TEI Correspondence SIG's task force
Sabine Seifert (Humboldt University Berlin, project 'Berlin Intellectuals 1800-1830')
Marcel Illetschko (Literature Archive, Austrian National Library, Vienna)
Peter Stadler (University of Paderborn, Carl Maria von Weber–Collected Works)

→ Prezi: http://prezi.com/e70bkx9iqiib/towards-a-correspondence-module-in-the-tei/

Towards a correspondence module in the TEI

Table of contents

Towards a correspondence module in the TEI

The main part of this paper is about how to put a theory of correspondence into practice and into TEI elements such as a new <corrispDesc>. → **SLIDE**
As an introduction, let's review briefly what has happened up to now. → **SLIDE**

1. Up to now

1.1. Some editions

For several years now, there have been digital projects and editions dealing with all sorts of correspondences and the encoding of letters has a long tradition in the TEI. You will probably know some of the correspondence projects, all using the TEI, from this short and definitely not complete overview:

1.1.1 Completed projects → **SLIDE**

Here are some completed projects (in alphabetical order), like DALF or Van Gogh The Letters.

1.1.2 Ongoing projects → **SLIDE**

Some ongoing projects are these, for example the Escher Letter Edition or the Weber - Collected Works.

1.1.3 Projects not yet online

Some projects were developed several years ago, others still await the launch of their web page. → **SLIDE**
Some German projects in development but not yet online are for example Epistolary networks or Schleiermacher in Berlin.

1.1.4 Non-TEI-projects

Of course, not each and every letter edition works with the TEI. → **SLIDE**
Interesting projects not working with the TEI are for example the Darwin Correspondence Project or the Correspondence of Thomas Bodley.

1.2. The main questions and the Correspondence SIG's task force

As we can see, a lot of development has been done in these projects and a lot of elaborate encodings have been produced. But still, the TEI Guidelines lack a comprehensive and coherent chapter on the encoding of this sort of texts. This results in much duplicate work and in hampered interchange of these texts due to different encoding schemes. → **SLIDE**
This leads us to two main questions: First, *how* should one encode correspondence with the TEI? And second, how can correspondence editions be *linked*?

To answer these questions, the Special Interest Group Correspondence put together a task force (that is Peter, Marcel, and me) at last years' TEI conference in Rome. The goal of this task force is to develop a full schema for encoding correspondence. Let me emphasize that for now we exclusively deal with correspondence-specific *meta* data within the TEI header, not with transcriptional elements that are part of the body. It is the long-term objective to reach a comprehensive model for encoding correspondence and a also best practice model. But for now we concentrate on correspondence-specific *meta* data.

1.3 Two inspiring examples → SLIDE

1.3.1 DALF

Our work was inspired hugely by the work of DALF and the edition Carl Maria von Weber–Collected Works. Due to the lack of specific encoding guidelines for correspondence within the TEI, these two letter projects have tried to address the problem by developing their own customizations for encoding meta data and text transcription. The Belgian DALF project, the Digital Archive of Letters by Flemish Authors and Composers from the 19th & 20th century[1] is a pioneer of encoding correspondence. In 2003, the DALF Guidelines were published and just last year the update from P4 to P5 was documented in the "Preliminary P5 Proposal"[2] at which we had a close look. Before their update to P5, the Carl Maria von Weber–Collected Works had developed and published a "Recommendation for the encoding of correspondence material" in TEI P5[3] in 2011. Both these customizations laid the basis for our work.

And both projects tried to keep the changes to the existing TEI elements as minimal as possible. They defined a special wrapper element for correspondence meta data within the source description to have these in one single place and, so to say, keep the rest of the TEI header 'clean' of these additions. These two basic assumptions were also the starting point for our work. The depth of the structure, however, and the introduced child elements of the wrapper element differ in both DALF and WEGA .

→ SLIDE: encoding example of Dalf in P5

In the DALF project, the wrapper element is called <dalf:letDesc> for "Letter Description" and its main child element is <dalf:letHeading> which contains - and that is DALF's definition - "a structured description of bibliographical information of a letter". These consist of the four key data

- author (please note the use of the term 'author' not 'sender'!) in the element <dalf:letAuthor>,
- addressee in the element <dalf:letAddressee>,
- place in the element <dalf:letPlace>
- and date in the element <dalf:letDate> of the letter.

