In the Tower of Babel
Modelling primary sources of multi-testimonial textual transmissions

Digital Classicist & Institute of Classical Studies Seminar 2012
London, 20 July 2012
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
The 'Babel issue' in a nutshell

- Each primary source uses a different writing (encoding, semiotic) system
- We want to compare the texts they bear
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
Comparing texts

- Each primary source uses a different writing (encoding, semiotic) system
- We want to compare the texts they bear
  - Textual Criticism
  - Processing (e.g. cross-corpus search)
Comparing texts

- Each primary source uses a different writing (encoding, semiotic) system
- We want to compare the texts they bear
  - Textual Criticism
  - Processing (e.g. cross-corpus search)
Comparing texts

- Each primary source uses a different writing (encoding, semiotic) system
- We want to compare the texts they bear
  - Textual Criticism
  - Processing (e.g. cross-corpus search)
Comparing texts

- Simplification: same writing system (no 'Babel' issue)
  - simple variant (at graphemic layer)

\[
\begin{align*}
\text{MS A} & \quad w_G \leftrightarrow w_G \quad \text{MS B} \\
& \quad y_G \leftrightarrow y_G \\
& \quad f_G \leftrightarrow f_G \\
& \quad \leftrightarrow f_G \\
& \quad e_G \leftrightarrow e_G
\end{align*}
\]
Comparing texts

- Simplification: same writing system (no 'Babel' issue)
  - no variant (at linguistic layer)
Comparing texts

- The same text (reading)?
  - Comparing texts at different layers

Comparing at linguistic layer

Comparing at graphemic layer
Comparing texts

- Types of edition in the print age

Critical edition (Pierazzo)
“Museum” edition (Vanhoutte)

Diplomatic edition (Pierazzo)
“Archive” edition (Vanhoutte)
Comparing texts

- **Digital edition**

Digital diplomatic edition (Pierazzo)
Comparing texts

- Digital edition

Open source critical edit. (Bodard-Garcés)
Comparing texts

- Digital edition

Digital edit. as model: sub-systems (Orlandi)
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
In the Tower

- Different Writing Systems (Graphemes):
  A: $\mathbf{p}_G, \mathbf{u}_G$ (no $u_G/v_G$ distinction)
  B: $u_G$ (no $u_G/v_G$ distinction)
  C: $u_G, v_G$

Different Writing Systems (Graphemes):
In the Tower

- Different Alphabets (Letters):
  A: $u_L$ (no $u_L/v_L$ distinction)
  B: $u_L$ (no $u_L/v_L$ distinction)
  C: $u_L, v_L$
In the Tower

- Same text at Linguistic layer (**Inflected word**)
  
  “Perius”, nom. sing. masc. of lemma “perius, -a, um”
In the Tower

- Digital edition: Formalisation
In the Tower

- Life in the Tower of Babel (recapitulation):
  - Each builder, a language
    - Each primary source, a semiotic system
  - ...yet we want builders to interact
    - Textual Criticism
    - Processing (e.g. cross-corpus search)
  - ...at different layers
    - Graphs, allophones, graphemes, linguistic...
  - ... all the sudden, all builders are robots!
    - Formalisation (no human intuition)
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
Two issues

A. Comparison at linguistic layer
   - Substantial readings \( \text{wyfe}_w \neq \text{lyf}_w \)
   - Critical edition
   - The 'text' (Inflected words)

B. Comparison at graphemic layer
   - Accidentals \( \text{wyfe}_G \neq \text{wyffe}_G \)
   - Diplomatic edition
   - The 'spelling' (orthography, Graphemes)
Two issues

A. Comparison at linguistic layer
   - How do you identify elements at linguistic layer (Inflected words) digitally?
     - The Canterbury Tales Project: “Regularized spelling”
     - Tito Orlandi: Linguistic entities
Two issues

B. Comparison at graphemic layer

- Can you compare graphemes through different graphemic systems?
  
