

ANOTACIÓN EN XML-TEI PARA LA EDICIÓN DE CARTAS: DOCUMENTACIÓN DEL PROYECTO CAREXIL-FR

Versión 2.0 (09/05/2024)

Marta López Izquierdo

SUMARIO:

1. [Introducción](#)

2. [<text>](#): el texto

2.1. [body](#): la organización textual

2.1.1. [opener](#) y [closer](#)

2.1.1.1. [placeName](#)

2.1.1.2. [date](#)

2.1.1.3. [salute](#)

2.1.1.4. [signed](#)

2.1.1.5. [seg](#)

- [@subtype](#)

- [@type](#)

2.1.2. [p](#): cuerpo

2.1.3. [postScript](#)

2.1.4. [seg @hand](#): anotación

2.1.5. [seg @type](#): fórmulas epistolares

a) exordio: “harangue”

b) conclusión: “peroration”

2.1.6. [list](#): listas dentro del texto

2.2. [Transcripción y edición digital múltiple](#):

2.2.1. [Separación de palabras](#)

2.2.2. [Puntuación](#)

2.2.2.1. Guiones

2.2.2.2. Comillas

2.2.3. [Abreviaturas](#)

2.2.4. [Paginación](#):

2.2.4.1. [<pb>](#) inicio de página: [@n @fac](#)s

2.2.4.2. [<lb>](#) inicio de línea: [@n](#)

2.2.5. [](#) texto cancelado: [@hand](#)

2.2.6. [<add>](#) texto añadido: [@hand @place](#)

- [above](#)

- [below](#)

- [inline](#)

- [margin-top](#)

- [margin-left](#)

- [margin-right](#)

- [back](#)

2.2.7. [<hi>](#) texto resaltado: [@rend @hand](#)

2.2.8. [<gap>](#) texto omitido: [@extent @reason](#)

- [damage](#)

- [hidden](#)

- [cancelled](#)

- [illegible](#)

- abandoned

- 2.2.9. <supplied> texto conjeturado: @resp @reason
- damage
- hidden
- cancelled
- illegible
- abandoned

2.2.10. <unclear>

2.2.11. <handshift> @resp

2.2.12. <xml:lang>

1. Introducción

Tras la anotación de los metadatos (cf. documento “Anotación de metadatos en XML-TEI...”), contenidos en el elemento <teiHeader>, se procede a la transcripción paleográfica del texto epistolar con marcado XML-TEI. La plantilla de referencia propuesta como documento de trabajo a los colaboradores de CAREXIL-FR tendrá que ser adaptada a las múltiples variaciones que podemos encontrar en las cartas del fondo editado¹.

Como para los metadatos, no se pueden crear etiquetas nuevas ni modificar las existentes de manera individual. Si se necesita anotar algún contenido que no se ha previsto en este documento, es necesario ponerse en contacto con la coordinadora del proyecto, que estudiará la posibilidad de introducir cambios y lo comunicará al conjunto del equipo.

2. <text>: el texto epistolar

El texto de la carta junto a su marcación xml se sitúa dentro del elemento <body> que está anidado en <text>. La marcación del texto dará cuenta, a través de su desarrollo sucesivo, de los diversos aspectos que se estudian en el proyecto CAREXIL: 1. organización textual, 2. transcripción y edición crítica, 3. rasgos sociolingüísticos y 4. información histórica.

2. 1. <body> : la organización textual

Siguiendo los estudios sobre textos epistolares, distinguimos los siguientes segmentos en una carta: encabezamiento, exordio (opcional), cuerpo, conclusión (opcional), despedida y postdata (opcional). Las etiquetas utilizadas dan cuenta de esta estructura textual.

Utilizamos cuatro elementos:

<opener>: agrupa la fecha, autoría, fórmula introductiva o de saludo utilizada al principio de una carta.

<p>: marca cada uno de los párrafos que componen el cuerpo de la carta.

