Summer School "Digitale Edition"

Erschließung geisteswissenschaftlicher Quellen
mit digitalen Methoden

5.-9. September 2016

Zentrum für Informationsmodellierung
Austrian Centre for Digital Humanities
Elisabethstraße 59/III, SR 81.31

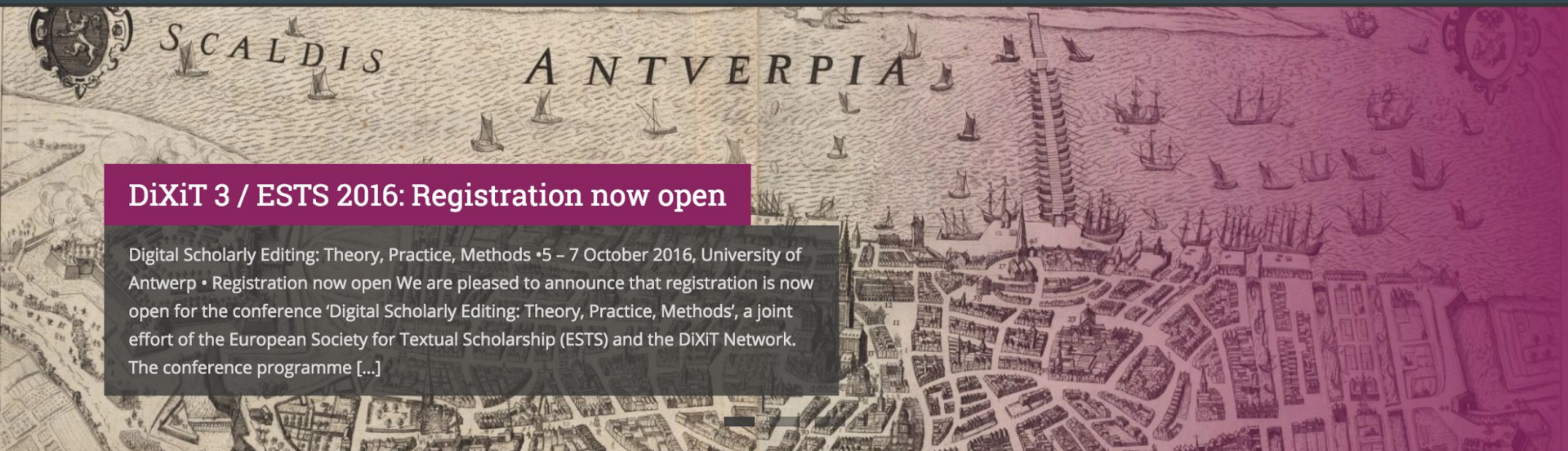# TEI-Vertiefung 2
Transkription, Text und Bild

• • •

Wout Dillen
Högskolan i Borås

@WoutDLN

ownCloud:
Dienstag/Dillen_Transkription_Tekst_und_Bild

01-transcription

DiXiT 3 / ESTS 2016: Registration now open

Digital Scholarly Editing: Theory, Practice, Methods •5 – 7 October 2016, University of Antwerp • Registration now open We are pleased to announce that registration is now open for the conference 'Digital Scholarly Editing: Theory, Practice, Methods', a joint effort of the European Society for Textual Scholarship (ESTS) and the DiXiT Network. The conference programme [...]

http://uahost.uantwerpen.be/estsdixit2016/

Home

Digital Scholarly Editing:
Theory, Practice, Methods

This thirteenth annual conference of the European Society for Textual Scholarship (ESTS), is organized in conjunction with the Digital Scholarly Editing Initial Training Network (DiXiT) and hosted by the Centre for Manuscript Genetics at the University of Antwerp, Belgium.

5-7 October 2016

Keynote speakers:

Kathryn Sutherland and Paul Eggert

Guests of honour:

Hans Walter Gabler and Peter Shillingsburg

As digital publications are reaching a stage of maturity and scholarly editors are becoming increasingly aware of the seemingly endless possibilities of hybrid or fully Digital Scholarly Editions, the impact of the digital medium on the field of Textual Criticism has become undeniable. As a result of this 'digital turn', textual scholars are now faced with new

CONTACT

Dirk Van Hulle

Wout Dillen

LINKS

Centre for Manuscript Genetics (CMG)

European Society for Textual Scholarship (ESTS)

DiXiT

TWITTER

@CMG_UA

@ESTS_eu

@DiXiT_eu

LOCAL ORGANIZERS

Bursaries

Thanks to the generous sponsorship of Digital Humanities Flanders (DHuF) and DiXiT, we will be able to offer five competitive bursaries. These bursaries take the form of a **reimbursement fee** that can be used to help cover travel and accommodation expenses.

Eligibility

The bursaries are intended to help early career scholars defray their costs of attending the conference. More specifically, we especially invite applications from undergraduate and master students, PhD students and candidates, and postdoctoral researchers who were awarded their PhD less than five years before the opening date of the conference.

