

Encodage du document ou encodage du
processus?
un nouveau modèle TEI pour l' édition
génétique.

Elena Pierazzo and Lou Burnard

Est-ce que les manuscrits modernes sont le même que les manuscrits médiévaux?

- Ils ont des fonctions différentes: privé / prise de notes par rapport à la «publication», la circulation au public (la plupart du temps)
- Ils ont l'air différents: non planifiés vs planifié
- Modern manuscripts contiennent un mélange de texte et d'autres choses dont la phénoménologie n'est pas toujours facile de reconduire aux manuscrits médiévaux
- Ils peuvent être organisés dans des séquences de versions
- **Est-ce le TEI capable de fournir un soutien suffisant pour l'encodage des manuscrits modernes?**

Le début

- L'idée a d'abord été soulevée à la première réunion du TEI MS SIG: Nancy 2003
- La question a été posée à nouveau avec une haute priorité dans l'agenda du MS SIG en 2007
- 2007: Sortie officielle P5
- Groupe de travail dans MS SIG a commence son travail en Novembre 2008 (Londres, MM TEI)

Groupe de travail

- Fotis Jannidis (chair)
- Elena Pierazzo
- Malte Rehbein
- + Lou Burnard
- (Paolo D'lorio)
- Gregor Middell
- Moritz Wissenbach

L'atelier

- Soutenu par ALLC, ACH, TEI, Moore Institute – University of Galway
- Paris, ENS – ITEM, 14-15 mai 2009
- Experts d'éditions et de critique génétique, pas forcément d'édition numérique
- Anne Bohnenkamp, John Bryant, *Aurèle Crasson*, Jean-Daniel Fekete, Daniel Ferrer, Hans Walter Gabler, Axel Gellhaus, Almuth Grésillon, Claus Huitfeldt, Dirk van Hulle, Jean-Louis Lebrave, Wolfgang Lukas, Kenneth M. Price, *Kathryn Sutherland*

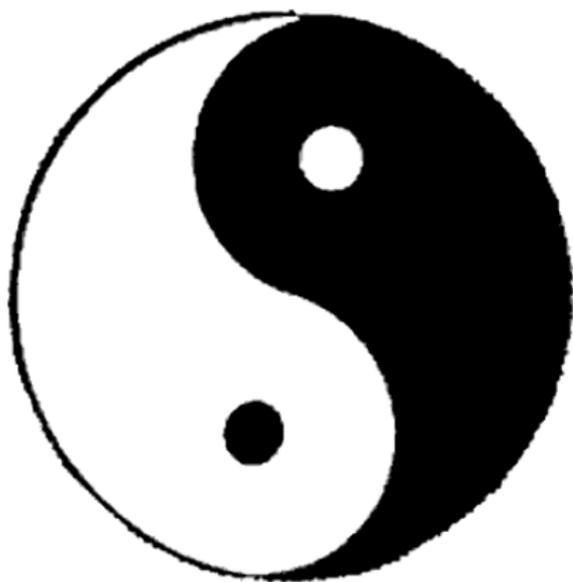
Quelque concepts de base

- Documents et textes
- Faits et interprétation
- État et processus
- Genèse: critique génétique, éditions génétiques
- Les actes d'écriture et les étapes du texte

Documents ou Texts



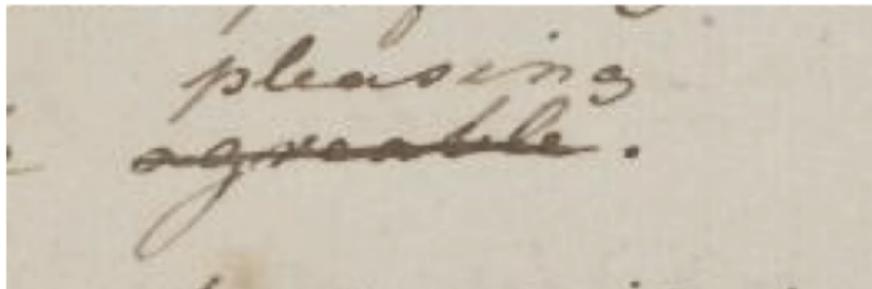
Documents et Texts, Yin et Yang, corps et âme