<closer>: agrupa saludos, fechas u otras expresiones que aparecen al final de la carta.

<postscript>: posdata o información complementaria situada al final de una carta, tras la firma.

¹ Este esquema se basa en gran medida en otros proyectos realizados con la plataforma TEITOK, en particular el esquema del proyecto Post-Scriptum, desarrollado por Gael Vaamonde (2019), disponible en <http://ps.clul.ul.pt/es/index.php?>. La anotación de los metadatos está disponible en el documento “Anotación de metadatos en XML-TEI para corpus epistolares: documentación del proyecto CAREXIL-FR”.

EJEMPLO:

```
<text>
  <body>
    <opener>
      <!--Texto de apertura de la carta aquí-->
    </opener>
    <p>cuerpo de la carta aquí</p>
    <closer>
      <!--Texto de cierre de la carta aquí-->
    </closer>
    <postscript>
      <p> Texto de la postdata aquí</p>
    </postscript>
  </body>
</text>
```

2.1.1. <opener> y <closer>:

Se descomponen en:

<placeName>, <date>, <personName>, <orgName>, <salute>, <signed>, <address>, <addrLine>...

Ejemplo:

```
<opener>
  <lb n="1"/>
  <placeName corresp="ner.xml#place-Miellin">
    Miellin
  </placeName>
  <date>
    7-1-40
  </date>
  <lb n="2"/>
  <orgName corresp="ner.xml#org-CAEERF" role="addressee">
    A. La. Comision. de Ayuda
  </orgName>
  <lb n="3" />
  a Los niños Españoles
  </orgName>
  <lb/>
  <lb n="4"/>
  <salute>
    Dis tin gidos Señores.
  </salute>
</opener>
```

Ejemplo:

```
<closer>
  <salute>
```

```

        <lb n="5"/>Esperando que Atenderan
        <lb n="6"/>Midemanda y dandoles
        <lb n="7"/>Mil gracias An ticipadas Les saludo
        <lb n="8"/>a tenta Mente
    </salute>
    <lb/>
    <signed>
        <persName corresp="ner.xml#pers-PP" role="author">
            Paula Perez
        </persName>
    </signed>
</closer>

```

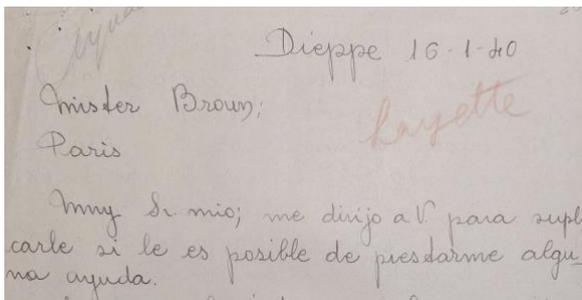
Como puede observarse en los ejemplos, los nombres de persona y de lugar llevan el atributo **@corresp** con la especificación de un código que remite al fichero central (ner.xml). Para cada persona, el código se compone con sus iniciales en mayúscula (dos o tres, según disponibilidad): Paula Pérez: id. **pers-PP**. El mismo sistema se usa con las organizaciones, que se citan por sus siglas en mayúsculas (CAEERF).

Para los nombres de lugar, se menciona el nombre de la localidad en su grafía contemporánea: **place-Miellin**.

Para cada nombre de persona y de lugar que se menciona en una carta, será necesario abrir un apartado en la sección correspondiente del fichero central, nerx.xml. (Cf. documento: “Anotación de entidades en XML-TEI para el fichero central de CAREXIL-FR”).

A veces, el saludo inicial o final se encuentran en el mismo bloque gráfico que el cuerpo. En esos casos, se incluyen en **<p>**, dentro de un elemento **<seg>**, que se especifica con los atributos **@subtype** (placeName, date, salute, signed) o **@type** (opener, closer).

