA-app*, se modifican también las entradas situadas al final de hsms.src. Para evitar esto, podemos pasar las entradas del final de hsms.src en la segunda sección <Entries> </Entries>

```

xarafe jarabe NCMS000
xerez Jerez NP000G0
xibia jibia NCFS000
</Entries>
<Entries>
anuncio anunciar VMIP1S0 anunciar VMIS3S0
cerastes ceraste NCFS000
auctoritates autoridad NCFP000
</Entries>
```

al final de la lista de palabras, en la primera sección <Entries> </Entries>

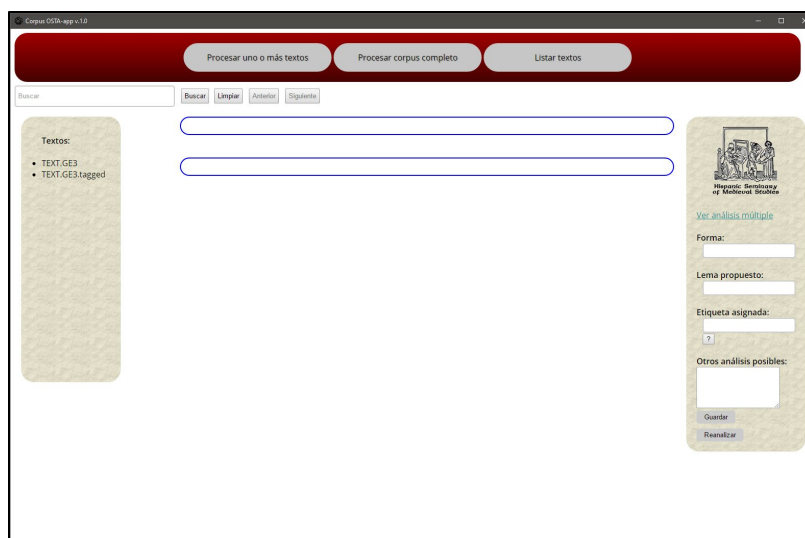
```

xarafe jarabe NCMS000
xerez Jerez NP000G0
xibia jibia NCFS000
anuncio anunciar VMIP1S0 anunciar VMIS3S0
cerastes ceraste NCFS000
auctoritates autoridad NCFP000
</Entries>
<Entries>
</Entries>
```

Pudiendo ahora eliminar los contenidos de additions-hsms.src. Una vez hecho esto, debemos comprobar que no haya entradas reduplicadas en hsms.src. Tras finalizar con este proceso, conviene volver a analizar un texto para comprobar el correcto funcionamiento del diccionario.

c. errores de formateado de las entradas

En algunas ocasiones *Corpus OSTA-app* nos muestra la siguiente pantalla tras reanalizar un texto.



Esto nos indica la existencia de un error en alguna de las entradas del diccionario:

- hay un espacio en blanco extra entre alguno de los componentes de una entrada:

forma lema1 PoS1

- hay un espacio en blanco extra al comienzo o final de una entrada:

forma lema1 PoS1
forma lema1 PoS1

- no hay espacio en blanco entre alguno de los componentes de una entrada:

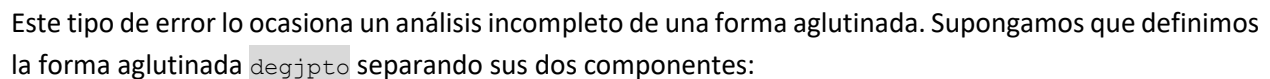
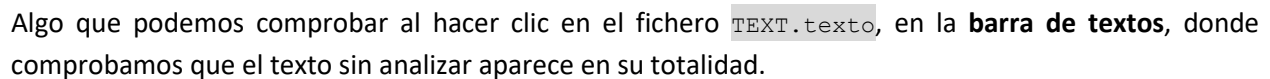
formalema1 PoS1
forma lema1PoS1

- en lugar de un espacio en blanco entre cada de los componentes de una entrada hay un guion bajo `_`, resultado de no haber usado `@` para separar lema y análisis morfológico en la caja **Otros análisis posibles**:

desamava desamar VMII1S0 desamar_VMII3S0

d. errores de definición de las formas aglutinadas

En algunas ocasiones *Corpus OSTA-app* no completa el análisis del texto, quedando la operación interrumpida. En la siguiente imagen se ve cómo la columna B no es analizada en su totalidad.