[1] http://www.kantl.be/ctb/project/dalf/
[2] http://ctb.kantl.be/project/dalf/P5/DALF_P5-p0.1.zip
[3] http://wiki.tei-c.org/index.php/File:CorrespondenceProposal.zip

Additional information, restricted to the formal classification of the letter, the occurrence of envelopes or illustrations, can be provided in other new child elements of <dalf:letDesc> (<dalf:type>, <dalf:envOcc>, <dalf:figOcc>).
(As you can see in this example, <dalf:letHeading> was made repeatable for several letters on one sheet of paper.)

1.3.2 WeGA → SLIDE: encoding example of WeGA

The WeGA offers a different structure and groups the information directly within the wrapper element called <wega:correspDesc>. There is also a greater variety of meta data included:

- sender (not author!) and addressee in <wega:sender> and <wega:addressee>,
- place and date where and when the letter was written as well as received in <wega:placeSender> and <wega:placeAddressee>,
- the incipit (<incipit>), and the
- position within the whole thread of correspondence if known, in <wega:context>.

The SIG Correspondence followed these two customizations not only in the basic assumption that the respective new wrapper element covers *correspondence-specific* meta data *exclusively* while all manuscript or rather source describing information including provenance, physical appearance, history, etc. still belong to <msDesc>. And as things like (postage) stamps, seals, water marks, etc. are not necessarily characteristics of correspondence but of manuscripts in general, they are not part of the proposed <ct:correspDesc> and also best encoded within <msDesc>. This also means that a *complete* description of a piece of correspondence is not provided by the new element <correspDesc> *alone* (which gives, as I said before, *selected* key data) but in conjunction with the manuscript description in <msDesc>.

We defined what meta data for a piece of correspondence we considered essential - and you will see that they in some aspects differ from the concepts of DALF and WEGA - and included them in our new wrapper element <ct:correspDesc>. And while doing this, we tried to only tentatively modify the TEI standard by merely adding a few elements, not changing already existing ones.
So much for this, let's now dive into the theory.

2 Theory → SLIDE

2.1 Theory of correspondence in general → SLIDE

The first thing was to settle on what we understand as a letter or any other piece of correspondence (like a postcard or a telegram) and we had a look at the theory of correspondence in general of which we now can only present some brief thoughts.[4]

2.1.1 A letter (or any other piece of correspondence) as an object

Usually, letters are considered as having a double nature as object *and* event. The aspect of objecthood predominantly, but not exclusively, is tied to their materiality. The facets of this

[4] The term "letter" in the following considerations means "piece of correspondence".

materiality are manifold and need not to be discussed here because the TEI Guidelines have elaborated instruments for describing them, especially within the manuscript description <msDesc>.

2.1.2 A letter (or any other piece of correspondence) as an event

More important than questions of objecthood and materiality for the development of our concept of *correspondence* was the aspect of eventness ("eventness" in a wide ontological sense). The Glasgow Classical Philologist Jan Stenger defines in his essay *The late antique letter as an event* "the temporal sequence of the Before and After with something happening in the middle of those" as a central element of epistolary eventness. "More important than the question of what happens is the fact that something happens, i.e. the mere carrying out."[5] Distancing oneself from the content and focussing on the event corresponds with the importance of non-object-specific meta data.

Already in late antiquity, correspondence did not mean merely written messages. Very often "the surrounding circumstances - sent works, presents, works of art, other oral messages - show that the letter is embedded within a whole ensemble of communication media that made correspondence to a multimedia process." (Stenger, p. 32) which reduces the importance of the letter in itself to just being a minor issue.

Hence, next to the writer or sender and to the addressee the method of transmission, possible involved participants, accompanying material, non-linguistic "communication objects" etc. are integral parts of the communication act. Only in the course of the 17th and 18th centuries, aspects concerning the writing began to be emphasized.

To sum it all up, the basic assumption is the understanding of correspondence in general as an event - connected to a document and/or an (electronical) text. What kind of 'event'? We tend to say: an act of communication.