  • The Canterbury Tales Project: Unicode; Corpus-wide modelling of the graphemic system
  
  • Tito Orlandi: MS-wide complete modelling of the graphemic system (and of other systems at other textual layers)
Two issues

- The Canterbury Tales Project
- Tito Orlandi, *Informatica Testuale*, Laterza: Roma 2010
- My own project of an experimental edition from the *Anthologia Latina*
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
A. Comparison at linguistic layer

- Why comparing *Words* (at linguistic layer)?

The Canterbury Tales Project:

- Textual criticism
  - Relations between MSS
- Processing
  - Indexing ("Spelling database")
A. Comparison at linguistic layer

- From Grapheme to Word, not Lemma
  - Inflected word ($\text{wyfe}_w \neq \text{wyves}_w$)
    - Not as element of language ($\text{langue}$)
    - But as as element of that text ($\text{parole}$)
  - “Substantial reading”
  - Critical edition
A. Comparison at linguistic layer

- From **Grapheme** to **Word**: semi-automatic
  - Not computable so far
  - Linguistic competence required

  - **Grapheme** → **Letter**: ambiguous graphemes
    
    Ion\_G → Jon\_L / Ion\_L → Jon\_W / Ion\_W
    pdo\_G → perdo\_L / prodo\_L → perdo\_W / prodo\_W
A. Comparison at linguistic layer

- From **Grapheme** to **Word**: semi-automatic
  - Not computable so far
  - Linguistic competence required

- Letter → Word: omographs
  - $est_G \rightarrow est_L \rightarrow est\ (she\ is)_W / est\ (she\ eats)_W$
A. Comparison at linguistic layer

- From **Grapheme** to **Word**: semi-automatic
  - Not computable so far
  - Linguistic competence required

- Letter → Word: different spellings
  - \( \text{wyfe}_w / \text{wyffe}_w \rightarrow \text{wyfe}_w / \text{wyffe}_w \rightarrow \text{wyfe}_w \)
A. Comparison at linguistic layer

- From **Grapheme** to **Word**: semi-automatic
  - Not computable so far
  - Linguistic competence required
  - 'Context': interpretation of the whole text
- Linguistic and **higher-level competence**
  (Jon / Ion)
A. Comparison at linguistic layer

- From **Grapheme** to **Word**: semi-automatic
  - Semi-automatic procedures (human-driven, computer-assisted)
  - "The computer collation program we are using (Collate) permits regularization as part of the collation process" [...] “regularized spelling" (Robinson-Solopova →)
A. Comparison at linguistic layer

- How do you identify elements at linguistic layer (Inflected words) digitally?
  - The Canterbury Tales Project: “Regularized spelling”
  - Tito Orlandi: Linguistic entities
A. Comparison at linguistic layer

- Canterbury: “Regularized spelling”

\[
\begin{array}{cccc}
\text{U0077}_C & \text{U0079}_C & \text{U0066}_C & \text{U0065}_C \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{w}_L \text{y}_L \text{f}_L \text{e}_L \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{w}_{G\text{\_reg}} \text{y}_{G\text{\_reg}} \text{f}_{G\text{\_reg}} \text{e}_{G\text{\_reg}} \\
\end{array}
\]

MS A

\[
\text{wyfe}_G \\
\]

MS B

\[
\text{wyffe}_G \\
\]
A. Comparison at linguistic layer

- Orlandi: Discrete linguistic entities in a database
A. Comparison at linguistic layer

- Orlandi: Discrete linguistic entities in a database → How do you identify them?
A. Comparison at linguistic layer

- Orlandi: Discrete linguistic entities in a database → How do you identify them?

\[
\begin{align*}
\text{deorum}_W & \quad \text{deum}_W \\
\text{deorum}_G & \quad \text{devm}_G
\end{align*}
\]
A. Comparison at linguistic layer

- Orlandi: Discrete linguistic entities in a database → How do you identify them?

\[
\begin{align*}
&\text{deorum}_G \\
&\text{deorum}_W \\
&[\text{deus1}]_C [\text{gen.pl.masc.1}]_C \\
&[\text{deus1}]_C [\text{gen.pl.masc.2}]_C \\
&\text{devm}_G \\
&\text{deum}_W \\
&\text{MS A} \\
&\text{MS B}
\end{align*}
\]
A. Comparison at linguistic layer

- Orlandi: Discrete linguistic entities in a database → How do you identify them?