Ejemplo:



Carta CAREXIL_5197030

```

<p>
    <lb n="4"/>
    <seg type="opener" subtype="salute">
        Muy Sr mio;
    </seg>
    me dirijo a V. para supli-
    <lb n="5"/>carle si le es posible de prestarme algu-
    <lb n="6"/>na ayuda.
</p>

```

Si la carta la firman varios autores, se sitúa cada firma dentro de un elemento `<signed>`, con marca de línea propia, `<lb/>`, si es el caso:

```
<closer>  
  <lb n="10"/><signed>E. Valverde</signed>  
  <lb n="11"/><signed> Miguel Martin</signed>  
  <lb n="12"/><signed> P. Guerrero</signed>  
</closer>
```

En las cartas de Carexil, no marcamos las rúbricas que acompañan las firmas. Estas podrán visualizarse directamente en el facsímil que acompaña cada transcripción en la plataforma.

2.1.2. `<p>`: cuerpo

Se considera que hay salto de párrafo cuando hay una o más líneas en blanco o cuando hay un renglón interrumpido.

2.1.3. `<postScript>`: postdata

Consideramos texto de la postdata el que aparece debajo de la firma que cierra la carta y es del mismo autor que esta, cuyos contenidos son distintos a los propios del closer (es decir, no es saludo, ni fecha, ni firma, ni nombre de lugar). Si es texto de otra mano, se marca como anotación con `<seg @hand>` (v. infra). No es necesario que el texto esté indicado en el original como P.D. o posdata para incluirla en esta categoría. La postdata puede incluir elementos de cierre, que se marcan con `<seg>` cuando no hay separación gráfica. Si encontramos separación gráfica, estos elementos se consideran un `<closer>` diferenciado.

2.1.4. `<seg @hand>`: anotación

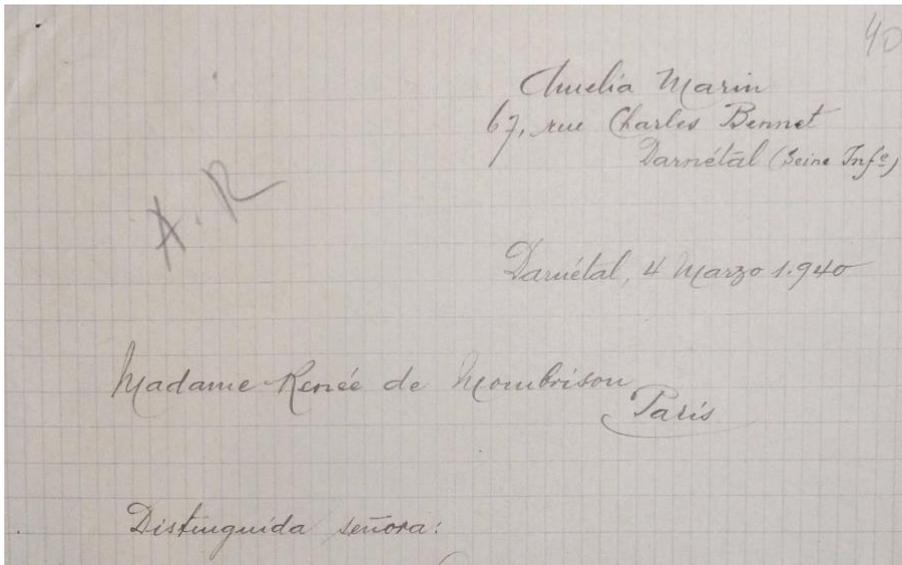
Para indicar un segmento escrito por otra mano, que no pertenece al texto de la carta, se usa el elemento `<seg>` seguido del atributo `@hand` y de su identificación siempre que es posible. Esta marca servirá esencialmente para las notas que añade la CAEERF a la recepción de la carta o antes de archivarla, o bien los números que ha añadido el archivo ruso (AM).