TEI fournit un moyen d'encoder des textes, mais parfois

- Le processus de création du document est au moins aussi important que le texte qu'il contient
- Le texte peut être déterminée par le document
- Le texte est présenté graphiquement
- Il n'y a pas de texte: textes non-linéaires sont des non-textes (D. Ferrer)

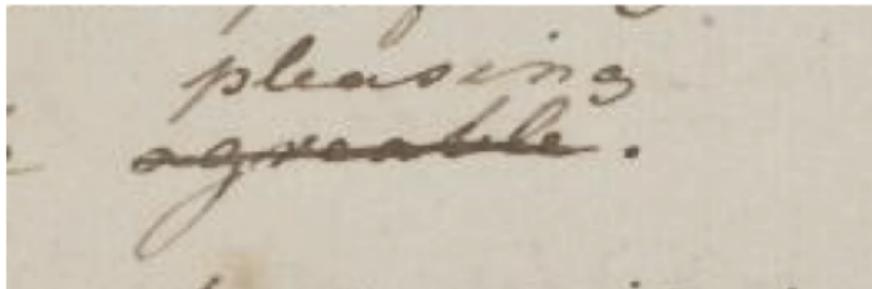
Que est-ce que c'est cela?



- 1 «agreable» est supprimé et remplacé par «pleasing»
- 2 «agreable» a une ligne au dessus, «pleasing» est écrit en haut dans l'espace interlinéaire.

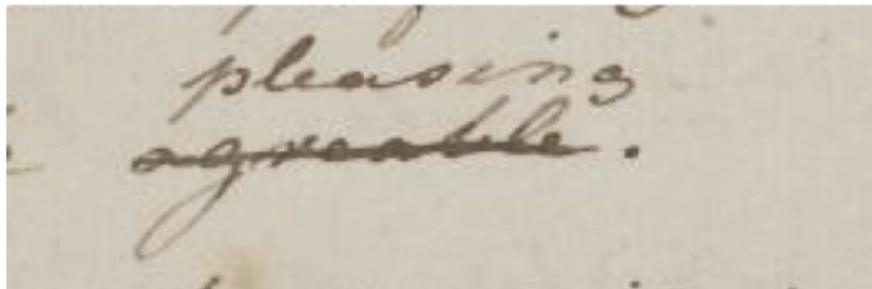
Peut-on vraiment les distinguer?

Que est-ce que c'est cela?



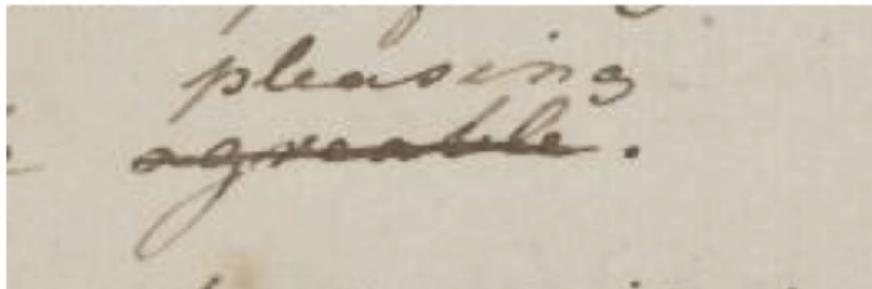
- 1 «agreable» est supprimé et remplacé par «pleasing»
- 2 «agreable» a une ligne au dessus, «pleasing» est écrit en haut dans l'espace interlinéaire.
- 3 La première version du texte contient «agreable», puis en un second moment le mot est supprimé et contextuellement «pleasing» est écrit dans l'espace interlinéaire.

Que est-ce que c'est cela?



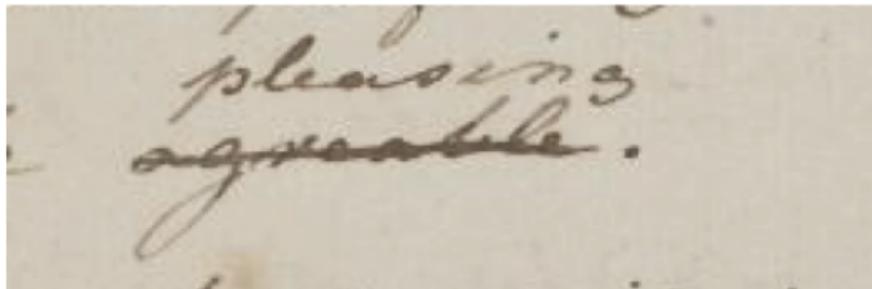
- 1 «agreable» est supprimé et remplacé par «pleasing»
- 2 «agreable» a une ligne au dessus, «pleasing» est écrit en haut dans l'espace interlinéaire.
- 3 La première version du texte contient «agreable», puis en un second moment le mot est supprimé et contextuellement «pleasing» est écrit dans l'espace interlinéaire.