<table>
<thead>
<tr>
<th>deorum</th>
<th>deum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{\text{G}_{\text{reg}}}$ U0064c U0065c U006Fc U0072c U0075c U006Dc</td>
<td>$^{\text{G}_{\text{reg}}}$ U0075c U006Dc</td>
</tr>
<tr>
<td>[deus1] [gen.pl.masc]C</td>
<td>[deus1] [gen.pl.masc]C</td>
</tr>
</tbody>
</table>

MS A

$\text{deorum}_W$

$\text{deorum}_G$

MS B

$\text{deum}_W$

$\text{devm}_G$
A. Comparison at linguistic layer

- Orlandi/Monella: is it worth it?
A. Comparison at linguistic layer

<table>
<thead>
<tr>
<th>Manuscript</th>
<th>Textual Form</th>
<th>Paradigm</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canterbury, MS A</td>
<td>deum_{G_reg}</td>
<td>[deus1] [gen.plur.masc]_{C}</td>
<td>MS A] Pater devm</td>
</tr>
<tr>
<td>Canterbury, MS B</td>
<td>deum_{G_reg}</td>
<td>[deus1] [acc.sing.masc]_{C}</td>
<td>MS B] Timeo devm</td>
</tr>
<tr>
<td>Orlandi/Monella, MS A</td>
<td>deum_{G_reg}</td>
<td>[deus1] [gen.plur.masc]_{C}</td>
<td></td>
</tr>
<tr>
<td>Orlandi/Monella, MS B</td>
<td>deum_{G_reg}</td>
<td>[deus1] [acc.sing.masc]_{C}</td>
<td></td>
</tr>
</tbody>
</table>

\[ deum_{W} \neq deum_{W} \]
A. Comparison at linguistic layer

- It's only worth it
  - if words are not their regularised spelling, i.e.
  - if homograph words exist

\[
\text{deum}_w \neq \text{deum}_w
\]

\[
Pater \text{ deum} \quad \text{Timeo deum}
\]
A. Comparison at linguistic layer

- It's only worth it
  - if substantial readings are not their regularised spelling, i.e.
  - if homograph substantial readings exist (do they? → Concept of “reading”)

\[ \text{deum} \neq \text{deum?} \]
A. Comparison at linguistic layer

Canterbury, MS A

\( \text{deum}_{G_{\text{reg}}} \)

Canterbury, MS B

\( \text{deum}_{G_{\text{reg}}} \)

Orlandi/Monella, MS A

\( \text{deum}_{G_{\text{reg}}} \)

[deus1] [gen.plur.masc]

Orlandi/Monella, MS B

\( \text{deum}_{G_{\text{reg}}} \)

[deus1] [acc.sing.masc]

MS A

Pater,
devm Neptvnvm odi

MS B

Pater deum,
Neptunum odi

\( \text{deum} \neq \text{deum} \)
A. Comparison at linguistic layer

Canterbury, MS A

\[\textit{deum}_{\text{G\_reg}}\]

Canterbury, MS B

\[\textit{deum}_{\text{G\_reg}}\]

Orlandi/Monella, MS A

\[\textit{deum}_{\text{G\_reg}} [\textit{deus1}] [\text{gen.plur.masc}]_{\text{C}}\]

Orlandi/Monella, MS B

\[\textit{deum}_{\text{G\_reg}} [\textit{deus1}] [\text{acc.sing.masc}]_{\text{C}}\]

\[\text{Pater, devm Neptunvm odi}\]

\[\text{Pater deum, Neptunum odi}\]

\[\textit{deum} = \textit{deum}\?\]
A. Comparison at linguistic layer

- Normally done via regularised spelling
- Inflected words are not their regularised spelling (omographs)
- Are there omograph substantial readings (*deum*)?
- If so:
  - No regularised spelling
  - But formal identifiers for inflected words (spelling, lemma, identifier)
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer
B. Comparison at graphemic layer

- Why comparing spellings (at graphemic layer)?
  - Historical Linguistics
    - Evidence for lexical, morphological and phonetic evolution
B. Comparison at graphemic layer