aunque no hay ningún error en la definición, la forma `egjpto` no está definida en el diccionario, por lo que *Corpus OSTA-app* no puede continuar procesando el texto cuando se encuentra con esta forma definida parcialmente. Para que no se produzca este tipo de error bastaría con añadir al diccionario el siguiente análisis:

Corpus OSTA-app – Manual de usuario - 24

Apéndice A: Formato de los ficheros

TEXT.original.txt

[fol. 27r]
{HD. \ xxvii}
{CB1.
% dy non temes los bramjdos
dela entrada temerosa
njin dela selua espantosa
los sus canes & ladridos
[^b] temer se deuen las cosas
q<ue> han poder
de nozir o mal fazer
otras non son pauorosas
[^f] % ya las terresçio theseo
& dubdolas el alc'ides
duques espertos en lides
(^dizes quando proserpjna)
(^fue Roba) & temjolas periteo
[^b] dizes q<u><<a>>(ua)[n]do proserpjna
fue Robada
non gozo desa vegada
la congregac'io<n> maligna
[^f] % delos dioses c'elestiales
las estigias son temjdas
non temes las eumenjdas
nj<n> los monstruos jnfernales
njin los ojos jnflamados
de caron
[^b] non njn toda la Region
do se penan los culpados}

TEXT.original.html

<DIV class="FOL">27r

<DIV class="HD" title="Heading"> \ xxvii</DIV>

<DIV class="CB1">
<DIV class="CB">
¶ dy non temes los bramjdos

dela entrada temerosa

njin dela selua espantosa

los sus canes & ladridos

b temer se deuen las cosas

que han poder

de nozir o mal fazer

otras non son pauorosas

f ¶ ya las terresçio theseo

& dubdolas el alcides

duques espertos en lides

dizes quando proserpjna

fue Roba & temjolas periteo

b dizes qu^auando proserpjna
 fue Robada
 non gozo desa vegada
 la congregaçion maljgna
 f ¶ delos dioses çelestiales
 las estigias son temjdas
 non temes las eumenjdas
 njn los mostruos jnfernales
 njn los ojos jnflamados
 de caron
 b non njn toda la Region
 do se penan los culpados
 </DIV>
 </DIV>
 </DIV>

TEXT.original.tagged.html

<DIV class="FOL">27r
 <DIV class="HD" title="Heading"> \ xxviiij </DIV>
 <DIV class="CB1">
 <DIV class="CB">
 ¶ dy non temes los bramjdos
 dela entrada temerosa
 njn dela selua espantosa
 los sus canes & ladridos
 b temer se deuen las cosas
 que han poder
 de nozir <span class="word2"

data-title="o: • o • CC" data2-title="o+NCF5000@donde+PR000000@donde+RG">o mal fazer

otras non son pauorosas

 f

¶ ya las terresçio theseo

& <span class="unknown2" data-title="dubdolas: • dudar+lo • VMIS3S0+PC3FPA00" data2-title="<dudar+lo>+VMIS3S0+PP3FPA00@<dudar+lo>+VMIP1S0+PP3FPA00">dubdolas el alçides

duques espertos en lides

 dizes quando <span class="unknown" data-title="proserpjna: • proserpjna • NCF5000" data2-title="<proserpjna>+AQ0FS0@<proserpjna>+NCF5000">proserpjna

 fue Roba & temjolas periteo

 b

dizes quando <span class="unknown" data-title="proserpjna: • proserpjna • AQ0FS0" data2-title="<proserpjna>+AQ0FS0@<proserpjna>+NCF5000">proserpjna