El segmento se coloca en la parte de la carta en la que aparezca: si está al principio, se sitúa en el opener, si en el cuerpo, en `<p>`, etc.

Para su diferente visualización en TEITOK, añadimos marca de cursiva: `<hi rend="italic">...</hi>`.

Especificamos si está en una lengua distinta al español con el atributo `@xml:lang`. A menudo, las anotaciones de la CAEERF están en francés: `<seg xml:id="fr">`.

Ejemplo:



Carta "CAREXIL_5197040".

```
<opener>
  <lb/>
  <seg hand="AM">
    <hi rend="italic">
      40
    </hi>
  </seg>
  <lb n="1"/>
  <persName corresp="ner.xml#pers-AMC" role="author">
    Amelia Marin
  </persName>
  <lb n="2"/>
  <address>
    <addrLine>
      67, rue Charles Bennet
    </addrLine>
  <lb n="3"/>
  <addrLine>
    <placeName corresp="ner.xml#place-Darnétal">
      Darnétal
    </placeName>(Seine Inf)
  </addrLine>
  </address>
  <lb/>
  <seg hand="CAEERF" xml:id="fr">
    <hi rend="italic">
      A.R
    </hi>
  </seg>
  <lb n="4"/>
  <placeName corresp="ner.xml#place-Darnétal">
    Darnétal,
```

```

</placeName>
<date>
4 Marzo 1.940
</date>
<lb/>
<lb/>
<lb/>
<lb n="5"/>
<persName corresp="ner.xml#pers-RDM" role="addressee">
Madame Renée de Mombrison
</persName>
<lb n="6"/>
<placeName corresp="ner.xml#place-Paris">
<hi rend="underlined">
París
</hi>
</placeName>
<lb/>
<lb/>
<lb/>
<lb n="7"/>
<salute>
Distinguida señora
</salute>:
<lb/>
</opener>

```

2.1.5. <seg @type="harangue/peroration"> fórmulas epistolares

Determinadas expresiones formularias se sitúan en posición inicial (exordios “harangue”) o final (conclusión “peroration”). Lo marcamos con el elemento <seg> seguido del atributo @type.

Ejemplo:

```

<p> <lb n="5"/> <seg type="harangue">Tiniendo el ha trebimiento de
<lb n="6"/>es cribirles Soloes por la necesida
<lb n="7"/>en que me encuentro.</seg>
</p>

```

2.1.6. <list>: listas dentro del texto

Las cartas pueden contener listas de personas, prendas, objetos, etc. Se marca toda la lista con el elemento <list> y cada una de las menciones de la lista con el elemento <item>. Téngase en cuenta que el elemento <list> no puede estar anidado en el elemento <p>².

Ejemplo:

² <https://tei-c.org/release/doc/tei-p5-doc/fr/html/ref-org.html>

```
<p>
  <lb n="1"/>Para la niña de 5 años
</p>

<list>
  <lb n="2"/><item>Mudas interiores </item>
  <lb n="3"/><item>medias ó calcetines </item>
  <lb n="4"/><item>refajo</item>
  <lb n="5"/><item>quimono de abrigo </item>
  <lb n="6"/><item>calzado </item>
  <lb n="7"/><item><hi rend= "underlined">sueter de </hi> abrigo</item>
</list>
```

Carta CAREXIL_5197050.

2.2. Transcripción y edición digital múltiple:

En CAREXIL realizamos una edición paleográfica de las cartas estudiadas. Respetamos separación de palabras, cambios de línea y párrafo, tachones, lagunas, adiciones de todo tipo, así como la ortografía, incluyendo el uso de minúsculas y mayúsculas, y la puntuación originales de los autores de las cartas.

En una segunda versión, normalizada, la separación de palabras y la ortografía para facilitar su lectura y permitir el tratamiento automatizado del texto. Mantenemos sin embargo la puntuación del texto original.

