CKCC Letter Format

Walter Ravenek
July 20, 2012

Introduction

Four of the correspondences in the CKCC corpus were digitized by the DBNL – Digitale Bibliotheek voor de Nederlandse Letteren (Digital Library of Dutch Literature). The letters were encoded in TEI format (Text Encoding Initiative, www.tei.c.org). Unfortunately the digitization of the DBNL focusses on the rendering of the letters and ignores many structural aspects of the letters. This made it necessary to do extensive preprocessing of the letters before they could be imported in our database. Therefore we have made up requirements for new material to be used in the CKCC project. We ask contributors to the project to comply with these requirements.

Document format

Each letter must be a separate document conforming to the TEI format described in the remainder of this section.

Basic structure

An example of the basic letter format is as follows:

```xml
<TEI>
<teiHeader>
<meta type="id" value="0084"/>
<meta type="date" value="1650-08-02"/>
<meta type="sender" value="huyg003"/>
<meta type="senderloc" value="denha004"/>
<meta type="recipient" value="huyg007"/>
<meta type="recipientloc" value="?"/>
<meta type="language" value="fr"/>
</teiHeader>
<text>
<body>
<div type="opener">
<p>Haije ce 2 Aug. 1650.</p>
</div>
<p>Nous n'avons pas reçu de vos nouvelles la semaine passée, ...
</p>
<p>Mais si d'aventure vous ne le savez pas encore nij aucune chose ...
I’espère que vous en ferez de mesme et croirez que je demeure</p>
<div type="closer">
<p>Vostre Tresaffectioné frere et serviteur Chr. Huijgens.</p>
</div>
</body>
</text>
</TEI>
```
The TEI starts with a teiHeader section in which the metadata for the letter is specified – see the proposed metadata standard for letters for an explanation of the metadata items.

The letter text is contained in the body element of the TEI document. It consists of
- an opener section (which may contain datelines, salutations and similar phrases appearing at the start of the letter),
- the main content of the letter with each paragraph containened in a p element, and
- a closer section which has similar content as the opener section, but now appearing at the end of the letter.

Additional structure
Some documents have more structure than just paragraphs. This structure can be encoded with div and header elements. For example

```
<div type="section">
<head>Propositio 5.</head>
<p>Si il y a tant de gravitez qu'on ...</p>
</div>
<div type="section">
<head>Propositio 6.</head>
<p>Eadem metodo probatur si AB, BC, CD ...</p>
</div>
```

The type of a div element may be anything that seems appropriate. However, certain types are handled specifically by our software:
- <div type="comment"> – rendered in our web application, but ignored by our analysis software.
- <div type="summary"> – rendered in our web application, but ignored by our analysis software.

Lists
The standard TEI elements list, item and label are supported; however any attribute of these elements is ignored by our software.

```
<list>
<label>1.</label><item>Securitas in finibus Turcicis.</item>
<label>2.</label><item>Stipendia limitaneis militibus a 1<hi rend="sup">ma</hi> Iunij ipsis debita solvenda.</item>
<label>3.</label><item>Securitatis in finibus Moscoviticis ratio.</item>
...
</list>
```

Tables
The standard TEI elements table, row and cell are supported. Except for the attributes role="header" and cols="n" of the cell element all attributes are ignored.

```
<table>
```
Poems

The standard TEI elements \texttt{lg} and \texttt{l} are supported; however any attribute of these elements is ignored by our software.

\begin{verbatim}
<lg type="poem">
<l>Qui habet in manebus</l>
<l>Et iectat sub pedebus,</l>
<l>Venient temporibus</l>
<l>Quod habebunt doloribus.</l>
<l>Deinde</l>
<l>Honneur ou mort.</l>
</lg>
\end{verbatim}

Figures

The standard TEI elements \texttt{figure}, \texttt{graphic} and \texttt{figDesc} are supported; however any attribute of these elements is ignored by our software.

\begin{verbatim}
<figure>
<graphic url="huyg003oeuv06ill68.gif"/>
<figDesc>Fig. 1.</figDesc>
</figure>
\end{verbatim}

Highlighting

For highlighting words or phrases the standard TEI element can be used. For rendering the following \texttt{rend} attributes are supported:

- \texttt{<hi rend="b">} – bold
- \texttt{<hi rend="i">} – italics
- \texttt{<hi rend="sc">} – small caps
- \texttt{<hi rend="sub">} – subscript
- \texttt{<hi rend="sup">} – superscript

The analysis software ignores the \texttt{hi} element (but its content is used).

Specials

Some additional TEI encodings have special meaning for our analysis software:

- \texttt{<p lang="en">} – denotes that the content of the paragraph is in the specified language; the language overrides the global language specification.
- \texttt{<persName>} – is used for indicating person names; similarly \texttt{<orgName>} and \texttt{<placeName>} are used for indicating organization and place names.
- `<formula>` – can be used to indicate mathematical formulas or symbols which are to be excluded from analysis, although they will be rendered in the web application.

  – o o o –