fue Robada

non gozo desa vegada

la congregaçion maljgna

 f

¶ delos dioses celestiales

las estigias son temjdas

non temes <span class="word2" data-title="las: • el • DA0FP0" data2-

title="lo+PP3FPA00@la+NCMP000">las eumenjdas

njin los mostruos jnfernales

njin los ojos jnflamados

de caron

 b

non njin toda la Region

do se penan los culpados

</DIV>

</DIV>

</DIV>

TEXT.original.vrt.html

<FOL>

<folio>27r</folio>

<HD>

<w data1='\•PUNCT•PUNCT'>\</w>

<w data1='xxviiij•28•Z'>xxviiij</w>

</HD>

<CB1>

<CB>

<LN id='1272'><LNC>1</LNC></LN>

<w data1='¶¶•Fz'>¶¶</w>

<w data1='dy•dar•VMIS1S0'>dy</w>

<w data1='non•no•RN'>non</w>

<w data1='temes•temer•VMIP2S0'>temes</w>

<w data1='los•el•DA0MP0'>los</w>

<w data1='bramjdos•bramido•NCMP000'>bramjdos</w>

<LN id='1273'><LNC>2</LNC></LN>

<w data1='dela•de•el•SPS00•DA0FS0'>dela</w>

<w data1='entrada•entrada•NCF5000'>entrada</w>

<w data1='temerosa•temeroso•AQ0FS0'>temerosa</w>

<LN id='1274'><LNC>3</LNC></LN>

<w data1='njin•ni•CC'>njin</w>

<w data1='dela•de•el•SPS00•DA0FS0'>dela</w>

<w data1='selua•selva•NCFS000'>selua</w>
 <w data1='espantosa•espantoso•AQ0FS0'>espantosa</w>

 <LN id='1275'><LNC>4</LNC></LN>
 <w data1='los•el•DA0MP0'>los</w>
 <w data1='sus•su•DP3CP0'>sus</w>
 <w data1='canes•can•NCMP000'>canes</w>
 <w data1='&•y•CC'>&</w>
 <w data1='ladridos•ladrido•NCMP000'>ladridos</w>

 <LN id='1276'><LNC>5</LNC></LN>
 <w data1='b•b•NCFS000'><COPADD>b</COPADD></w>
 <w data1='temer•temer•VMN0000'>temer</w>
 <w data1='se•se•PC3CN000'>se</w>
 <w data1='deuen•deber•VMIP3P0'>deuen</w>
 <w data1='las•el•DA0FP0'>las</w>
 <w data1='cosas•cosa•NCFP000'>cosas</w>

 <LN id='1277'><LNC>6</LNC></LN>
 <w data1='que•que•PROCN000'>q<ABB>ue</ABB></w>
 <w data1='han•haber•VAIP3P0'>han</w>
 <w data1='poder•poder•NCMS000'>poder</w>

 <LN id='1278'><LNC>7</LNC></LN>
 <w data1='de•de•SPS00'>de</w>
 <w data1='nozir•nocir•VMN0000'>nozir</w>
 <w data1='o•o•CC'>o</w>
 <w data1='mal•mal•RG'>mal</w>
 <w data1='fazer•hacer•VMN0000'>fazer</w>

 <LN id='1279'><LNC>8</LNC></LN>
 <w data1='otras•otro•PIOFP000'>otras</w>
 <w data1='non•no•RN'>non</w>
 <w data1='son•ser•VSIP3P0'>son</w>
 <w data1='pauorosas•pavoroso•AQ0FP0'>pauorosas</w>

 <LN id='1280'><LNC>9</LNC></LN>
 <w data1='f•f•NCFS000'><COPADD>f</COPADD></w>
 <w data1='¶•¶•Fz'>¶</w>
 <w data1='ya•ya•CS'>ya</w>
 <w data1='las•lo•PC3FPA00'>las</w>
 <w data1='terresçio•UNK•VMIS3S0'>terresçio</w>
 <w data1='theseo•Teseo•NP000P0'>theseo</w>

 <LN id='1281'><LNC>10</LNC></LN>
 <w data1='&•y•CC'>&</w>
 <w data1='dubdolas•dudar•lo•VMIS3S0•PC3FPA00'>dubdolas</w>
 <w data1='el•el•DA0MS0'>el</w>
 <w data1='alçides•Alcides•NP000P0'>alçides</w>

 <LN id='1282'><LNC>11</LNC></LN>

<w data1='duques•duque•NCMP000'>duques</w>
 <w data1='espertos•experto•AQ0MP0'>espertos</w>
 <w data1='en•en•SPS00'>en</w>
 <w data1='lides•lid•NCMP000'>lides</w>

 <LN id='1283'><LNC>12</LNC></LN>
 <COPDEL>
 <w data1='dizes•decir•VMIP250'>dizes</w>
 <w data1='quando•cuando•CS'>quando</w>
 <w data1='proserpjna•UNK•NCFS000'>proserpjna</w>
 </COPDEL>

 <LN id='1284'><LNC>13</LNC></LN>
 <COPDEL>
 <w data1='fue•ser•VSIS350'>fue</w>
 <w data1='Roba•robar•VMIP350'>Roba</w>
 </COPDEL>
 <w data1='&•y•CC'>&</w>
 <w data1='temjolas•temer-lo•VMIS350•PC3FPA00'>temjolas</w>
 <w data1='periteo•Periteo•NP000P0'>periteo</w>