2.2.1. La separación de palabras

En CAREXIL es importante mantener la separación de palabras presente en los originales pues consideramos que es un rasgo determinante para caracterizar el grado de alfabetización de los autores. Así pues, aunque en la versión normalizada (v. *infra*) sí corregimos este aspecto, no lo reflejamos en la edición paleográfica.

La hiposegmentación o la hipersegmentación son dos de los parámetros que tendremos en cuenta en nuestro análisis lingüístico y en la clasificación de los documentos.

2.2.2. La puntuación

Las cartas que recoge CAREXIL hacen uso de signos de puntuación bastante convencionales (punto, coma, dos puntos, guión...), aunque no siempre se utilizan de manera convencional. Respetamos la puntuación tal y como aparece en las cartas originales para la edición paleográfica y la conservamos en la versión normalizada.

2.2.2.1. Guiones

El tratamiento de guiones (-), guiones largos (—) y rayas (—) debe intentar reflejar el uso que hacen los refugiados de estos tres símbolos. Se recomienda sin embargo preferir guiones (-) cuando aparecen en fechas, y siempre a final de línea para marcar separación de palabras (no usar guiones bajos, aunque lo utilicen los refugiados). los guiones a final de línea para separar palabra se pegan siempre a la palabra cortada.

Asimismo, preferir guiones largos cuando el refugiado elabora listas en las que separa distintas informaciones sobre lo que pide. Incluir espacios delante y detrás del guión largo:

María – 3 años – calzado número 20

En las fechas, los espacios delante o detrás de guion respetan lo que aparezca en la carta. En TEITOK, cada cifra y cada guion deberá ir identificado por su propio token.

Por último, la raya se usa solo cuando el valor es parentético. Usarla según normativa ortográfica actual: sin espacio delante y detrás del texto que se encierra entre rayas, con espacio delante y detrás del texto que precede y sigue:

me dijo —no sé por qué— que no vendría

2.2.2.2. Comillas

Usar siempre comillas inglesas (“...”) cuando se introduce texto entre ellas.

Usar comillas rectas " " para señalar repetición de elementos:

Edad	8	Nº	30
" "	10	" "	33

2.2.3. Las abreviaturas

En la transcripción paleográfica, recogemos las abreviaturas sin desarrollar. Su desarrollo y normalización se llevan a cabo en una fase posterior, directamente en la plataforma TEITOK, que propone una interfaz ad-hoc para ello.

Si la abreviatura es un grupo de palabras, se transcribe cada palabra abreviada con su punto, seguido de un espacio:

que besa su mano → q. b. s. m.

En TEITOK, cada palabra del grupo será un token distinto con su propio etiquetado. El desarrollo de la abreviatura se realiza a nivel de “Full form”.

2.2.4. Paginación y líneas

Utilizamos este tipo de elementos para subdividir el texto en unidades como página o línea.

2.2.4.1. <pb>: inicio de página

Indica inicio de página (‘page beginning’). Se acompaña de dos atributos:

- @n: introduce el número de página del documento, dentro de una numeración interna (que no coincide con el número atribuido por el archivo de Moscú, que sigue una numeración propia). Así, una carta escrita a una sola cara, tendrá una sola página : <pb n="1">, a dos caras, dos páginas (n="1" y n="2"), etc.

- @facs (facsimil): indica la imagen que se corresponde con este elemento. El contenido de este atributo debe presentar la forma exacta del nombre del archivo de imagen, incluida su extensión.