Terms

Successful applicants will be awarded a bursary of **€250** to help defray expenses incurred in attending the conference. The bursary can be used to cover **registration**, **travel**, and **accommodation** expenses paid by the recipient. To be reimbursed for their expenses, recipients will be required to produce all relevant original bills, receipts, etc. to be attached to a signed expense form provided by the host institution. Recipients will be reimbursed as soon as possible after the

CONTACT

Dirk Van Hulle

Wout Dillen

LINKS

Centre for Manuscript Genetics (CMG)

European Society for Textual Scholarship (ESTS)

DiXiT

What we have learnt:

- Basics of XML structure
- Basic TEI vocabulary and transcriptions: <teiHeader>, <text>, core elements

What we have learnt:

- Basic TEI structure and transcriptions: <teiHeader>, <text>, core elements

What's up next:

- More tags
- Link transcription to documents

More Tags

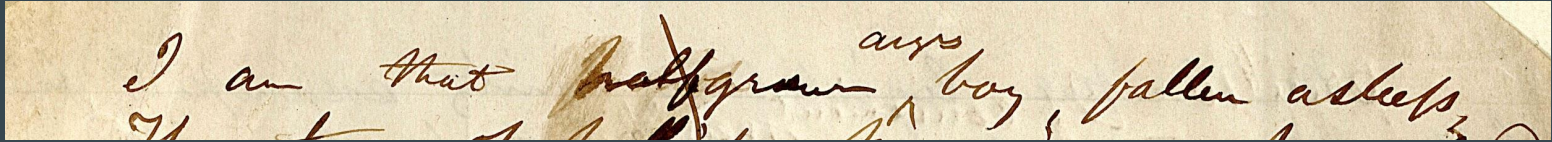focus: handwritten materials

Disclaimer:

- This is the TEI we're talking about. There is not ONE WAY of encoding things. So this is not THE WAY of transcribing documents: it is A WAY. You can do it differently if you want. Just:
  - Be consistent
  - Document everything

- This is a selection of tags, with a selection of attributes. Most of these tags have many more attributes, and many more uses.
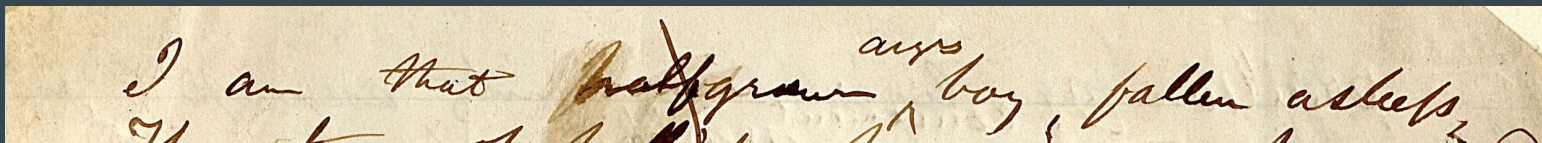
I am that half-grown^aays boy, fallen asleep,

‹unclear›

‹unclear› contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

‹unclear›

‹unclear› contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

I am that ‹unclear›halfgrown‹/unclear› angry boy, fallen asleep

\<unclear\>

\<unclear\> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.
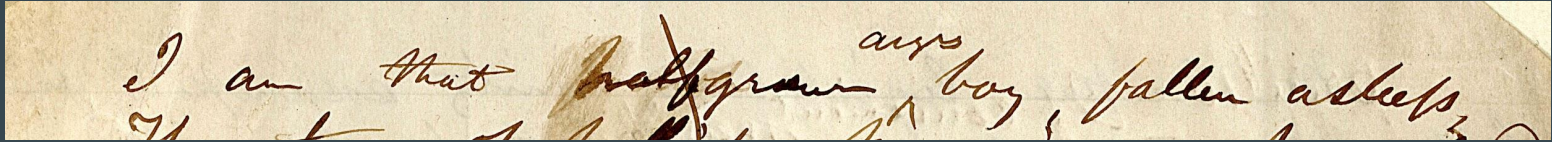


```
I am that <unclear>halfgrown</unclear> angry boy, fallen asleep
```
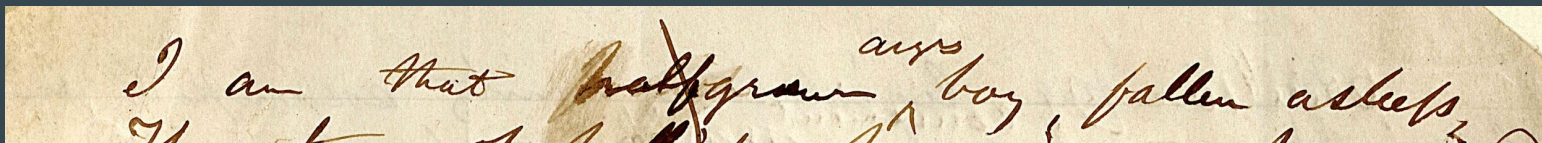
Optional: @reason (why illegible? Smudge, ink stain, crossed out, bad handwriting, …)

```
I am that <unclear reason="smudge">halfgrown</unclear> angry boy, fallen asleep
```
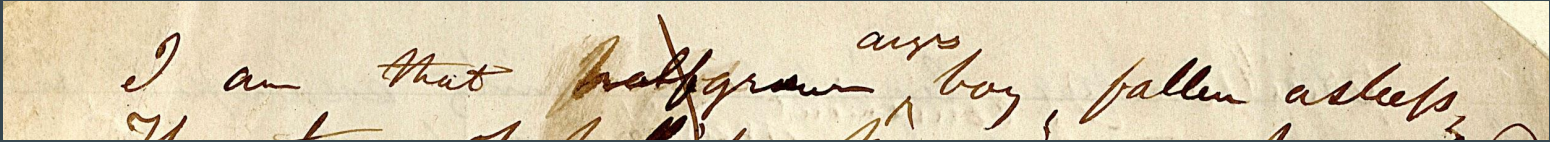
<gap>

<gap> indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible.

