Correspondence Encoding

(at the Centre for Textual Editing and Document Studies)

ctb

Ron Van den Branden
ron.vandenbranden at kantl.be

A Bit of Context

1886: ° Royal Academy of Dutch Language and Literature (KANTL)

2000: ° Centre for Textual Editing and Document Studies (CTB)

→ Aim: making available valuable text(collection)s for language and literature studies

→ Products:

- Print: scholarly editions, reading editions, studies
- Online: digital scholarly editions, language corpora, digital image collections

→ Methodology: TEI XML encoding, XSLT/XQuery processing in eXist-db-based web applications

Correspondence

From the start, CTB has had a strong interest in editions of correspondence:

- Interesting study material for lots of research purposes
- Typically physically dispersed over different archives and holding institutions
- Ideal for offering in a centralized, digital edition

Aim: Digital Archive of Letters in Flanders: expanding text-base of correspondence editions

Early on, a theoretical effort was made to investigate how digital editions of correspondence could be encoded in TEI.

A Bit of Chronology

- 2001: TEI P4 released
- 2003: DALF Guidelines for the Description and Encoding of Modern Correspondence Material, Version 1.0

DALF - Analysis of P4

1. Lack of teiHeader elements for meta-description of correspondence:
    - Lacuna: description of **unpublished manuscripts** (physical description, archival information, …)
    - Lacuna: description of **letter-specific** metadata (sender, addressee, date, place of writing)

    → TEI P4: <teiHeader> aimed at description of published texts

2. Lack of text elements for transcription of correspondence-specific features:
    - envelope
    - postscript
    - calculations (business-like correspondence)
    - pre-printed text

DALF - Approach

- Defined lots of <teiHeader> elements, inside a specific <letDesc> section inside <sourceDesc>:
  - Letter-specific metadata: sender, addressee, date, place of writing
  - Manuscript description: modeled on MASTER (Manuscript Access through Standards for Electronic Records)
- Defined some specific <text> elements:
  - Envelope: <envelope> (as sibling to <body>)
  - Postscript: <ps>
  - Calculations: <calc>: <arg>, <oper> <result>
  - Pre-printed text: <print>

DALF

→ http://ctb.kantl.be/project/dalf/dalfdoc/index.html

Defined as DTD customization of TEI P4

- defined 60 new elements (47 header + 13 text)
- Modified 8 TEI elements

A Bit of Chronology

- 2001: TEI P4 released
- 2003: DALF Guidelines for the Description and Encoding of Modern Correspondence Material, Version 1.0

- 2011: DALF edition: Van Nu en Straks Letters

- 2014: DALF edition: Streuvels and his Dutch-Speaking Publishers

- 2016: DALF edition: Streuvels and his German-Speaking Publishers

A Bite of Chronology

- 2001: TEI P4 released
- 2003: DALF Guidelines for the Description and Encoding of Modern Correspondence Material, Version 1.0

- 2011: DALF edition: Van Nu en Straks Letters

- **November 2012: † support for TEI P4**
- 2014: DALF edition: Streuvels and his Dutch-Speaking Publishers

- 2016: DALF edition: Streuvels and his German-Speaking Publishers

A Bit of Chronology

- 2001: TEI P4 released
- 2003: DALF Guidelines for the Description and Encoding of Modern Correspondence Material, Version 1.0
- 2007: TEI P5-1.0.0 released, including <msDesc>
- 2008: ° TEI Correspondence SIG
- 2011: DALF edition: Van Nu en Straks Letters
- 2012: ° TEI Corresp SIG proposal for correspondence TEI module
- November 2012: † support for TEI P
- 2014: DALF edition: Streuvels and his Dutch-Speaking Publishers
- 2015: TEI P5-2.8.0 released, including <correspDesc>
- 2016: DALF edition: Streuvels and his German-Speaking Publishers

TEI P5 and Correspondence Encoding

- //teiHeader//profileDesc/corrspDesc: letter-specific metadata:
  - <correspAction>: structured description of aspects of a correspondence action (writing, sending, reception, ...)
  - <correspContext>: links to previous / following letters in the correspondence
- //teiHeader//sourceDesc/msDesc: manuscript-specific metadata:
  - <msIdentifier>: identification of the manuscript (archival identification)
  - <msContents>: description of the content
  - <physDesc>: description of the physical realization and state of the manuscript (binding, writing hands / typefaces, accompanying materials, condition, layout, supporting materials, ...)
  - <history>: description of the history of a manuscript (origin, provenance, acquisition)
  - <additional>: extra administrative information and description of other representations
- 1 correspondence-specific text element: <postscript>

DALF 2.0 and / or TEI P5?