2.1.3 A letter (or any other piece of correspondence) as an act of communication

This idea of an act of communication (sender, transmission, receiver) is not yet included in the TEI guidelines. But the topos of correspondence being a "half dialogue", a "conversation amongst absentees" is very often found in theoretical articles about epistolography. Janet Gurkin Altman for example speaks about "the letter's function as a connector between two distant points, as a bridge between sender and receiver".[6]

Although, of course, not every letter follows this basic pattern, it nevertheless seems reasonable to concentrate on these points:

- WHO writes - TO WHOM - WHERE FROM - WHERE TO - ANSWERING WHAT - AT WHAT DATE? - HOW is the message TRANSMITTED and (very often unknown) - AT WHAT DATE is it being received?

We therefore tried to implement the basic principles of the communication model by Shannon and Weaver into our concept (sender - channel/message - receiver) → **SLIDE** Hence, we need to provide information about persons (or organisations) as sender, receiver or messenger.

---

[5] Stenger, J. (2010) Reich an Worten, arm an Inhalt? Der spätantike Brief als Ereignis. In: Wiethölter, W. and Bohnenkamp, A. (eds.) Der Brief – Ereignis & Objekt: Frankfurter Tagung. Stroemfeld, Frankfurt, Germany, pp. 26-41, here p. 30.

[6] Altman, J. G. (1982) Epistolarity. Approaches to a Form, Columbus, p. 13.

Adding to this, we need to provide the respective dates and places. Additionally, we need a mechanism for pointing at (or: referencing) preceding and subsequent messages. → **SLIDE**

2.2. (Our) Theory of correspondence in terms of the TEI

2.2.1 Persons

An act of communication has a sender of a message, who is not necessarily identical to the author of the text of the message.

2.2.1.1 Sender or Author?

In the TEI-guidelines, <author> in a bibliographic reference contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority. Part of this definition are specific concepts of "authorship" and "work". But of course it is also possible to send a Shakespeare poem instead of writing an original love letter, it is possible for a writer to send a publishing contract in order to discuss it with a friend, it is possible to send a cut ear in order to threaten an enemy instead of writing a text at all. Still correspondence, communication happens. That's why we want to differentiate between author and sender (if author and sender are the same person - which in most letters will be the case - we can use a pointer to another part of the TEI-header to indicate this: <sender sameAs="#author"/>).

2.2.1.2 Addressee/Receiver

Furthermore, an act of communication generally has no unknown audience/unspecified recipient but one or more (intended, perhaps fictional) addressees, "spoken" to by the piece of correspondence.

2.2.2 Dates

A piece of correspondence generally has a certain time of sending—even though it may be unknown or indeterminate. And of course one could argue that in some cases also the date of receiving was documented and should be encoded as well.[7]

2.2.3 Places

Similarly, a piece of correspondence was generally sent from a particular location, which may or may not be the same as the place of writing, to a specific location. Very often these bits of information are evident in the text of the message, in an address line, postmark, electronically generated sending meta data, etc.

---

[7] "Briefedition im digitalen Zeitalter", S. 100: "Im Rahmen seiner Konzeption einer „Ökonomie des Briefs" gelangt Strobel somit zur Forderung nach editorischer Dokumentation dieser „Spuren der brieflichen Ökonomie". Nicht nur die „Schreib-", sondern auch die „Empfangsszenen" sollten vom Editor rekonstruiert werden: „Erst die Rekonstruktion von Schreib- und Empfangsszenen, von Postierung und Archivierung, kurz: der wenigstens teilweise materialiter sicht- bar gemachten Spuren brieflicher Ökonomie, belehrt den nachgeborenen Leser, dass Briefe viel mehr sind als Dialoge zwischen Abwesenden, viel mehr als nur Texte."[30]

2.2.4 Transmission

A piece of correspondence generally has a process or a medium and/or one or more executor(s) of transfer (messenger, carrier, postman, carriage, fax machine, internet, etc.).

2.2.5 Context

Very often a particular message is sent in order to trigger an answer (in form of one or more pieces of correspondence). A single piece of correspondence therefore is normally not a secluded entity but a (written) act of communication in a communication continuum (Bohnenkamp/Richter 2013, p.4), in which it is defined by its relative position between messages sent "before" and "after". This information should be gathered in one place. E.g. if the delivery of letters overlaps – and the answer to a letter by person A is not person B's letter of the next day but maybe the one three weeks later because it got stuck. Dates could be helpful – but finding out the correct chronology is part of the editing process, and they should be given in a specific element.