<gap>

<gap> indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible.

I am that <gap unit="words" quantity="1"/> angry boy, fallen asleep

‹gap›

‹gap› indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible.
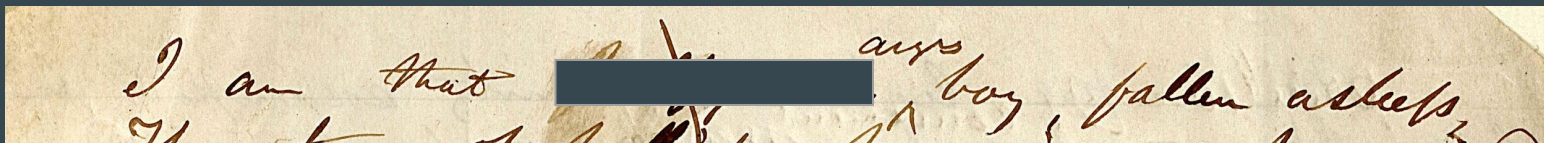


```
I am that <gap unit="words" quantity="1"/> angry boy, fallen asleep
```
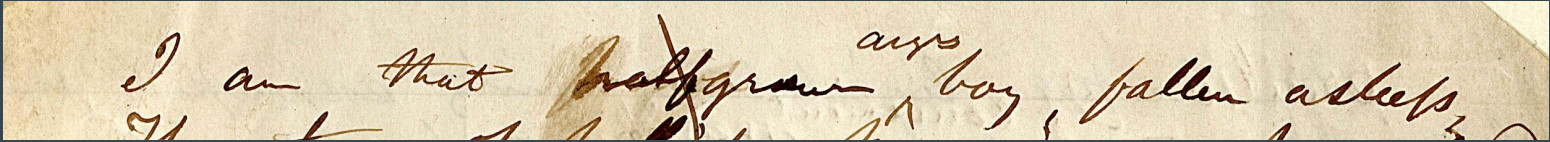
Optional: @reason (why omitted? Illegible, sampling, irrelevant, cancelled, …)

```
I am that <gap unit="words" quantity="1" reason="illegible"/> angry boy, fallen asleep
```

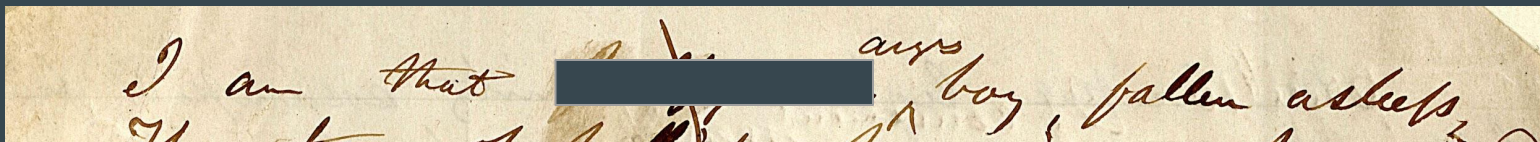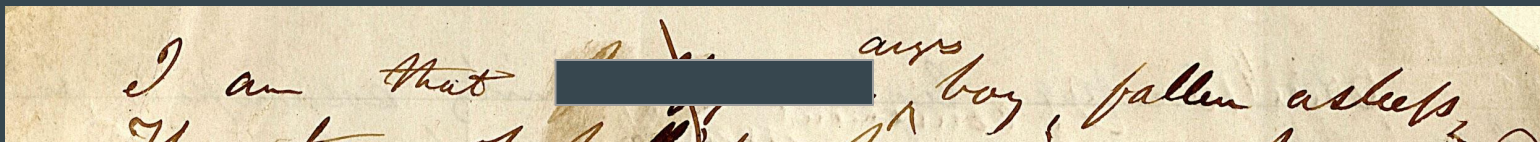# <damage>

<damage> contains an area of damage to the text witness.

<damage>

<damage> contains an area of damage to the text witness.

I am that `<damage><gap unit="words" quantity="1"/></damage>` angry boy, fallen asleep

<damage>

<damage> contains an area of damage to the text witness.

I am that <damage><gap unit="words" quantity="1"/></damage> angry boy, fallen asleep

Optional: @agent (cause of damage? Fire, water, ink, cut, tear, …)

I am that <damage agent="cut"><gap unit="words" quantity="1"/></damage> angry boy, fallen asleep
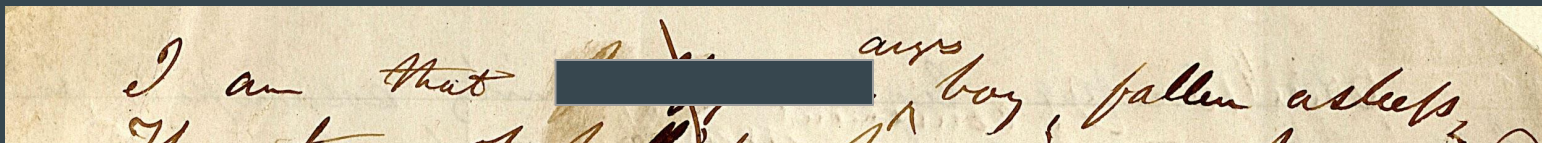
<supplied>

<supplied> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.