 <LN id='1285'><LNC>14</LNC></LN>
 <w data1='b•b•NCFS000'><COPADD>b</COPADD></w>
 <w data1='dizes•decir•VMIP250'>dizes</w>
 <w data1='quando•cuando•CS'>q<ABB>u</ABB>^a<EDITDEL>ua</EDITDEL><EDITADD>n</EDITADD>do</w>
 <w data1='proserpjna•UNK•AQ0FS0'>proserpjna</w>

 <LN id='1286'><LNC>15</LNC></LN>
 <w data1='fue•ser•VSIS350'>fue</w>
 <w data1='Robada•robar•VMP00SF'>Robada</w>

 <LN id='1287'><LNC>16</LNC></LN>
 <w data1='non•no•RN'>non</w>
 <w data1='gozo•gozar•VMIP150'>gozo</w>
 <w data1='desa•de-ese•SPS00•DD0FS0'>desa</w>
 <w data1='vegada•vegada•NCFS000'>vegada</w>

 <LN id='1288'><LNC>17</LNC></LN>
 <w data1='la•el•DA0FS0'>la</w>
 <w data1='congregaçion•congregación•NCFS000'>congregaçio<ABB>n</ABB></w>
 <w data1='maligna•maligno•AQ0FS0'>maligna</w>

 <LN id='1289'><LNC>18</LNC></LN>
 <w data1='f•f•NCFS000'><COPADD>f</COPADD></w>
 <w data1='¶•¶•Fz'>¶</w>
 <w data1='delos•de-el•SPS00•DA0MP0'>delos</w>
 <w data1='dioses•dios•NCMP000'>dioses</w>
 <w data1='çelestiales•celestial•AQ0CP0'>çelestiales</w>

 <LN id='1290'><LNC>19</LNC></LN>
 <w data1='las•el•DA0FP0'>las</w>

<w data1='estigias•estigio•AQ0FP0'>estigias</w>
 <w data1='son•ser•VSIP3P0'>son</w>
 <w data1='temjdas•temer•VMP00PF'>temjdas</w>

 <LN id='1291'><LNC>20</LNC></LN>
 <w data1='non•no•RN'>non</w>
 <w data1='temes•temer•VMIP2S0'>temes</w>
 <w data1='las•el•DA0FP0'>las</w>
 <w data1='eumenjdas•UNK•VMP00PF'>eumenjdas</w>

 <LN id='1292'><LNC>21</LNC></LN>
 <w data1='njn•ni•CC'>nj<ABB>n</ABB></w>
 <w data1='los•el•DA0MP0'>los</w>
 <w data1='mostruos•monstruo•NCMP000'>mostruos</w>
 <w data1='jnfernales•infernai•AQ0CP0'>jnfernales</w>

 <LN id='1293'><LNC>22</LNC></LN>
 <w data1='njn•ni•CC'>njn</w>
 <w data1='los•el•DA0MP0'>los</w>
 <w data1='ojos•ojo•NCMP000'>ojos</w>
 <w data1='jnflamados•inflamar•VMP00PM'>jnflamados</w>

 <LN id='1294'><LNC>23</LNC></LN>
 <w data1='de•de•SPS00'>de</w>
 <w data1='caron•Carón•NP000P0'>caron</w>

 <LN id='1295'><LNC>24</LNC></LN>
 <w data1='b•b•NCF5000'><COPADD>b</COPADD></w>
 <w data1='non•no•RN'>non</w>
 <w data1='njn•ni•CC'>njn</w>
 <w data1='toda•todo•DIOFS0'>toda</w>
 <w data1='la•el•DA0FS0'>la</w>
 <w data1='Region•región•NCF5000'>Region</w>

 <LN id='1296'><LNC>25</LNC></LN>
 <w data1='do•donde•PR000000'>do</w>
 <w data1='se•se•PC3CN000'>se</w>
 <w data1='penan•penar•VMIP3P0'>penan</w>
 <w data1='los•el•DA0MP0'>los</w>
 <w data1='culpados•culpado•NCMP000'>culpados</w>
 </CB>
 </CB1>
 </FOL>