3. Putting Theory into Practice → SLIDE

3.1 <correspDesc>

So, how to encode correspondence with TEI?!

The direct implementation of our concept into elements, in which one category of information becomes one element, could look like this: → **SLIDE <correspDesc>: explain single elements**

To differentiate between the place of writing and the place of sending one could use attributes, e.g.: <placeSender @type='writing'>.

3.2 Ongoing Discussions, e.g. sender and receiver → SLIDE

3.2.1 Element Names

'Sender' as a new element or as an attribute? → **SLIDE**

Even if one recognizes the necessity of a 'sender' in the context of encoding correspondence one may not introduce a new element. Alternatively the code could look like this:

The fundamental difference is that our proposed elements are specifically designed for meta data which are usually more constrained than those elements that can occur within the <body>. E.g. <editor> or <author> are very similar in this respect to our <ct:sender> - being somehow syntactic sugar for <persName role=„editor|author|sender"> - but with only a subset of attribute classes and a limited usage (i.e. the contexts in which those elements may occur). We are not trying to say „Because we have <editor> we want <sender>" but want to make the point that <ct:sender> is not like <persName>, rather it's more like <editor|author|principal|…>. Having a clear divide between meta data elements and text elements is IMHO a very good thing as other discussions have already shown.

<participant> instead of <sender>, <addressee> (and part of <transmission>) → **SLIDE**

Some contributors on the TEI SIG list argued, that it would be more appropriate to understand correspondence as involvement of various persons, resp. functions, rather than just sender, receiver (and transmitter → <transmission>) and therefore one should use the term <participant>.One could have @role=author, sender, signer, co-signer and many other variants - a semi-open list with some suggested values.

In the <sender> or <addressee> (or what have you) element we mainly focus on the 'two sides' of the communicative act and don't care about the other parties maybe involved in the process. Very often the <transmission> element will be more suitable for information like this. Another option could be to use <msDesc> to cover people/functions like typists etc.

3.2.2 Element Structure → SLIDE

Another argument was made to wrap all information about the sending (or receiving) side in one wrapper element; <sender> and <addressee> respectively. Names, dates and places are encoded using standard (existing) TEI elements.

Our impression is that the semantic structure of our proposal is more accurate and therefore very useful for interchange. But at present we are still evaluating those different approaches and welcome any feedback.

3.3 CIF - Correspondence Interchange Format → SLIDE

A crucial point in the development of our proposal has always been the wish for facilitating interchange (or even interoperability) of the meta data of this sort of texts because there is a general and growing demand of correspondence projects for interchange and linked data facilities. To provide for this, we developed a model of the <ct:correspDesc> element in concentrated form that heavily relies on authority files and standard formats. An example of such an abbreviated form might look like this (by Patrick Sahle): → **SLIDE: encoding example CIF**

As you can see, authority files (GND) are used for persons and places, and standardizations like ISO-codes for the encoding of dates. Including such standards again poses new problems when there are several authority files for one information, or when there are none.
This here is just a preliminary example based on <correspDesc> as it looks now. When the structure and implementation of <correspDesc> is finalized, then this short format will be adapted.
Let us just mention in passing, that along with developing this *Correspondence Interchange Format* we also strive for an accordant web service to demonstrate the envisioned benefits.→ **SLIDE** A first result, set up by Stefan Dumont, is "correspSearch". This web service makes the correspondence-specific meta data of currently three German-language correspondence editions searchable with one query and gives an idea of what becomes possible when correspondence projects can be linked.

4. Links & Thank you → **SLIDE**

We hope we could give you some insights into our ongoing work on “correspDesc”.
In everything we did, we tried to be as transparent as possible and put everything online, at the TEI wiki page of the SIG and on GitHub. Here you can find the proposal, the RNG schema, examples and other stuff. We are very thankful for *any* feedback! Right after this session and the coffee break, there is the Correspondence SIG meeting in the Halgate Room to which we warmly welcome you. Thank you!