<supplied>

<supplied> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.


```
I am that <damage><supplied>halfgrown</supplied></damage> angry boy, fallen asleep
```

‹supplied›

‹supplied› signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.


```
I am that <damage><supplied>halfgrown</supplied></damage> angry boy, fallen asleep
```

Optional: @reason (Why supplied? Damage, omission, faded ink, lost folio, …)

Summary

<unclear>

> *"There's something there, and I think I can read it, but I'm not sure."*

<gap>

> *"There's something there, but I've omitted it (e.g. because I can't read it at all)."*

<damage>

> *"The document is damaged, which makes it difficult to determine the original text (or if there even is any)."*
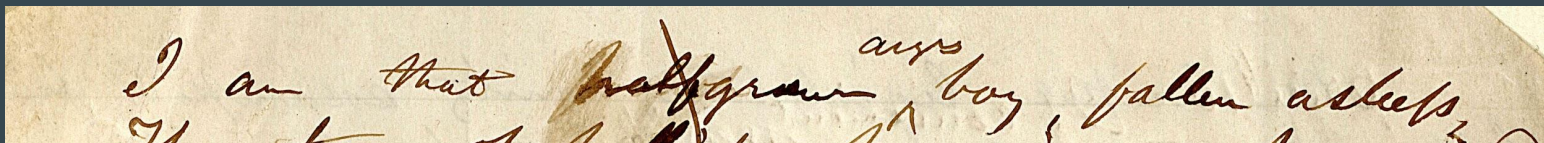
<supplied>

> *"As an editor, I've decided to add this piece of text, that isn't actually on the document."*
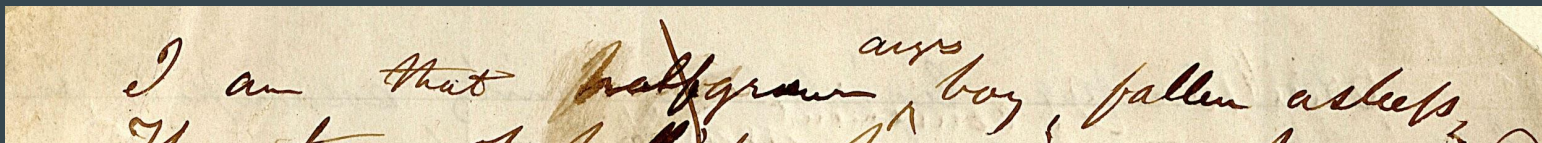
<del>

<del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector.

<del>

<del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector.

I am that `<del>halfgrown</del>` angry boy, fallen asleep

‹del›

‹del› (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector.
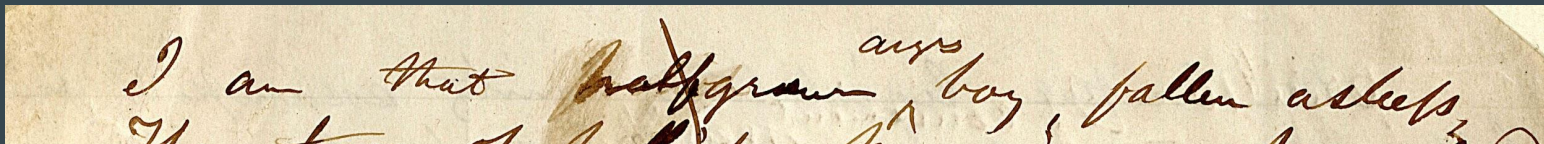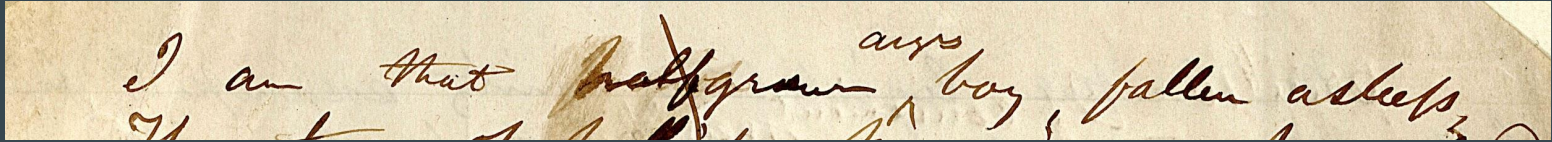


```
I am that <del>halfgrown</del> angry boy, fallen asleep
```

Optional: @type (what kind of deletion? Crossed out, overwritten, underlined, underdot, erased …)

```
I am that <del type="crossedOut">halfgrown</del> angry boy, fallen asleep
```
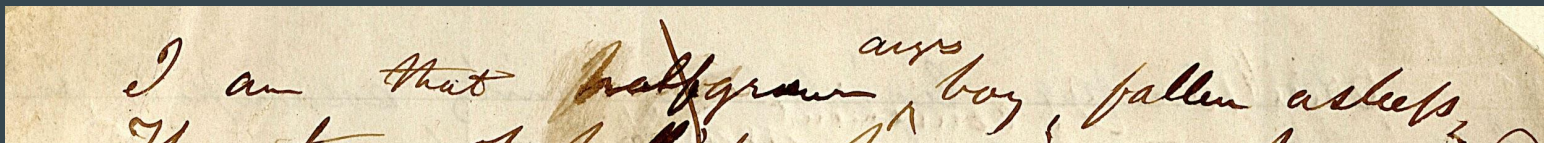
<add>