Phenomenon	DALF	TEI P5
Metadata	<letDesc>	<correspDesc> + <msDesc>
Postscript	<ps>	<postscript>
Description + documentation of physical layers in a letter	<layer>, <layerShift/>	Object-oriented transcription of a document in <sourceDoc>
Calculations	<calc>, <arg>, <oper>, <result>	Inclusion of MathML in TEI P5 schema (if needed at all)?
Envelope	<envelope>	?
Pre-printed text	<print>	?
Letterhead	<print type=”letheading”>	?

TEI P5 and Correspondence Encoding

- Metadata: ✓
- Text: ?

→ Little attention to encoding of specific letter structures:

  - Envelope
  - Letterheads
  - Pre-printed text

→ Little ***guidance*** on structuring different correspondence types

TEI P5 and Correspondence Encoding

Attempt: start from scratch, see how far we get with P5

- benefiting from:
  - insights gained in previous efforts (DALF)
  - current state of TEI P5 art (<msDesc> / <correspDesc> / <postscript>)
- staying as close to TEI P5 as possible
  - only customize when necessary
- contributing back:
  - propose customizations where needed
  - present encoding guidance for correspondence encoding

DALF Lesson: Correspondence Types

- What is correspondence?
  - Letter
  - Picture postcard
  - Postal card
  - Telegram
  - Aerogram
  - (email?)
  - ...
- Formal + structural variety

DALF Lesson: Correspondence Types

- What is correspondence?
- Formal + structural variety
  - Envelope:
    - physical object (regular, windowed)
    - address section on a postcard, telegram, aerogram
  - Image sections:
    - image side of postcard
    - decorations on telegrams
  - Actual text content:
    - opener
    - text paragraphs / lists / …
    - closer / postscript