Que est-ce que c'est cela?



- 1 «agreable» est supprimé et remplacé par «pleasing»
- 2 «agreable» a une ligne au dessus, «pleasing» est écrit en haut dans l'espace interlinéaire.
- 3 La première version du texte contient «agreable», puis en un second moment le mot est supprimé et contextuellement «pleasing» est écrit dans l'espace interlinéaire.

Que est-ce que c'est cela?



- 1 «agreable» est supprimé et remplacé par «pleasing»
- 2 «agreable» a une ligne au dessus, «pleasing» est écrit en haut dans l'espace interlinéaire.
- 3 La première version du texte contient «agreable», puis en un second moment le mot est supprimé et contextuellement «pleasing» est écrit dans l'espace interlinéaire.

Genèse de quoi

- Documents? (déchirures, coupures, taches ...)
- Textes? (corrections, modifications, ajouts, suppressions, transpositions?)

Les deux, bien sûr!

La balise <document>

- un frère de <text> et <facsimile>
- représente la structure physique d'un document, en termes de <surface>s écrites, et <zone>s de l'écriture.
- un nouvel élément <line> est nécessaire pour représenter les lignes topographiques
- un nouvel élément <patch> est nécessaire pour représenter des surfaces fragmentaires (les paperoles)

Samaria is a Greek
strand of water that
comes from the natural
springs of stilos, in
Crete.

Feed Birds in the park today.
Might write an article about
the Thick-billed Warbler.

1 April 2009

Samaria is a Greek
stand of water that
comes from the natural
springs of Stilos, in
Crete.

1 April 2009

Feeding birds in the park today.
Might write an article about
the Thick-billed Warbler.

```

<text>
  <body>
    <div type="diary-entry">
      <dateline>
        <date value="2009-04-01"> 1 April
2009 </date>
      </dateline>
      <p><lb/>Fed Birds in the park
today.<lb/> Might write an article about
<lb/>
        the Thick-billed Warbler. </p>
    </div>
    <div type="note" rend="rotated">
      <p><lb/>Samaria is a Greek <lb/>
brand of water that <lb/> comes from the
natural <lb/> springs of Stilos, in <lb/>
Crete</p>
    </div>
  </body>
</text>

```

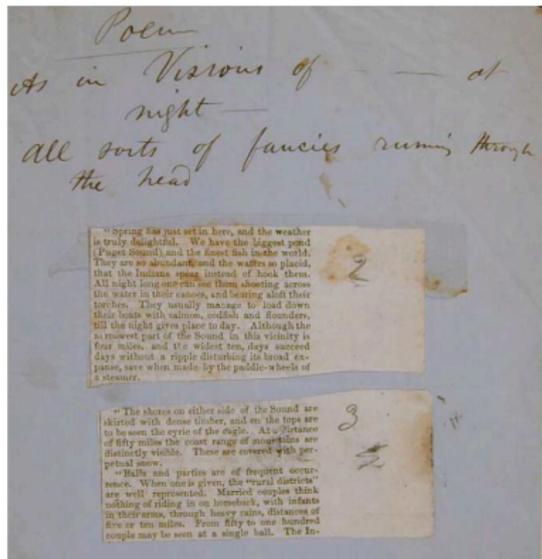
```

<ge:document>
  <surface ulx="0" uly="0" lrx="200" lry="300">
    <zone ulx="10" uly="43" lrx="185" lry="84"
rotate="0">
      <zone>
        <ge:line rend="right"> 1 April 2009 </ge:line>
      </zone>
      <zone>
        <ge:line>Fed Birds in the park today.</ge:line>
        <ge:line>Might write an article about </ge:line>
        <ge:line>the Thick-billed Warbler. </ge:line>
      </zone>
    </zone>
    <zone ulx="9" uly="20" lrx="70" lry="60" rotate="90">
      <ge:line>Samaria is a Greek </ge:line>
      <ge:line>brand of water that</ge:line>
      <ge:line>comes from the natural</ge:line>
      <ge:line>springs of Stilos, in </ge:line>
      <ge:line>Crete </ge:line>
    </zone>
  </surface>
</ge:document>