• Why comparing spellings (at graphemic layer)?
  – Palaeography
    • Indirect evidence for letters (Benskin: letter þ thorn →)
      – Overlapping with Historical Linguistics
        (Emiliano: “Scripto-linguistic change” →)
    • Though only graphemes
      – Not allographs (graphetes, s/ſ) or graphs (bitmap)
B. Comparison at graphemic layer

- Why comparing spellings (at graphemic layer)?
  - Textual Criticism
    - 'Orthographic' apparatus (The Hengwrt Chaucer Digital Facsimile, Collation function →)
    - "Although for most manuscripts collation of the regularized text will produce sufficient information to place those manuscripts in genetic relation to one another [...]" (Robinson-Solopova →)
B. Comparison at graphemic layer

- *The loue of love*
- *Loue*: “A hill or mountain” ≠ *Love*: “Love”
- Middle English alphabet: u ≠ v
B. Comparison at graphemic layer

The loue of love

MS A

The loue of loue

MS B
B. Comparison at graphemic layer

The loue of love

\[ a_L \]
\[ u_L \]
\[ v_L \]

MS A

The loue of loue

\[ a_L \]
\[ u_L \]
\[ v_L \]

MS B
B. Comparison at graphemic layer

The $\text{lou}_G^G$ of $\text{lov}_G^G$  

The $\text{lou}_G^G$ of $\text{lou}_G^G$
B. Comparison at graphemic layer

- Back in the Tower

The $\text{lou}_G$ e of $\text{lov}_G$ e

The $\text{lou}_G$ e of $\text{lou}_G$ e
B. Comparison at graphemic layer

- \( a_G \) \( a_L \) \( a_G \)
- \( u_G \) \( u_L \) \( u_G \)
- \( v_G \) \( v_L \) \( u_G \)

- Back in the Tower
  - “Grapheme. (1) A minimally distinctive unit of writing in the context of a particular writing system” (Unicode Glossary →).
B. Comparison at graphemic layer

- Back in the Tower
  - De Saussure: relational nature of signs within a semiotic system

The \(\text{lou}_G\) e of \(\text{lov}_G\) e

The \(\text{lou}_G\) e of \(\text{lou}_G\) e
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th>$a_G$</th>
<th>$a_L$</th>
<th>$a_G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_G$</td>
<td>$u_L$</td>
<td>$u_G$</td>
</tr>
<tr>
<td>$v_G$</td>
<td>$v_L$</td>
<td></td>
</tr>
</tbody>
</table>

- Back in the Tower
  - De Saussure: relational nature of signs *within* a semiotic system
  - This is a different grapheme than this, as the former is in contrast with this, while the latter is not

The $\text{lou}_G$ of $\text{lov}_G$ e; The $\text{lou}_G$ e of $\text{lou}_G$ e

MS A  MS B
B. Comparison at graphemic layer

- Back in the Tower
  - De Saussure: relational nature of signs within a semiotic system
  - This is a different grapheme than this, as the former is in contrast with this, while the latter is not

The $\text{lou}_G$e of $\text{lov}_G$e

The $\text{lou}_G$e of $\text{lou}_G$e
B. Comparison at graphemic layer

- Back in the Tower
  - De Saussure: relational nature of signs within a semiotic system
  - This is a different grapheme than this, as the former is in contrast with this, while the latter is not

The $\text{lou}_G \text{e}$ of $\text{lov}_G \text{e}$  The $\text{lou}_G \text{e}$ of $\text{lou}_G \text{e}$
B. Comparison at graphemic layer

How do you encode graphemes digitally?