Apéndice B: Etiquetas EAGLES

El conjunto de etiquetas morfológicas utilizadas en OSTA se basa en las etiquetas propuestas por el grupo *Expert Advisory Group on Language Engineering Standards* (EAGLES)¹⁹. A continuación se presentan cada una de las etiquetas definidas en una tabla y el conjunto de etiquetas utilizadas en OSTA. Para cada categoría se presentan los atributos, valores y códigos que puede tomar, así como algunos ejemplos relevantes. Las tablas en la que se presentan las etiquetas tienen el siguiente aspecto:

Categoría			
Posición	Atributo	Valor	Código
columna 1	columna 2	columna 3	columna 4

En la *columna 1* encontramos un número que hace referencia al orden y posición en que aparecen los atributos. La *columna 2* hace referencia a los atributos, el número de los cuales varía dependiendo de la categoría. En la *columna 3* encontramos los valores que puede tomar cada atributo y, finalmente, la *columna 4* representa los códigos que se han establecido para su representación. Las etiquetas en sí sólo son los códigos (columna 4) y se sabe a qué atributo pertenecen por la posición (columna 1) en la que se encuentran. Así NCMS000, correspondiente a *nombre* (N), *común* (C), *masculino* (M), *singular* (S), solo tiene cuatro códigos, mientras que VMIS2P0, *verbo* (V), *principal* (M), *pretérito perfecto* (I), *simple* (S), *2 persona* (2), *plural* (P), tiene seis.

Adjetivos			
Posición	Atributo	Valor	Código
1	Categoría	Adjetivo	A
2	Tipo	Calificativo	Q
		Ordinal	O
3	Grado	Aumentativo	A
		Diminutivo	D
		Comparativo	C
		Superlativo	S
4	Género	Masculino	M
		Femenino	F
		Común	C
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Función	-	O
		Participio	P

¹⁹ <http://www.ilc.cnr.it/EAGLES/browse.html>

Adverbios			
<i>Posición</i>	<i>Atributo</i>	<i>Valor</i>	<i>Código</i>
1	Categoría	Adverbio	R
2	Tipo	General	G
		Negativo	N
		-mente	M

Conjunciones			
<i>Posición</i>	<i>Atributo</i>	<i>Valor</i>	<i>Código</i>
1	Categoría	Conjunción	C
2	Tipo	Coordinada	C
		Subordinada	S

Determinantes			
<i>Posición</i>	<i>Atributo</i>	<i>Valor</i>	<i>Código</i>
1	Categoría	Determinante	D
2	Tipo	Demostrativo	D
		Posesivo	P
		Interrogativo	T
		Exclamativo	E
		Indefinido	I
		Artículo	A
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Poseedor	Singular	S
		Plural	P

Interjecciones			
<i>Posición</i>	<i>Atributo</i>	<i>Valor</i>	<i>Código</i>
1	Categoría	Interjección	I

Nombres			
Posición	Atributo	Valor	Código
1	Categoría	Nombre	N
2	Tipo	Común	C
		Propio	P
3	Género	Masculino	M
		Femenino	F
		Común	C
4	Número	Singular	S
		Plural	P
		Invariable	N
5-6	Clasificación semántica	Persona	P0
		Lugar	G0
		Organización	O0
		Otros	V0
7	Grado	Aumentativo	A
		Diminutivo	D

Numerales			
Posición	Atributo	Valor	Código
1	Categoría	Cifra	Z

Preposiciones			
Posición	Atributo	Valor	Código
1	Categoría	Adposición	S
2	Tipo	Preposición	P
3	Forma	Simple	S
		Contraída	C
4	Género	Masculino	M
5	Número	Singular	S

Pronombres			
Posición	Atributo	Valor	Código
1	Categoría	Pronombre	P
2	Tipo	Personal	P
		Demostrativo	D
		Posesivo	X
		Indefinido	I
		Interrogativo	T
		Relativo	R

		Exclamativo	E
		Clítico	C
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Impersonal Invariable	N
6	Caso	Nominativo	N
		Acusativo	A
		Dativo	D
		Oblicuo	O
7	Poseedor	Singular	S
		Plural	P
8	Politeness	Polite	P

Verbos			
Posición	Atributo	Valor	Código
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
4	Tiempo	Presente	P
		Imperfecto	I
		Futuro	F
		Pasado	S
		Condicional	C
5	Persona	Primera	1
		Segunda	2