```


Les paperoles

```
<surface>
  <zone>
    <ge:line>Poem</ge:line>
    <ge:line>As in Visions of — at</ge:line>
    <ge:line>night —</ge:line>
    <ge:line>All sorts of fancies running through</
ge:line>
    <ge:line>the head</ge:line>
  </zone>
  <ge:patch
    type="newsprint"
    binder="glue"
    height="40"
    width="90"> Spring has
    just set in here, and the weather.... a steamer
  </zone>
    <ge:metaMark function="sequence">2</
ge:metaMark>
  </zone></ge:patch>
  <ge:patch
    type="newsprint"
    binder="glue"
    height="35"
    width="90"> "The shores
    on either side of the Sound are... The In- <zone>
    <ge:metaMark function="sequence">3</
ge:metaMark>
  </zone></ge:patch>
</surface>
```



La transcription des sources primaires

Quelles caractéristiques d'une source primaire pourrait-on vouloir exprimer dans un travail de transcription ?

- la manière dont le texte est inscrit sur la page
- l'orthographe
- le rendu formel original (s'il a une importance)
- la ponctuation originale
- les abréviations
- les ajouts et suppressions
- les erreurs et omissions
- éventuellement, les illustrations, dessins ou graphiques illustrant le texte
- la structure logique du texte
- peut-être d'autres choses, selon le point de vue qu'on adopte et les objectifs du projet...

Objectifs d'un projet de transcription

Il pourrait y en avoir plusieurs...

- mettre à disposition du plus grand nombre de chercheurs -- ou curieux -- des sources primaires
- ... en les rendant compréhensibles à un public moderne
- établir un texte définitif à partir de plusieurs témoins, compléter la transcription avec un stemma ou tableau de la tradition, et des notes d'apparat critique
- mettre en place un paratexte fournissant des notes explicatives ou historiques pour identifier un lieu, une personne, etc.

Mais en chaque cas, la transcription devrait s'appuyer sur des règles éditoriales qui seront explicitées, et qui seront elles-mêmes fondées sur des recommandations reconnues dans la discipline scientifique concernée.

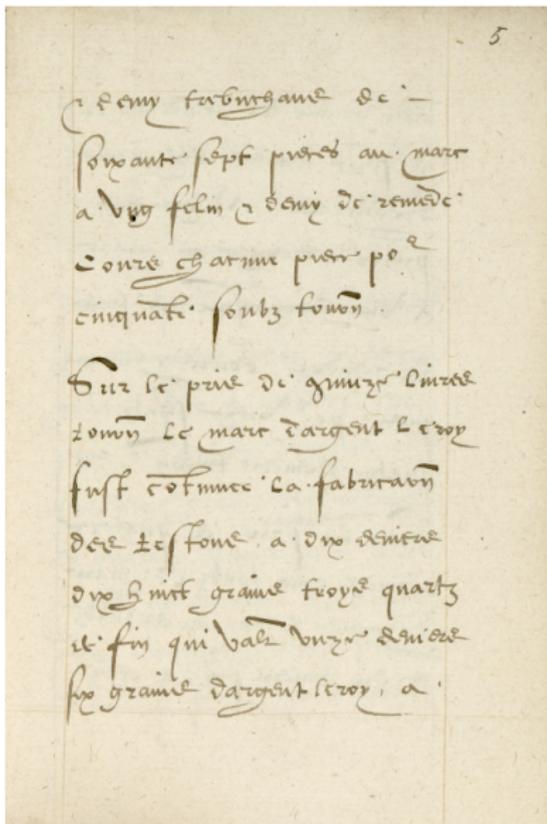
Les principaux éléments TEI utiles pour la transcription

Définis dans le module 'core' : <abbr>, <add>, <choice>, <corr>, , <expan>, <gap>, <sic>

Définis dans le module 'transcr' : <addSpan>, <am>, <damage>, <damageSpan>, <delSpan>, <ex>, <facsimile>, <fw>, <handNotes>, <handShift>, <restore><space>, <subst>, <supplied>, <surface>, <zone>

Définis dans le module 'textcrit' : <app>, <lacunaEnd>, <lacunaStart>, <lem>, <listWit>, <rdg>, <rdgGrp>, <variantEncoding>, <wit>, <witDetail>, <witEnd>, <witStart>, <witness>