The $lou_G$ e of $lov_G$ e

The $lou_G$ e of $lou_G$ e
B. Comparison at graphemic layer

- TEI
  - Unicode → →
- The Canterbury Tales Project
  - Corpus-wide definition of the graphemic system →

\[
\begin{array}{ccc}
  a_G & a_L & a_G \\
  u_G & u_L & u_G \\
  v_G & v_L & u_G \\
\end{array}
\]
B. Comparison at graphemic layer

The love of love

<table>
<thead>
<tr>
<th>$a_G$</th>
<th>$a_L$</th>
<th>$a_G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_G$</td>
<td>$u_L$</td>
<td>$u_G$</td>
</tr>
<tr>
<td>$v_G$</td>
<td>$v_L$</td>
<td></td>
</tr>
</tbody>
</table>

$\text{U0075}_c$ $\text{U0076}_c$ $\text{U0075}_c$ $\text{U0075}_c$

The love of love

MS A MS B
B. Comparison at graphemic layer

The lo\textsubscript{G} u\textsubscript{G} e of lo\textsubscript{L} v\textsubscript{L} e

No match: different graphemes

The lo\textsubscript{G} u\textsubscript{G} e of lo\textsubscript{L} v\textsubscript{L} e

The lo\textsubscript{G} u\textsubscript{G} e of lo\textsubscript{L} v\textsubscript{L} e

<table>
<thead>
<tr>
<th>a\textsubscript{G}</th>
<th>a\textsubscript{L}</th>
<th>a\textsubscript{G}</th>
</tr>
</thead>
<tbody>
<tr>
<td>u\textsubscript{G}</td>
<td>u\textsubscript{L}</td>
<td>u\textsubscript{G}</td>
</tr>
<tr>
<td>v\textsubscript{G}</td>
<td>v\textsubscript{L}</td>
<td>u\textsubscript{G}</td>
</tr>
</tbody>
</table>
B. Comparison at graphemic layer

Match:
same grapheme

The love of love

The love of love
B. Comparison at graphemic layer

- Orlandi
  - MS-wide *complete* definition of graphemic system
    - Not corpus-wide (Canterbury)
    - Not world-wide (Unicode)
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th>a_G</th>
<th>a_L</th>
<th>a_G</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_G</td>
<td>u_L</td>
<td>u_G</td>
</tr>
<tr>
<td>v_G</td>
<td>v_L</td>
<td>u_G</td>
</tr>
</tbody>
</table>

<teiHeader>
<encodingDesc>
<charDecl>

Graphemes

Allographs

The lou_G e of lov_G e

The lou_G e of lou_G e

MS A

MS B
B. Comparison at graphemic layer

The $\text{lou}_G$ of $\text{lov}_G$  The $\text{lou}_G$ of $\text{lou}_G$

<body>
<p>
<g ref="uv" />
<!-- or, better: -->
<!ENTITY uv 'g ref="#uv" />
&uv;

The $\text{lou}_G$ of $\text{lov}_G$  The $\text{lou}_G$ of $\text{lou}_G$
B. Comparison at graphemic layer

The $\text{lo}_G \text{u}_G$ of $\text{lo}_G \text{y}_G$
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th></th>
<th>a_G</th>
<th>a_L</th>
<th>a_G</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_G</td>
<td>u_L</td>
<td>u_G</td>
<td></td>
</tr>
<tr>
<td>v_G</td>
<td>v_L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The \text{lo}u_{G} e of \text{lo}v_{G} e

The \text{lo}u_{G} e of \text{lo}u_{G} e
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th>a_G</th>
<th>a_L</th>
<th>a_G</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_G</td>
<td>u_L</td>
<td>u_G</td>
</tr>
<tr>
<td>v_G</td>
<td>v_L</td>
<td>u_G</td>
</tr>
</tbody>
</table>

The `lou` of `lov` in MS A

The `lou` of `lou` in MS B

<charDesc>
<char xml:id="uv">
  <charName>SMALL LATIN U OR V</charName>
  <desc>U-shaped when lowercase, V-shaped when uppercase. Content: either small letter Latin u or small letter Latin v</desc>
</char>

MS A

MS B
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th>a_G</th>
<th>a_L</th>
<th>a_G</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_G</td>
<td>u_L</td>
<td>u_G</td>
</tr>
<tr>
<td>v_G</td>
<td>v_L</td>
<td>u_G</td>
</tr>
</tbody>
</table>

*The lous e of lous e*

---

<charDesc>
<char xml:id= "uv">
  <charProp>
    <localName>Expression</localName>
    <value>U+0075</value>
    <localName>Content</localName>
    <value>u|v</value>
  </charProp>
</char>
B. Comparison at graphemic layer

The $\text{lou}_G$ of $\text{lov}_G$

The $\text{lou}_G$ of $\text{lov}_G$
B. Comparison at graphemic layer

Two minims, lowercase

**Expression** (shape): indistinguishable.
**Content** (letter): n or u.

The reader does not identify the grapheme from its shape, but guessing its content.