		Tercera	3
6	Número	Singular	S
		Plural	P
7	Género	Masculino	M
		Femenino	F

Signos de puntuación			
<i>Posición</i>	<i>Atributo</i>	<i>Valor</i>	<i>Código</i>
1	Categoría	Puntuación	F

ADJETIVOS**Ordinal**

AO0MS0
AO0MP0
AO0FS0
AO0FP0

Calificativo

AQ0MS0
AQ0MP0
AQ0FS0
AQ0FP0
AQ0CS0
AQ0CP0
AQ0CN0

ADVERBIOS**General**

RG

-mente

RM

Negativo

RN

CONJUNCIONES**Coordinada**

CC

Subordinada

CS

DETERMINANTES**Artículo²⁰**

DA0MS0
DA0MP0
DA0FS0
DA0FP0
DA0NS0

Demostrativo

DD0MS0
DD0MP0
DD0FS0
DD0FP0
DD0CS0
DD0CP0
DD2FS0
DD2MS0

Exclamativo

DE0MS0
DE0MP0
DE0FS0

DE0FP0

DE0CS0

DE0CP0

DE0CN0

Indefinido

DI0MS0

DI0MP0

DI0FS0

DI0FP0

DI0CS0

DI0CP0

Posesivo

DP1CSS

DP1CPS

DP2CSS

DP2CPS

DP3CS0

DP3CP0

DP1MSP

DP1MPP

DP1FSP

DP1FPP

DP2MSP

DP2MPP

DP2FSP

DP2FPP

Interrogativo

DT0MS0

DT0MP0

DT0FS0

DT0FP0

DT0CS0

DT0CP0

DT0CN0

INTERJECCIONES

I

NOMBRES**Común**

NCMS000

NCMP000

NCFS000

NCFP000

NCMN000

NCFN000

NCCS000

NCCP000

NCCN000

Topónimo

NP000G0

Antropónimo

NP000P0

Extranjerismo

NCEX000

NUMERALES

Z

PREPOSICIONES

SPS00

PRONOMBRES**General**

P0000000

P01CP000

P02CP000

P03CN000

Clíticos

PC1CS000

PC2CS000

PC3FSA00

PC3MSA00

PC3CN000

PC3CSD00

PC0CN000

PC1CP000

PC2CP000

PC3MPA00

PC3FPA00

PC3CPD00

Demostrativo

PD0MS000

PD0MP000

PD0FS000

PD0FP000

PD0CS000

PD0CP000

PD0NS000

Exclamativo

PE000000

PE0MS000

PE0MP000

PE0FS000

PE0FP000

PE0CS000

PE0CP000

Indefinido

PI0MS000

PI0MP000

²⁰ Sólo se consideran las formas del artículo *definido*. El artículo *indefinido* se trata como

determinante indefinido o como *numeral*.

PI0FS000	PX1FS0S0	Futuro de indicativo
PI0FP000	PX1FP0S0	VAIF1S0
PI0CS000	PX1MS0P0	VAIF2S0
PI0CP000	PX1MP0P0	VAIF3S0
Personal	PX1NS0P0	VAIF1P0
PP1CSN00	PX1FS0P0	VAIF2P0
PP1CS000	PX1FP0P0	VAIF3P0
PP1CS000	PX2MS0S0	Condicional
PP1CP000	PX2MP0S0	VAIC1S0
PP1MP000	PX2NS0S0	VAIC2S0
PP1FP000	PX2FS0S0	VAIC3S0
PP2CSN00	PX2FP0S0	VAIC1P0
PP2CS000	PX2MS0P0	VAIC2P0
PP2CS000	PX2MP0P0	VAIC3P0
PP2CP000	PX2NS0P0	Presente de subjuntivo
PP2CP00P	PX2FS0P0	VASP1S0
PP2CS00P	PX2FP0P0	VASP2S0
PP2MP000	PX3MS000	VASP3S0
PP2FP000	PX3MP000	VASP1P0
PP3MS000	PX3FS000	VASP2P0
PP3FS000	PX3FP000	VASP3P0
PP3MP000	PX3NS000	Pretérito imperfecto de subjuntivo
PP3FP000		VASI1S0
PP3NS000	VERBOS (AUXILIAR)	VASI2S0
PP3FSA00	Infinitivo	VASI3S0
PP3FPA00	VAN0000	VASI1P0
PP3MSA00	Gerundio	VASI2P0
PP3CNA00	VAG0000	VASI3P0
PP3MPA00	Participio pasado	Futuro de subjuntivo
PP3CSD00	VAP00SF	VASF1S0
PP3CPD00	VAP00SM	VASF2S0
PP3CN000	VAP00PF	VASF3S0
PP3CN000	VAP00PM	VASF1P0
Relativo	Presente de indicativo	VASF2P0
PR000000	VAIP1S0	VASF3P0
PR0MS000	VAIP2S0	Imperativo
PR0MP000	VAIP3S0	VAM02S0
PR0FS000	VAIP1P0	VAM03S0
PR0FP000	VAIP2P0	VAM01P0
PR0CS000	VAIP3P0	VAM02P0
PR0CP000	Pretérito imperfecto	VAM03P0
PR0CN000	VAII1S0	
Interrogativo	VAII2S0	
PT000000	VAII3S0	VERBOS (PRINCIPAL)
PT0MS000	VAII1P0	Infinitivo
PT0MP000	VAII2P0	VMN0000
PT0FS000	VAII3P0	Gerundio
PT0FP000		VMG0000
PT0CS000	Pretérito perfecto simple	Participio presente
PT0CP000	VAIS1S0	VMPP0SC
PT0CN000	VAIS2S0	VMPP0PC
Posesivo	VAIS3S0	Participio pasado
PX1MS0S0	VAIS1P0	VMP00SF
PX1MP0S0	VAIS2P0	VMP00SM
PX1NS0S0	VAIS3P0	