Encodage de la structure physique : exemple

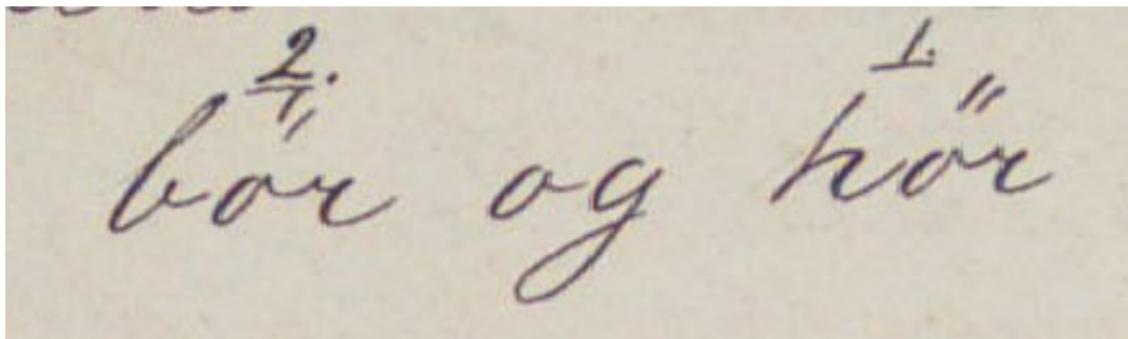


```
<p>
<!-- ... -->
<pb n="5r"/><fw>5</fw>
<lb/><expan>et</expan> demy trebuchans de
<lb/>soixante sept pieces au marc
<lb/>a ung felin <expan>et</expan> demy de
remede
<lb/>Cours chacune piece
<expan>pour</expan>
<lb/><expan>cinquante</expan> soubz
<expan>tournois</expan> <pc>.</pc>
</p> <p>
<lb/>Sur le pris de quinze livres
<lb/><expan>tournois</expan> le
marc dargent le roy
<lb/>fust <expan>continuee</expan> la
<expan>fabricacion</expan>
<lb/>des testons a dix deniers
<lb/>dix huict grains troys quartz
<lb/>de fin qui <expan>valent</expan> unze
deniers
<lb/>six grains dargent le roy<pc>,</pc> a
```

Balises supplémentaires pour la transcription

- `<mod>` et `<modSpan>` indiquent, de manière générique, un lieu de modification dans le document
- `<metaMark>` un signal quelconque déterminant la lecture du document
 - `<used>` un cas spécifique de metaMark, indiquant une partie rayée parce que déjà utilisée
- `<undo>`, `<redo>`, `<rewrite>` indications respectivement de: corrections supprimées, corrections réaffirmées, réécritures
- `<transpose>` et `<transposeGrp>` transpositions

<metaMark> exemplaire



```
<line>
  <seg xml:id="ib01">bör</seg>
  <metaMark rend="underline" function="transposition" tar-
get="#ib1" place="above">2.</metaMark> og
  <seg xml:id="ib02">hör</seg>
  <metaMark rend="underline" function="transposition" tar-
get="#ib02" place="above">1.
  </metaMark>
</line>
```

Transposition exemplaire

```
<transposeGrp>  
  <transpose>  
    <ptr target="#ib02"/>  
    <ptr target="#ib01"/>  
  </transpose>  
</transposeGrp>
```


<used> encodage

```
<used spanTo="#endOfZone1" type="rature"/>
<zone xml:id="zone1">
  <line> 28) le court de tennis. Les tribunes sont
    maintenant pleines de supporteurs. Deux joueurs</line>
<!-- ... -->
  <metaMark xml:id="metaMark" function="rature">tapé</metaMark>
<!-- ... -->
  <line>3<subst stage="#red">
    <del>3</del>
    <add>4</add>
  </subst>)
  Le mot FIN apparaît sur l'écran.</line>
</zone>
<anchor xml:id="endOfZone1"/>
```

Balisage des étapes d'écriture

On propose un attribut *@stage* qui associe n'importe quel élément de la transcription avec une étape identifiée dans l'évolution du document

L'identification d'une étape est au coeur du problématique génétique: on ne peut pas vous aider là!

Les étapes elles-mêmes sont à documenter dans le [<profileDesc>](#) de l'entête TEI.