Correspondence Types

- What is correspondence?
- Formal + structural variety

⇒ Distinguish correspondence “genres”, that

- share common functional building blocks
- organize them differently

Functional Blocks

- Content section
- Envelope section
- Figure section

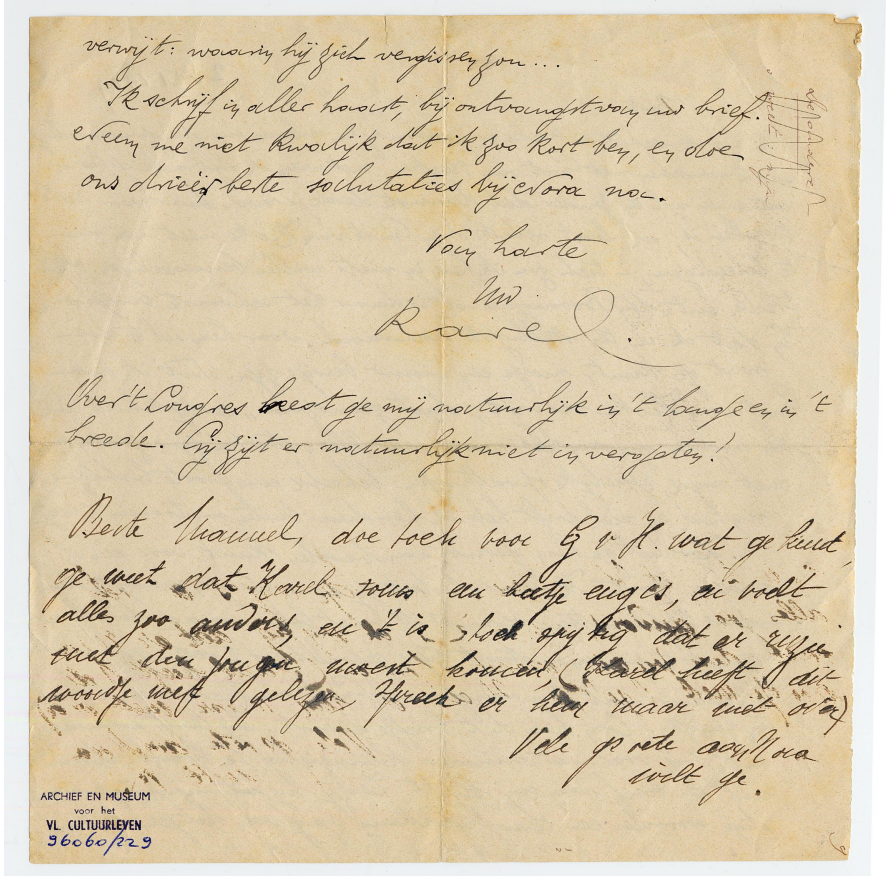
Functional Blocks (1): Content Section

Text Phenomenon	Tag
Opener stuff	<opener> containing <salute>, <dateline>, <address>, ...
Actual text	<p>, <list>, <table>
Closer stuff	<closer> containing <salute>, <dateline>, <address>
Postscript	<postscript>

possibly organized in further subdivisions:

Functional Blocks (1): Content Section

```
<div type="corresp.content">
  <!-- first subsection -->
  <div>
    <opener><!-- opener stuff --></opener>
    <p><!-- the actual text --></p>
    <closer><!-- closer stuff--></closer>
    <postscript><!-- postscript --></postscript>
  </div>
  <!-- second subsection -->
  <div>
    <p><!-- further text --></p>
    <p><!-- further text --></p>
    <closer><!-- closer stuff --></closer>
  </div>
</div>
```



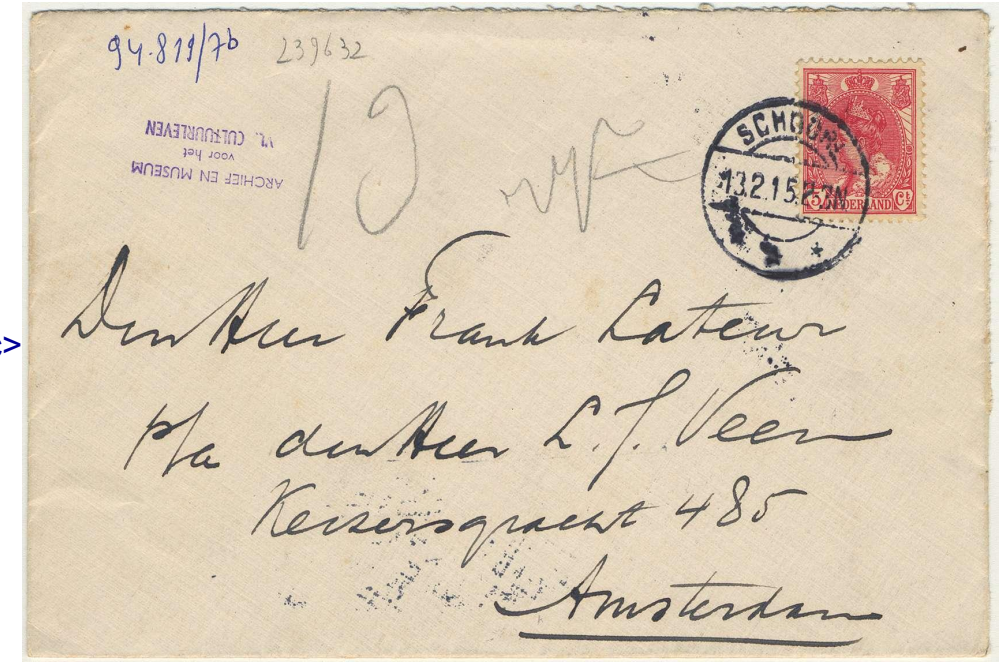
Functional Blocks (2): Envelope Section

Most common contents of an envelope will be:

Text Phenomenon	Tag
stamp	<figure type="stamp">
postmark	<figure type="postmark">
address	<ab type="address.sender"> <ab type="address.recipient">
side marker	<pb/>

Functional Blocks (2): Envelope Section

```
<div type="envelope">
  <figure type="stamp">
    <figDesc><!-- stamp description --></figDesc>
    <graphic url=""><!-- stamp facsimile --></graphic>
    <p><!-- stamp text (if any) --></p>
  </figure>
  <figure type="postmark">
    <figDesc><!-- postmark description --></figDesc>
    <graphic url=""><!-- postmark facsimile --></graphic>
    <p><!-- postmark text (if any) --></p>
  </figure>
  <ab type="address.recipient">
    <address>
      <addrLine><!-- address line(s) --></addrLine>
    </address>
  </ab>
</div>
```



Functional Blocks (3): Figure Section

Most common contents of a figure section will be:

