Paolo Monella

A Saussurean approach to graphemes declaration in charDecl for manuscripts encoding

paolo.monella@gmx.net

TEI Conference - Rome, October 3, 2013
Outline

- Saussure
Outline

• Saussure
Outline

- Saussure
- Currently
Outline

- Saussure
- Currently
- Proposed change
Outline

- Saussure
- Currently
- Proposed change
- Interoperability
Outline

- Saussure
- Currently
- Proposed change
- Interoperability
- “Why don't you use...?”
Outline

- Saussure
- Currently
- Proposed change
- Interoperability
- “Why don't you use...?”
- What does it take?
Saussure

Ferdinand De Saussure

Relational nature of signs within a semiotic system
Saussure
Saussure
he said: alas

yes: sir

Saussure
Saussure

Every noun

≠

U+0075

U+0076

E
v
r
y

n
o
u
n

Every noun

MS A

MS B
Currently

- Unicode (a “u” is a “u”)

![Screen capture of the Text Encoding Initiative (TEI) website, showing the Table of Contents and the 5.1 section on Non-standard Characters and Glyphs.](image-url)
Currently

- Unicode (a “u” is a “u”)
- Corpus-wide normalization (Canterbury Tales)
Currently

- Unicode (a “u” is a “u”)
- Corpus-wide normalization (Canterbury Tales)
- Documentation

2. The practice of this transcription

A specially designed computer screen font was used for the transcription.

```
abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
```

These were supplemented by the following Middle English characters:

```
³ ³ ³
```

The following signs of abbreviation were available:

```
³³ ³³ ³³
```

These characters are available as superscripts:

```
a e i r t u
```

Characters usually occurring at word ends are:

```
γ ẞ ẞ
```

Marks of punctuation are:

```
¶ , ; : . ()
```
Currently

2. The practice of this transcription

A specially designed computer screen font was used for the transcription:

```
abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
```

Those were supplemented by the following Middle English characters:

```
\&\%\'\``
```
```
\&\%\'\``
```

The following signs of abbreviation were available:

```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
\&\%\'\``
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
Currently

<charDecl>

a
b
<glyph id="fun-b">

a
b
fun-b
c
d
...

2. The practice of this transcription

A specially designed computer screen font was used for the transcription:

abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ

Those were supplemented by the following Middle English characters:

j p y f

The following signs of abbreviation were available:

godp

Those characters are available as superscripts:

1 2 3 4

Characters usually occurring at word ends are:

a b e

Marks of punctuation are:

, . : ; ( )
Proposed change: `<char>`

```
<charDecl>
  a
  b
  <glyph id="fun-b"/>
</charDecl>

<charDecl>
  <char xml:id="a"/>
  <char xml:id="b"/>
  <glyph id="fun-b"/>
</charDecl>
```

Guidelines:

```
 a
 b
 fun-b
 c
d
...`

Proposed change: <char>

<a>

<glyph id="fun-b">

<charDecl>

<char xml:id="v">

<charName>Premodern Latin uncial lowercase v</charName>

<charProp>

<localName>Expression</localName>

</localName>

/value>U+0076</value>

<localName>Content</localName>

/value>v</value>

<charProp>

<mapping type="Unicode">v</mapping>

<graphic url="v.jpg"/>

</char>

</charDecl>

<charDecl>
Proposed change: `<char>`
Proposed change: <char>

<charDecl>
  a
  b
  <glyph id="fun-b">
</charDecl>

<body>
  a
  b
  <g ref="#fun-b">
</body>

<charDecl>
  <char xml:id="a">
  <char xml:id="b">
  <glyph id="fun-b">
</charDecl>

Guidelines
Proposed change: <char>
Proposed change: `<char>`

```xml
<charDecl>
  a
  b
  <glyph id="fun-b">
</charDecl>

<body>
  a
  b
  <g ref="#fun-b">
</body>
```

```xml
<charDecl>
  <char xml:id="a">
  <char xml:id="b">
    <glyph id="fun-b">
  </glyph>
</charDecl>

<body>
  <g ref="#a">
  <g ref="#b">
    <g ref="#fun-b">
  </g>
</body>
```
Proposed change: `<char>`

```
<charDecl>
  a
  b
  <glyph id="fun-b">
</charDecl>

<body>
  a
  b
  <g ref="#fun-b">
</body>
```

Guide-lines

```
<charDecl>
  <char xml:id="a">
  <char xml:id="b">
  <glyph id="fun-b">
</charDecl>

<body>
  <g ref="#a">
  <g ref="#b">
  <g ref="#fun-b">
</body>
```
Proposed change: `<char>`

**<charDecl>**
- `<char xml:id="a">`<char xml:id="b">`
- `<glyph id="fun-b">`

**<body>**
- `<g ref="#a">`<g ref="#b">`
- `<g ref="#fun-b">`
Proposed change: no <g>

Proposed change:

```xml
<charDecl>
  <a/>
  <b/>
  <glyph id="fun-b"/>
</charDecl>

<body>
  a
  b
  <g ref="#fun-b"/>
</body>
```

Guide-lines

```xml
<charDecl>
  <char xml:id="a"/>
  <char xml:id="b"/>
  <glyph id="fun-b"/>
</charDecl>

<body>
  a
  b
  <g ref="#fun-b"/>
</body>
```

Techn.

```xml
<body>
  a
  b
  <g ref="#fun-b"/>
</body>
```
Interoperability

<charDecl>
  <char xml:id="y">
    <mapping type="Unicode">
      y
    </mapping>
  </char>
</charDecl>

<body>
every y noun
</body>

<charDecl>
  <char xml:id="y">
    <mapping type="Unicode">
      y
    </mapping>
  </char>
</charDecl>

<body>
every y noun
</body>
Interoperability

MS A

<charDecl>
  <char xml:id="u">
    <mapping type="mapto_u">
      u
    </mapping>
    <mapping type="mapto_v">
      v
    </mapping>
  </char>
</charDecl>

<body>
every noun
</body>

MS B

<charDecl>
  <char xml:id="u">[...]
  <char xml:id="v">
    <mapping type="standard">
      v
    </mapping>
  </char>
</charDecl>

<body>
every noun
</body>
“Why don't you use...?”

• Medieval Unicode Font Initiative
“Why don't you use...?”

- Medieval Unicode Font Initiative
- Unicode / Private Use Area (PUA)
“Why don't you use...?”

- Medieval Unicode Font Initiative
- Unicode / Private Use Area (PUA)
- Gaiji / <glyph>
“Why don't you use...?”

- Medieval Unicode Font Initiative
- Unicode / Private Use Area (PUA)
- Gaiji / <glyph>
- SGML / Writing System Declaration (WSD)
What does it take?

<charDecl>
  a
  b
  <glyph id="fun-b">
</charDecl>

<body>
  a
  b
  <g ref="#fun-b">
</body>

<charDecl>
  <char xml:id="a"> a
  <char xml:id="b"> b
  <glyph id="fun-b">
</charDecl>

<body>
  <g ref="#a"> a
  <g ref="#b"> b
  <g ref="#fun-b">
</body>

Guidelines

Techn.
What does it take?

- More lines of code
What does it take?

- More lines of code
- Guidelines: `<char>`
What does it take?

- More lines of code
- Guidelines: `<char>`

- Technical: no `<g>`
What does it take?

- More lines of code
- Guidelines: <char>

- Technical: no <g>
- Interoperability: <mapping>

Guidelines

<charDecl>
<a>
<b>
<glyph id="fun-b">
</body>
What does it take?

- More lines of code
- Guidelines: `<char>`

Technical: no `<g>`

Interoperability: `<mapping>`