Ejemplo:

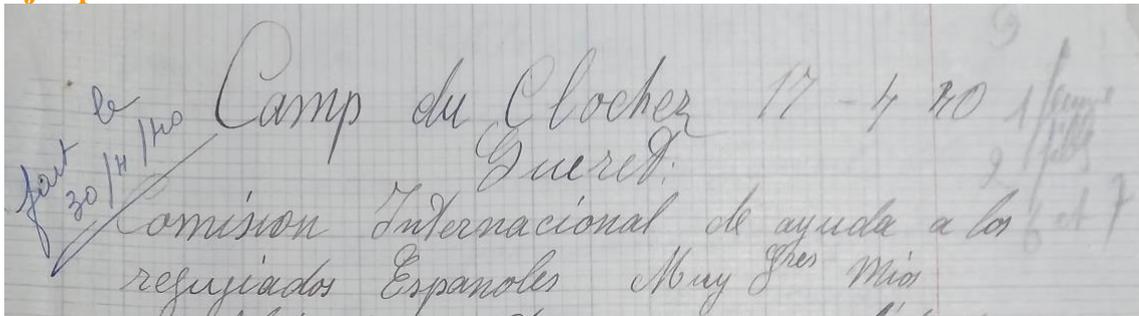
<pb n="1" facs="CAREXIL_3107028r.jpg"/>

2.2.4.2. <lb>: inicio de línea

Indica inicio de línea ('line beginning').

Como en el caso de <pb>, el atributo @n indica el número de la línea: solo se indica número en las líneas con texto escrito por el autor de la carta. Cuando las líneas están en blanco, se marca salto de línea sin numerar: <lb/>. No se numeran las líneas de las anotaciones hechas por otra mano (el Archivo de Moscú o la CAEERF, por ejemplo), si no forman parte del texto de la carta; solo se marca salto de línea.

Ejemplo:



Carta CAREXIL_4139009

```
<opener>
  <lb/>
  <seg hand="AM">
    <hi rend="italic">
      9
    </hi>
  </seg>
  <lb/>
  <seg hand="CAEERF" xml:lang="fr">
    <hi rend="italic">
      fait le
    </hi>
  </seg>
  <lb/>
  30/4/40
  </hi>
</placeName>
  Camp du Clocher
</placeName>
<date>
```

```

17-4-40
</date>
</lb>
<seg hand="CAEERF" xml:lang="fr">
  <hi rend="italic">
    1 femme 2 filles 6 et 7
  </hi>
</seg>
</lb>
<placeName>
  Gueret
</placeName>
</lb>
<orgName>
  Comision Internacional de ayuda a los
</lb>
  Refugiados Españoles
</orgName>
<salute>
  Muy Sres Mios
</salute>
</opener>

```

2.2.5. : texto cancelado

El elemento marca un segmento de texto que ha sido cancelado pero que es posible descifrar. Este elemento lleva el atributo @hand, que permite indicar la referencia del autor de la cancelación por medio de sus siglas:

Ejemplo:

CS = Cesaria Salgado
 <del hand="#CS"> ese dia

Este elemento se aplica al segmento que esté cancelado: palabra, secuencia o parte de palabra.

2.2.6. <add>: texto añadido

El elemento <add> sirve para marcar texto que se ha añadido a la carta propiamente dicha. Lleva dos atributos: @hand, para indicar el responsable de la adición; @place: especifica dónde está colocado este texto. Debe tenerse en cuenta que se trata siempre de añadidos internos de mano del autor del texto epistolar, y no de añadidos de una tercera persona que comentan, subrayan o modifican el texto epistolar del autor o autores (estas intervenciones se consideran anotaciones y se marcan con <seg> (cf. supra). .

Los lugares de la adición pueden ser:

- above: por encima de la línea
- below: por debajo de la línea
- inline: a la altura de la línea

margin-top: margen superior
margin-left: margen izquierdo
margin-right: margen derecho
margin-bottom: margen inferior
back: reverso

Si se borra un segmento para añadir otro, se marca con la combinación de elementos: `` para lo que se borra y `<add>` para lo que se añade. Si lo añadido está en el renglón, no se marca con `<add>`.

Los subrayados de otra mano que aparecen en el texto de la carta se incluyen como `<hi @rend="underlined" @hand="CAEERF">` en el punto de inserción del texto afectado. V. supra 2.2.7.