<add> (addition) contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.
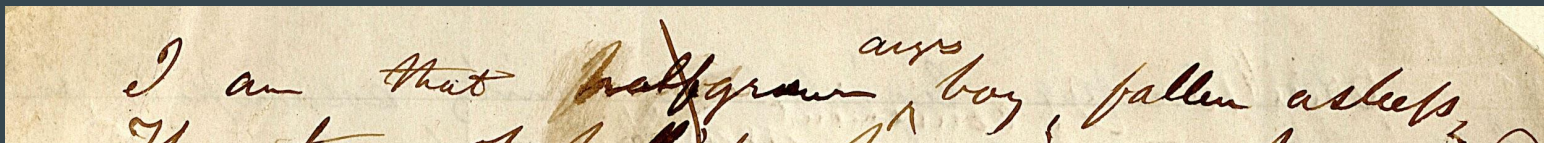
‹add›

‹add› (addition) contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.

I am that halfgrown ‹add›angry‹/add› boy, fallen asleep

<add>

<add> (addition) contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.

I am that halfgrown <add>angry</add> boy, fallen asleep

Optional: @place (where inserted? Supralinear, sublinear, margin, facing leaf, …)

I am that halfgrown <add place="supralinear">angry</add> boy, fallen asleep
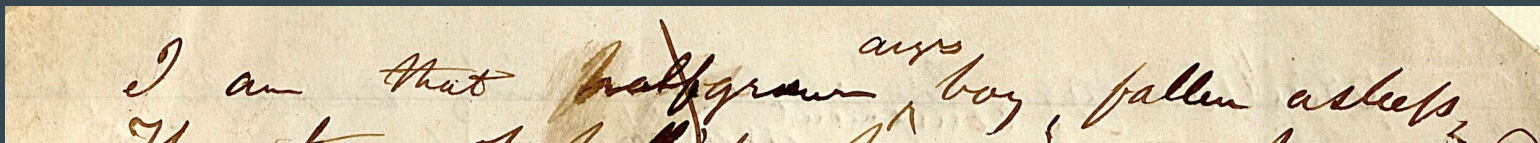
<subst>

<subst> (substitution) groups one or more deletions with one or more additions when the combination is to be regarded as a single intervention in the text.

<subst>
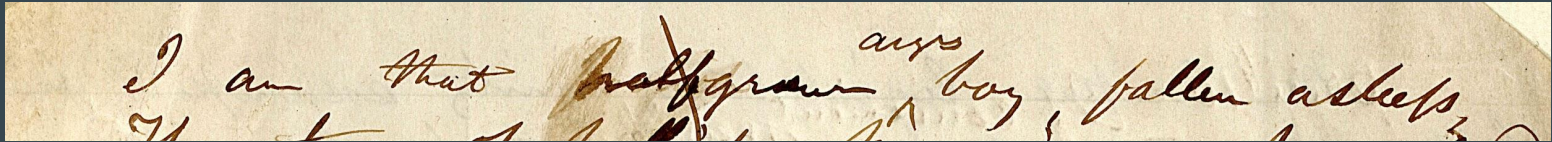
<subst> (substitution) groups one or more deletions with one or more additions when the combination is to be regarded as a single intervention in the text.


```
I am that
    <subst>
        <del type="crossedOut">halfgrown</del>
        <add place="supralinear">angry</add>
    </subst>
boy, fallen asleep
```

<restore>

<restore> indicates restoration of text to an earlier state by cancellation of an editorial or authorial marking or instruction.
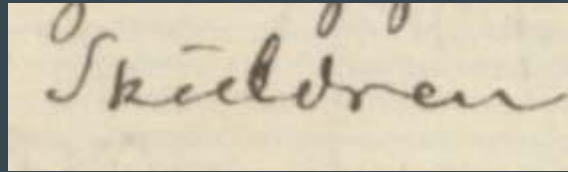
<restore>

<restore> indicates restoration of text to an earlier state by cancellation of an editorial or authorial marking or instruction.

Where now? When now? Who now? <restore>~~Unquestioning~~<restore>.

‹retrace›

‹retrace› contains a sequence of writing which has been retraced, for example by over-inking, to clarify or fix it.

<retrace>

<retrace> contains a sequence of writing which has been retraced, for example by over-inking, to clarify or fix it.
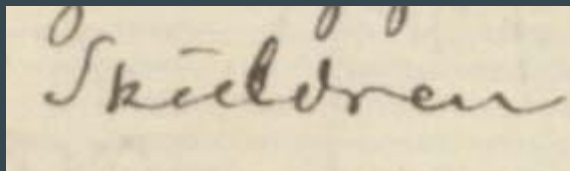
Sku<retrace>l</retrace>dren

Summary

<del> (or <add>)

"This text has been deleted (or added)."

<subst>

"As an editor, I interpret this combination of addition(s) and deletion(s) as a substitution."

<restore>

"The original reading has been restored (e.g. by deleting a deletion)."

<retrace>

"The same thing was written on top of itself (e.g. to make it clearer)"

Spans

Spans

<delSpan> (deleted span of text) marks the beginning of a longer sequence of text deleted, marked as deleted, or otherwise signaled as superfluous or spurious by an author, scribe, annotator, or corrector.

<addSpan> (added span of text) marks the beginning of a longer sequence of text added by an author, scribe, annotator or corrector (see also add).

<damageSpan> (damaged span of text) marks the beginning of a longer sequence of text which is damaged in some way but still legible.

Spans

= Solution for problem of overlapping hierarchies.