Exemple continue

28) Le comte de Tercis. Le comte n'est pas né à Tercis, mais à Paris. Son père avait hérité de la comté. Il est allé à la cour, à Paris, mais n'a pas pu aller à Tercis (à la cour ? en France ?).

héritier.
charte
d'usage

29) A Paris le comte a hérité de son père. Mais il n'a pas pu aller à Tercis. Il a hérité de son père. Mais il n'a pas pu aller à Tercis. Il a hérité de son père. Mais il n'a pas pu aller à Tercis.

30) L'usage de Tercis est différent de celui de Paris. L'usage de Tercis est différent de celui de Paris. L'usage de Tercis est différent de celui de Paris.

usage de Paris
= coutume

31) La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis.

Président

32) La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis.

33) La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis.

Donc ce cas figure
de moyen âge
d'usage de Paris
ou son usage
de Paris
de Paris

En France
Fin

Par exemple

usage de Paris
resté d'usage

usage de Paris
resté d'usage

La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis. La coutume de Paris est différente de celle de Tercis.



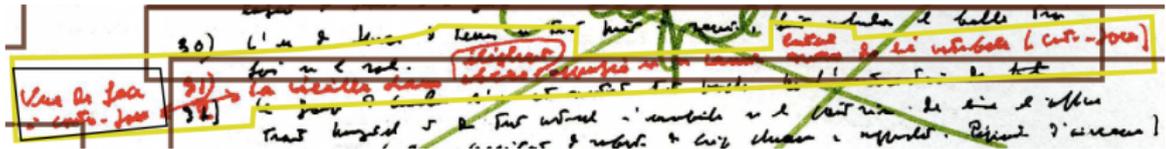
Encodage des étapes

```
<profileDesc>
  <creation>
    <stageNotes ordered="true">
      <stageNote xml:id="ST-1">Premier
        brouillon, en encre noir </stageNote>
      <stageNote xml:id="ST-2">Corrections et
        revisions en rouge</stageNote>
      <stageNote xml:id="ST-3">Corrections et
        revisions en bleu</stageNote>
      <stageNote xml:id="ST-4">Ratures et
        indications d'usage en vert</stageNote>
    </stageNotes>
  </creation>
</profileDesc>
```

Association d'etape avec la transcription

```
<zone xml:id="zone1" stage="#ST-1">
  <line> 28) le court de tennis. Les tribunes sont
    ... Deux joueurs</line>
<!-- ... -->
  <line>30) l'un des joueurs de tennis se tient ... trois</line>
  <line>fois sur le sol</line>
  <zone stage="#ST-2">
    <line>31) </line>
    <line>Vue de face</line>
    <line> a contre jour</line>
    <metaMark function="add"/>
    <line>la vieille dame ... dans le vestibule
      (contre-jour)</line>
  </zone>
  <line>3<subst stage="#ST-2">
    <del>1</del>
    <add>2</add>
  </subst>)<line>Le
    groupe de cavaliers ...</line>
<!-- ... -->
  <line>... Le mot FIN apparaît sur l'écran.</line>
</zone>
<zone stage="ST-3">
  <line>Dans ce cas...</line>
</zone>
```

Traitement d'un crux



```
<zone facs="#path4052" stage="#red">
  <line>31) <add>
    <metaMark function="add" target="#add1"/>
  </add> la vieille dame <seg xml:id="alt1">éléphant</seg>
  <seg rend="suplerlinear" xml:id="alt2">
    <metaMark function="add" target="#alt2" facs="#path6183"/>obèse</seg>
appuyée
  sur ses cannes <unclear>nous<add>latel</add>
  </unclear> dans le vestibule
  (contre-jour)</line>
</zone>
<zone facs="#path3070" xml:id="add1">
  <line>Vue de face</line>
  <line>à
    contre-jour</line>
</zone>
<alt targets="#alt1 #alt2" weights="0.5 0.5" mode="excl"/>
```

Où en sommes-nous?

- Premier rapport au TEI Members Meeting à Ann Arbor (Michigan) en 2009
- Béta test jusqu'à la fin de Janvier 2010
- Révisions : février - mars 2010: r
- Proposition quasiment complète soumise au TEI Council : avril 2010
- Ateliers et test au TEI Members Meeting à Zadar (Croatie): novembre 2010

<http://www.tei-c.org/Activities/Council/Working/tcw19.html>