Text Phenomenon	Tag
figure	<figure>
other text	<p> , <ab>

Functional Blocks (3): Figure Section

```
<div type="figure">
  <figure>
    <figDesc><!-- figure description --></figDesc>
    <graphic url=""><!-- figure facsimile --></graphic>
    <head type="caption" place="verso">
       <!-- caption (possibly on verso side) -->
    </head>
    <head type="copyright" place="verso">
       <!-- copyright statement (possibly on verso side) -->
    </head>
    <p><!-- figure text (if any) --></p>
  </figure>
  <!-- other text / structures in the figure section -->
</div>
```

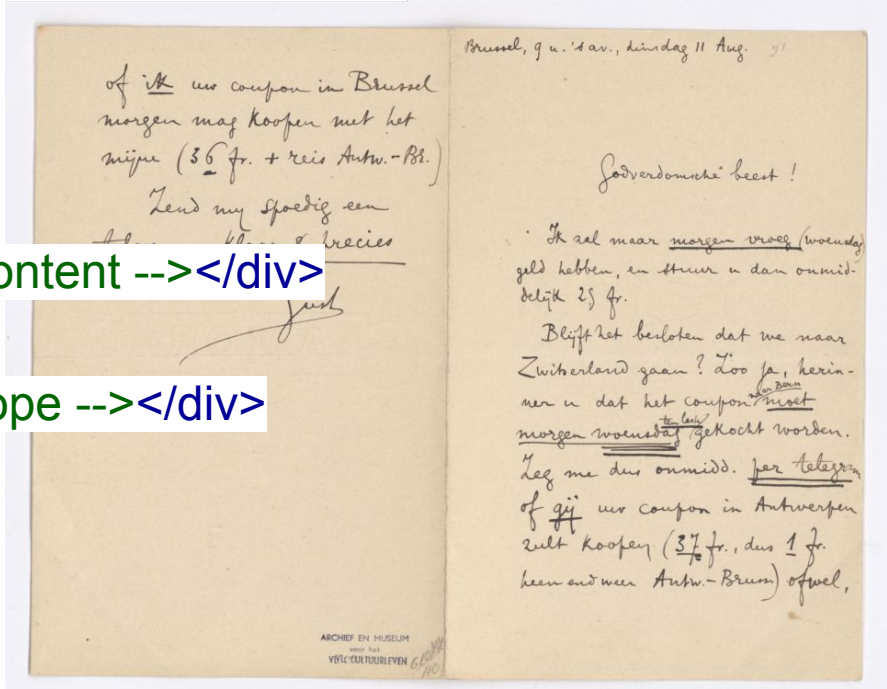
Correspondence Types

- Letter
- Postcard
- Aerogram
- Telegram

⇒ combining and structuring similar functional building blocks differently

Correspondence Types (1): Letter

- Typically mainly text
- Typically wrapped in a physical envelope




```
<text>
  <body>
    <div type="letter">
      <div type="corresp.content"><!-- letter content --></div>
    </div>
    <div type="envelope"><!-- physical envelope --></div>
  </body>
</text>
```

Correspondence Types (2): Postcard

- Typically text + figure section
- Typically with an “inline” envelope section; possibly in physical envelope

Inline envelope	External envelope
<text> <body> <!-- figure section --> <!-- pb marks 'side boundaries' --> <pb/> <!-- postcard content --> <!-- envelope section --> </body> </text>	<text> <body> <!-- physical envelope --> <!-- figure section --> <!-- pb marks 'side boundaries' --> <pb/> <!-- postcard content --> </body> </text>

Correspondence Types (2): Postcard

```
<text>
  <body>
    <div type="postcard">
      <div type="figure"><!-- figure section --></div>
      <!-- pb marks 'side boundaries' -->
      <pb/>
      <div type="corresp.content"><!-- postcard content --></div>
      <div type="envelope"><!-- envelope section --></div>
    </div>
  </body>
</text>
```