    => start & end point = 'empty element' (like <lb/>, <pb/>, <cb/>, <gap/>, …)

START: <delSpan/> (or: <addSpan/>, or <damageSpan/>)

STOP:

Spans

```
<delSpan/>

        <lg> I am that halfgrown <add>angry</add> boy, fallen asleep> [...]</lg>

        <lg> text of line group 2</lg>

        <lg> text of line group 3</lg>

<anchor/>
```

Spans

```
<delSpan/>

    <lg> I am that halfgrown <add>angry</add> boy, fallen asleep> [...]</lg>

    <lg> text of line group 2</lg>

    <lg> text of line group 3</lg>

<anchor/>
```

BUT: we need to link the two together.

=> <anchor/> gets @xml:id -- <delSpan> gets @spanTo

Spans

Linking with @xml:id and @spanTo

    Step 1: identify the <anchor/> with @xml:id

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

Formal requirements:

- No fancy characters: letters (a-z; A-Z), numbers (0-9), underscore (_), dash (-)
- Starts with a letter or underscore
- Has to be unique (no 2 identical xml:id's in 1 document)
- Length may vary

(oXygen will tell you if you're doing it wrong)

Spans

Linking with @xml:id and @spanTo

Step 1: identify the <anchor/> with @xml:id

```
<delSpan/>

    <lg> I am that halfgrown <add>angry</add> boy, fallen asleep> [...]</lg>

    <lg> text of line group 2</lg>

    <lg> text of line group 3</lg>

<anchor xml:id="anchor1"/>
```

Spans

Linking with @xml:id and @spanTo

     Step 2: Refer to <anchor/> in @spanTo

Spans

Linking with @xml:id and @spanTo

    Step 2: Refer to <anchor/> in @spanTo

Value of @spanTo = referenced xml:id *preceded by a hashtag (#) !*

```
<delSpan spanTo="#anchor1"/>
```

Spans

Linking with @xml:id and @spanTo

```
<delSpan spanTo="#anchor1"/>

    <lg> I am that halfgrown <add>angry</add> boy, fallen asleep> [...]</lg>

    <lg> text of line group 2</lg>

    <lg> text of line group 3</lg>

<anchor xml:id="anchor1"/>
```

Spans

Empty elements = good way of avoiding overlapping hierarchies

```
<lg> I am that <delSpan spanTo="#anchor1"/>halfgrown <add>angry</add> boy, fallen asleep> [...]</lg>

<lg> text of line group 2</lg>

<lg> text of <anchor xml:id="anchor1"/>line group 3</lg>
```

Exercise

Work in pairs to decypher and encode the entire passage, using the tags that we've seen so far.

I am that *halfgrown* boy, fallen asleep,
The tears of foolish passion yet undried
upon my cheeks,

years with all their events pass for me,
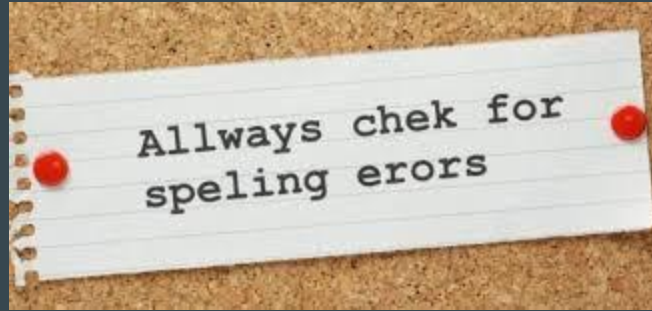Some are spent in travel ~ some in the ~~~~
usual hunt *after* fortune.

I pass through *the* travels and ~~fortunes~~ of *thirty*
years, and become old,
Each in its due order comes and goes,
And then a message for me comes.

The

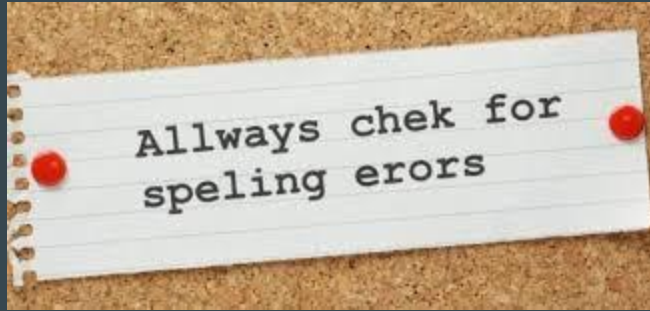Entered ~ yes

<sic>

<sic> (Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.
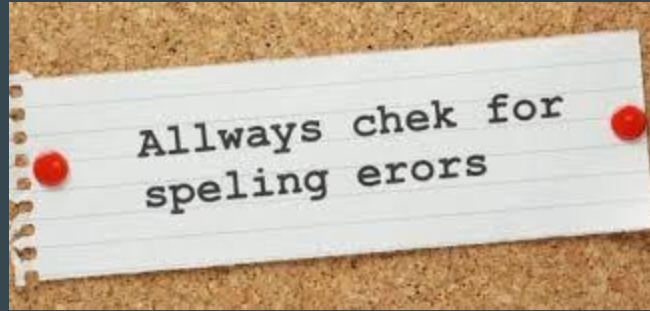
<sic>

<sic> (Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

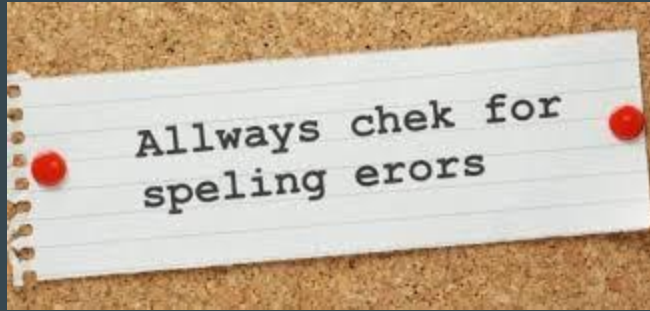`<sic>Allways</sic> <sic>chek</sic> for <sic>speling</sic> <sic>erors</sic>`

\<corr\>

\<corr\> (correction) contains the correct form of a passage apparently erroneous in the copy text.