Graphematic information is not conveyed by graphic information, but by linguistic information (context), so the scribe was confident in our Linguistic competence to tell the graphemes apart.

The \( \text{lo}_G \text{u}_G \text{e} \) of \( \text{lov}_G \text{e} \)  The \( \text{lo}_G \text{u}_G \text{e} \) of \( \text{lo}_G \text{u}_G \text{e} \)
B. Comparison at graphemic layer

The $lou_1$ e of $lov_1$ e  

The $lou_2$ e of $lov_2$ e

MS A

MS B
B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th></th>
<th>a₀</th>
<th>a₀</th>
<th>a₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>u₀</td>
<td>u₀</td>
<td>u₀</td>
<td>u₀</td>
</tr>
<tr>
<td>v₀</td>
<td>v₀</td>
<td>v₀</td>
<td>v₀</td>
</tr>
</tbody>
</table>

The \text{lo}u\varepsilon \text{ of } \text{lo}v\varepsilon

\text{MS A}

U0075_c \quad U0076_c

The \text{lo}u\varepsilon \text{ of } \text{lo}v\varepsilon

\text{MS A}

\&u-u_c \quad \&u-v_c

\text{MS B}
B. Comparison at graphemic layer

- Fun
- But how do you compare graphemes now?

The $\text{love}_G$ of $\text{love}_G$

The $\text{love}_G$ of $\text{love}_G$
B. Comparison at graphemic layer

The $\text{lo}_g\text{u}_e$ of $\text{lo}_g\text{v}_e$

The $\text{lo}_g\text{u}_e$ of $\text{lo}_g\text{u}_e$

MS A

MS B
B. Comparison at graphemic layer

- When the game gets tough...
  - Different alphabets
    (Middle English, thorn letter)
B. Comparison at graphemic layer

- When the game gets tough...
  - Different alphabets, shared phonetics (Middle English, thorn letter)
B. Comparison at graphemic layer

- When the game gets tougher...
  - No 1:1 grapheme / letter ratio (Huitfeldt → )
    - Brevigraphs as graphemes (up to 40%)
    - “Grapheme. (1) A minimally distinctive unit of writing in the context of a particular writing system” (Unicode Glossary → ).
### B. Comparison at graphemic layer

<table>
<thead>
<tr>
<th>(a_G)</th>
<th>(a_L)</th>
<th>(a_G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_G)</td>
<td>(p_L)</td>
<td>(p_G)</td>
</tr>
<tr>
<td>(e_G)</td>
<td>(e_L)</td>
<td>(e_G)</td>
</tr>
<tr>
<td>(r_G)</td>
<td>(r_L)</td>
<td>(r_G)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p_G)</td>
</tr>
</tbody>
</table>

- When the game gets tougher...
  - No 1:1 grapheme / letter ratio (Huitfeldt →)
    - Brevigraphs as graphemes (up to 40%)
    - “Grapheme. (1) A minimally distinctive unit of writing in the context of a particular writing system” (Unicode Glossary →).
B. Comparison at graphemic layer

- When the game gets most tough...
  - (Classics start to play)
  - Different alphabets, different phonetics (Ancient Latin)
B. Comparison at graphemic layer

- When the game gets most tough...
  - (Classicists start to play)
  - Different alphabets, different phonetics (Ancient Latin)
  - Still possible to formalise relationships
B. Comparison at graphemic layer

- Description of graphemic system
  - MS-wide
  - Complete
- Formalisation of relationships between graphemes in different graphemic systems
  - Involving higher-level entities (letters, phonemes)
- Intelligent searches, intelligent results
Outline

• The 'Babel issue'
  – The 'Babel issue' in a nutshell
  – Comparing texts
  – In the Tower

• Two issues
  A. Comparison at the linguistic layer
  B. Comparison at the graphemic layer