Correspondence Types (3): Aerogram

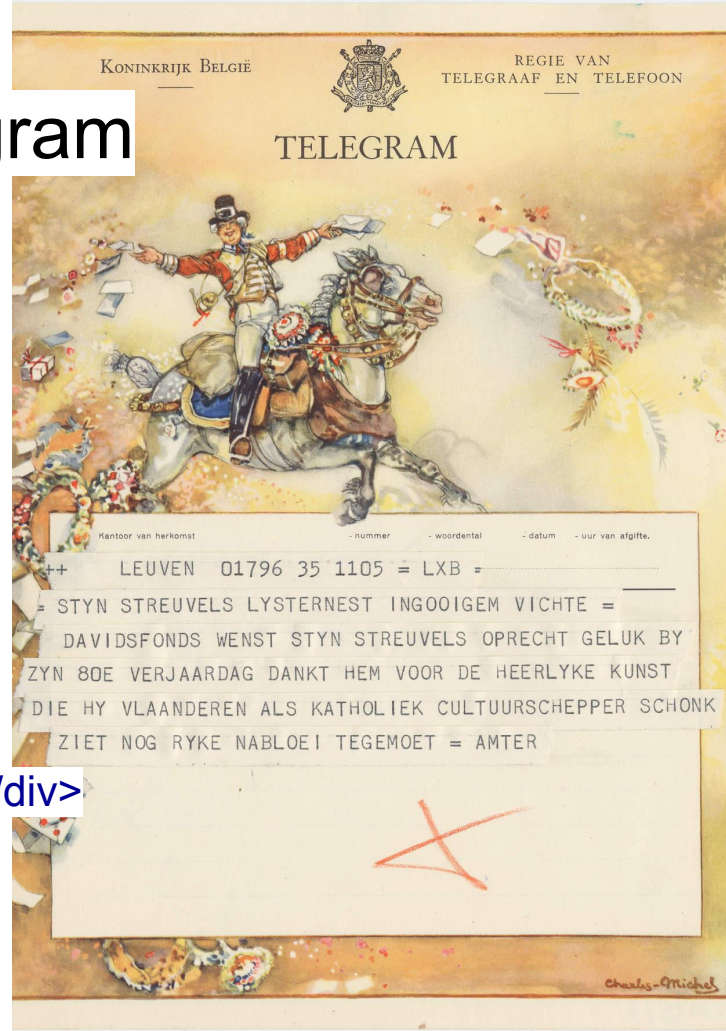
- Typically mainly text
- Dedicated address section on the letter, folded as envelope
- Possibly decorated

```
<text>
  <body>
    <div type="aerogram">
      <div type="corresp.content"><!-- aerogram content --></div>
      <div type="figure"><!-- figure section --></div>
      <div type="envelope"><!-- envelope section --></div>
    </div>
  </body>
</text>
```

Correspondence Types (4): Telegram

- Typically mainly text
- Dedicated address section
- Possibly decorated

```
<text>
  <body>
    <div type="telegram">
      <div type="envelope"><!-- envelope section --></div>
      <div type="figure"><!-- figure section --></div>
      <div type="corresp.content"><!-- telegram content --></div>
    </div>
  </body>
</text>
```



Other Common Text Features in Correspondence

- Letterhead:<fw type="letterhead">
  - flexible distribution: member of model.milestoneLike
    ⇒ may occur before <opener> / after <closer>
  - content model limited (macro.phraseSeq), but allowing for most common structures: <address> and <figure>
  - AND/OR possible to describe letterhead as part of support material inside <physDesc> section in header
- Pre-printed text: <seg type="print"> or <ab type="print">
  - block or phrase level, depending on context
  - perhaps @hand or @rend could be more appropriate than @type

Summary

- Correspondence metadata: TEI P5 provides sufficient + well-documented means
- Correspondence transcription: TEI P5 provides sufficient means (in theory); could use better documentation / guidance on the encoding of specific correspondence features
- Attempt at TEI P5 correspondence encoding proposal:
  - Identify different functional blocks
  - Compose these blocks according to the structural characteristics of the different correspondence types

What’s Next

- Prove this pudding:
  - Create sensible ODD file(s)
  - See if they (can) cover reality
- Do the eating: encode actual letters in a pilot project

  → challenge: context-specific **validation / authoring guidance**

  Vanilla TEI = impossible to restrict e.g. div/@type values for different contexts at ODD level (e.g.: prohibit div[@type='envelope'] inside div[@type='corresp.content'] or vice versa)

  ⇒ possible solutions:
  - Contextual **Schematron** constraints (e.g. only allow div[@type='corresp.content'] inside div[@type=('letter', 'postcard', 'telegram', 'aerogram')])
  - Still resort to a **customized authoring ODD** (defining e.g. <envelope> element as equivalent to), and convert ‘working’ authoring files to final TEI files?
- Share recipes