<corr>

<corr> (correction) contains the correct form of a passage apparently erroneous in the copy text.
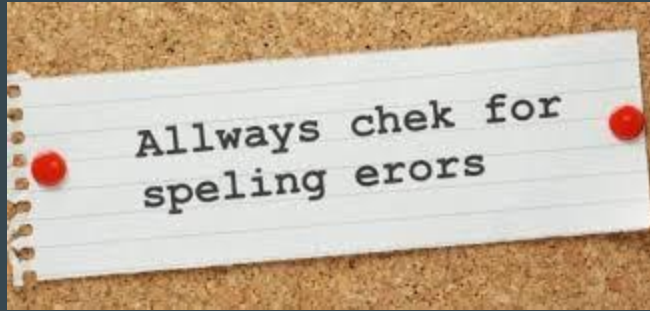


```
<corr>Always</corr> <corr>check</corr> for <corr>spelling</corr> <corr>errors</corr>
```
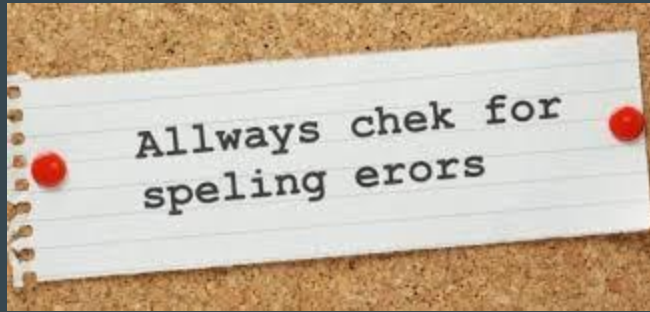
<choice> – corrections

<choice> groups a number of alternative encodings for the same point in a text.

\<choice> – corrections

\<choice> groups a number of alternative encodings for the same point in a text.


```
<choice>
    <sic>Allways</sic>
    <corr>Always</corr>
</choice>
```

<choice> - corrections

```
<choice>
      <sic>Allways</sic>
      <corr>Always</corr>
</choice>
<choice>
      <sic>chek</sic>
      <corr>check</corr>
</choice>
for
<choice>
      <sic>speling</sic>
      <corr>spelling</corr>
</choice>
<choice>
      <sic>erors</sic>
      <corr>errors</corr>
</choice>
```

<choice> – abbreviations

We can also use <choice> for transcribing abbreviations their expansions:

<abbr> (abbreviation) contains an abbreviation of any sort.

<expan> (expansion) contains the expansion of an abbreviation.

\<choice\> – abbreviations

We can also use <choice> for transcribing abbreviations their expansions:

<abbr> (abbreviation) contains an abbreviation of any sort.

<expan> (expansion) contains the expansion of an abbreviation.


```
<choice>
     <abbr>i|d|e</abbr>
     <expan>Institut für Dokumentologie und Editorik</expan>
</choice>
```

<choice> – regularizations

Or for regularizations (e.g. old spelling):

<orig> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

<reg> (regularization) contains a reading which has been regularized or normalized in some sense.

"how godly a dede it is to overthrowe so wicked a race the world may judge: for my part I thinke there canot be a greater sacryfice to God."

Exercise

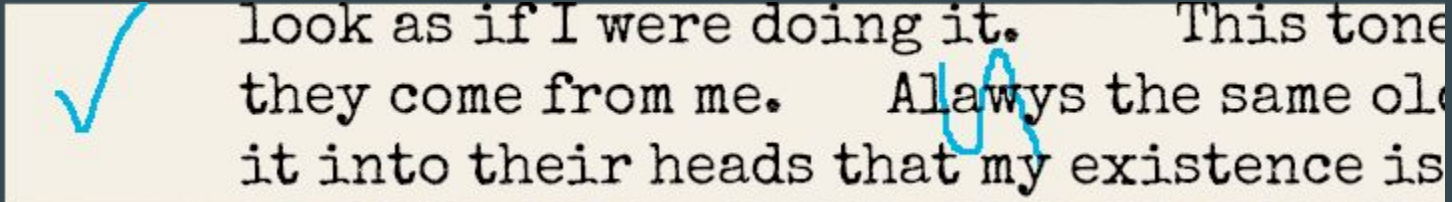Mr. Greene did o'erthrowe Colonel Mustard with a kitchen naif.

<metamark>

<metamark> contains or describes any kind of graphic or written signal within a document the function of which is to determine how it should be read rather than forming part of the actual content of the document.

<metamark>

<metamark> contains or describes any kind of graphic or written signal within a document the function of which is to determine how it should be read rather than forming part of the actual content of the document.