2.2.7. `<hi>` : texto resaltado

Con el elemento `<hi>` se marcan segmentos subrayados en el texto. El atributo `@rend` especifica de qué manera se ha resaltado el fragmento en cuestión.

En nuestras cartas, es muy frecuente que aparezcan palabras o frases subrayadas a lápiz de color azul o rojo, probablemente por parte de la CAEERF, destinatario de la correspondencia. Indicamos el origen del subrayado con el atributo `@hand`.

Ejemplo:

si pueden mandarnos zapatos y ropa se lo agradeceríamos...

```
<p>  
  si pueden mandarnos  
  <hi hand="CAEERF" rend="underlined">  
    zapatos y ropa  
  </hi>  
  se lo agradeceríamos  
</p>
```

2.2.8. `<gap>`: texto omitido

El elemento `<gap>` se utiliza para marcar segmentos omitidos en la transcripción, casi siempre por ser ilegibles o incomprensibles. El atributo `@reason` declara la causa de la omisión, y `@extent` la extensión estimada de texto omitido (en cantidad de líneas, palabras o letras, según la importancia del pasaje en cuestión).

Causas posibles de la omisión:

- **damage**: daño en el soporte (papel roto, manchado...)
- **hidden**: texto oculto por algún elemento (encuadernación...)
- **cancelled**: tachado que no permite leer lo que hay debajo
- **illegible**: indescifrable
- **abandoned**: omitido por estar incompleto

Ejemplo:

```
<gap extent="1 word" reason="deleted"/>
```

2.2.9. <supplied>: texto conjeturado

Cuando un segmento de texto presenta lectura dudosa se marca con <supplied>, seguido del atributo @resp, que incluye las siglas del nombre del agente responsable de la interpretación (el editor o el transcriptor, generalmente), y del atributo @reason, para indicar la causa de dicha dificultad. Las razones son muy similares a las indicadas *supra* para las omisiones:

damage, hidden, cancelled, illegible, abandoned.

La conjetura se aplica a una palabra entera, aunque solo una parte de ella sea de lectura dudosa. Si se conjetura más de una palabra, es necesario marcar cada una de ellas con el elemento <supplied>.

Ejemplo:

```
<supplied reason="hidden" resp="#MLI">mucho</supplied>
```

2.2.10. <unclear>: texto incomprensible

<unclear> es el elemento con el que marcamos segmentos de texto que no se puede leer con total certeza o que, aunque se pueden leer, resultan incomprensibles. Se diferencia de <gap> en que este elemento se reserva para contenidos ilegibles. Se utiliza para palabras o grupos de palabras.

Ejemplo:

lo vemos en el <unclear> arijo </unclear> de todos.

2.2.11. <handshift>: texto de nueva mano

El elemento <handshift> aparece delante de un segmento de texto en que cambia la mano del escriba (pero sin que estemos ante un nuevo texto). Puede también indicarse con él un cambio de estilo, tinta o letra.

Si se conoce su identidad, se indica con el atributo @resp y las siglas que lo identifican. Es algo que puede aparecer en nuestras cartas, escritas por una mano y firmadas por otra, seguramente porque el escaso dominio de la escritura no permite al autor de la carta más que firmarla y se pide ayuda a otro refugiado para que la escriba en su nombre.

Ejemplo:

```
</><signed>E. Valverde</signed>  
</><signed> <handShift resp="MM"/>Miguel Martin</signed>  
</><signed> <handShift resp="PG"/>P. Guerrero</signed>
```

2.2.12. xml:lang="fr": texto en francés

El atributo @**xml:lang** se usa dentro de los elementos <word> (<tok> en TEITOK), <seg>, etc, para marcar una palabra o todo un segmento cuando están en francés (e igualmente si está en inglés "en").

Ejemplo:

```
<seg hand="CAEERF" xml:lang="fr">  
  <hi rend="italic">demander pointure</hi>  
</seg>
```