VMP00PF
VMP00PM

Presente de indicativo

VMIP1S0
VMIP2S0
VMIP3S0
VMIP1P0
VMIP2P0
VMIP3P0

Pretérito imperfecto

VMII1S0
VMII2S0
VMII3S0
VMII1P0
VMII2P0
VMII3P0

Pretérito perfecto simple

VMIS1S0
VMIS2S0
VMIS3S0
VMIS1P0
VMIS2P0
VMIS3P0

Futuro de indicativo

VMIF1S0
VMIF2S0
VMIF3S0
VMIF1P0
VMIF2P0
VMIF3P0

Condicional

VMIC1S0
VMIC2S0
VMIC3S0
VMIC1P0
VMIC2P0
VMIC3P0

Presente de subjuntivo

VMSP1S0
VMSP2S0
VMSP3S0
VMSP1P0
VMSP2P0
VMSP3P0

Pretérito imperfecto de subjuntivo

VMSI1S0
VMSI2S0
VMSI3S0
VMSI1P0
VMSI2P0
VMSI3P0

Futuro de subjuntivo

VMSF1S0

VMSF2S0
VMSF3S0
VMSF1P0
VMSF2P0
VMSF3P0

Imperativo

VMM02S0
VMM03S0
VMM01P0
VMM02P0
VMM03P0

VERBOS (SEMIAUXILIAR)

Infinitivo

VSN0000

Gerundio

VSG0000

Participio pasado

VSP00SF
VSP00SM
VSP00PF
VSP00PM

Presente de indicativo

VSIP1S0
VSIP2S0
VSIP3S0
VSIP1P0
VSIP2P0
VSIP3P0

Pretérito imperfecto

VSII1S0
VSII2S0
VSII3S0
VSII1P0
VSII2P0
VSII3P0

Pretérito perfecto simple

VSIS1S0
VSIS2S0
VSIS3S0
VSIS1P0
VSIS2P0
VSIS3P0

Futuro de indicativo

VSIF1S0
VSIF2S0
VSIF3S0
VSIF1P0
VSIF2P0
VSIF3P0

Condicional

VSIC1S0
VSIC2S0

VSIC3S0
VSIC2P0
VSIC1P0
VSIC3P0

Presente de subjuntivo

VSSP1S0
VSSP2S0
VSSP3S0
VSSP1P0
VSSP2P0
VSSP3P0

Pretérito imperfecto de subjuntivo

VSSI1S0
VSSI2S0
VSSI3S0
VSSI1P0
VSSI2P0
VSSI3P0

Futuro de subjuntivo

VSSF1S0
VSSF2S0
VSSF3S0
VSSF2P0
VSSF1P0
VSSF3P0

Imperativo
VSM02S0
VSM03S0

VSM01P0
VSM02P0
VSM03P0

SIGNOS DE PUNTUACIÓN
F



Ysopete-Zaragoza, 1489

WWW.HISPANICSEMINARY.ORG