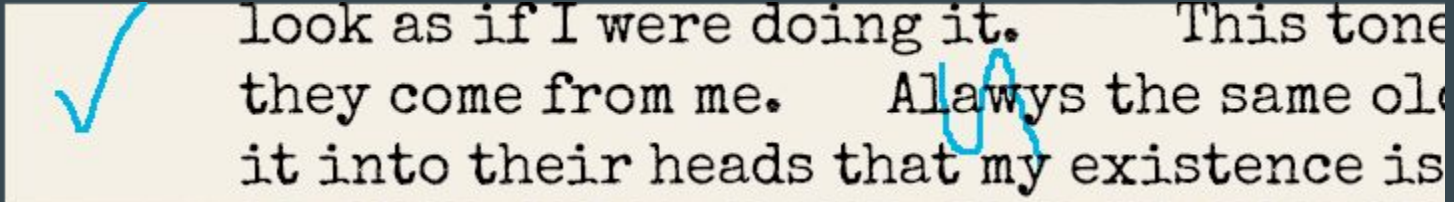`<metamark function="proofreading">checkmark</metamark>`

<metamark>

<metamark> contains or describes any kind of graphic or written signal within a document the function of which is to determine how it should be read rather than forming part of the actual content of the document.

<metamark function="proofreading">checkmark</metamark>

But where do we put it?

<metamark>

E.g. treat it like an addition:

they come from me. <add place="marginleft"><metamark function="proofreading">checkmark

</metamark></add> Alawys the same old

<transpose>

<transpose> describes a single textual transposition as an ordered list of at least two pointers specifying the order in which the elements indicated should be re-combined.

<transpose>

<transpose> describes a single textual transposition as an ordered list of at least two pointers specifying the order in which the elements indicated should be re-combined.

\<transpose\>

\<seg\> (arbitrary segment) represents any segmentation of text below the 'chunk' level.

Step 1: wrapping a \<seg\> around the individual parts of the transposition

Al\<seg\>a\</seg\>\<seg\>w\</seg\>ys

‹transpose›

<seg> (arbitrary segment) represents any segmentation of text below the 'chunk' level.

Step 2: identifying the <seg>s with @xml:id

Al<seg xml:id="trans1">a</seg><seg xml:id="trans2">w</seg>ys

‹transpose›

<ptr> (pointer) defines a pointer to another location.

Step 3: listing the transposition in <transpose> with <ptr>s

```
<listTranspose>

    <transpose>

        <ptr target= "#trans2"/>

        <ptr target= "#trans1"/>

    </transpose>

</listTranspose>
```

\<transpose\>

\<ptr\> (pointer) defines a pointer to another location.

Step 3: listing the transposition in \<transpose\> with \<ptr\>s

```
<listTranspose>

    <transpose>

        <ptr target= "#trans2"/>


        <ptr target= "#trans1"/>

    </transpose>

</listTranspose>
```

Mind!

- Hashtags
- First 2, then 1 (corrected order)

\<transpose>

Where to put it?

\<listTranspose> can be nested inside a lot of different elements (including \<p>), but you may want to put them all together (= list), so it makes sense to put it in \<teiHeader>. You can put it as a direct child of \<profileDesc>.

Alternative Readings (open variants)

What is it?

When a document offers two (or more) alternatives for the same reading, this is called an open variant. At this stage in the writing process, the decision which of the variants will eventually make it into the final, published text is left open.

Example:

Alternative Readings (open variants)

How to encode?

Step 1: wrap tags around individual variants

Alternative Readings (open variants)

How to encode?

Step 1: wrap tags around individual variants

That is <seg type="alternative">an easy thing to say.</seg> <add type="alternative">soon said</add>

Alternative Readings (open variants)

How to encode?

Step 2: identify the elements with @xml:id


```
That is <seg type="alternative" xml:id="alt1">an easy thing to say.</seg>
<add type="alternative" xml:id="alt2">soon said</add>
```

Alternative Readings (open variants)

<alt> (alternation) identifies an alternation or a set of choices among elements or passages.

Step 3: mark the alteration of both variants with <alt/>


```
That is <seg type="alternative" xml:id="alt1">an easy thing to say.</seg>
<add type="alternative" xml:id="alt2">soon said</add>

<alt target="#alt1 #alt2"/>
```

Exercise

Where now? When now? Who now? ~~Unquestioning.~~ ~~without asking~~ stet

What

A Word of Advice:
Think Before You Transcribe

Think Before you Transcribe

It's not because it's there, and there's a way to encode it, that you should.

➔   What do I want to say about the document?
➔   What do I want to do with the elements and attributes that I have encoded?

Examples:

Shelley-Godwin Archive (http://shelleygodwinarchive.org/)

Beckett Digital Manuscript Project (http://www.beckettarchive.org/)

References

Most images in the examples were taken from the TEI Guidelines, which are released under a CC+BY license (see: http://www.tei-c.org/Guidelines/access.xml). All element descriptions were also quoted from the same TEI Guidelines.

TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.0.0. Last modified on 29th March 2016. TEI Consortium. http://www.tei-c.org/Guidelines/P5/ (Accessed on 13 October 2016).

More specifically, the images taken from the TEI Guidelines can be found in the 11th chapter of the guidelines, titled 'Representation of primary documents,' which can be found here: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html. The specific images copied from this chapter are attributed in the slides.

The images on slides 33-34, 61-68, 72-77 were created by the author of the slides.

This work is licensed under a .

All works of other author cited here are their intellectual property and are used for academic teaching purpose only.