



# *Ursus Beneventanus – Data pitch*

Graphic markup (punctuation, space),  
abbreviations  
and the “e caudatum” conundrum

## **Paolo Monella**

*Die (hyper-)diplomatische Transkription und ihre Erkenntnispotentiale / Hackathon*  
Bergische Universität Wuppertal (BUW), 7. Februar 2020

# **Graphic markup (punctuation, space)**

# Punctuation as “graphic markup”

**<w>infiniti</w>**

**<pc n=“,”>▪</pc>**

**<w>mutata</w>**

**<pc n="quote">▪</pc>**

**<w>tum</w>**

**<pc n="quote">▪</pc>**

# Space as “graphic markup”

`<w>in</w>`

`<w>nomi`

`<lb break="no" n="1r.a.14" />`

`natiuo</w>`

`<pc n=", ">·</pc>`

`<w>et</w>`

`<w>non</w>`

`<pc n="space"> </pc>`

`<w>habet</w>`

[...]

`<w>unius`

`<pc n="space"> </pc>`

`cuiusque</w>`

Punctuation as  
“graphic markup”

Space as  
“graphic markup”

# Tokenization

## “Graphic word” vs “linguistic word”

`<w>in</w>`

`<w>nomi`

`<lb break="no" n="1r.a.14" />`

`natiuo</w>`

`<pc n=",">·</pc>`

`<w>et</w>`

`<w>non</w>`

`<pc n="space"> </pc>`

`<w>habet</w>`

[...]

`<w>unius`

`<pc n="space"> </pc>`

`cuiusque</w>`

Distinction between  
“graphic word” and  
“linguistic word”

# Tokenization

## “Graphic word” vs “linguistic word”

`<w>in</w>`

`<w>nomi`

`<lb break="no" n="1r.a.14" />`

`natiuo</w>`

`<pc n=", ">.</pc>`

`<w>et</w>`

`<w>non</w>`

`<pc n="space"> </pc>`

`<w>habet</w>`

[...]

`<w>unius`

`<pc n="space"> </pc>`

`cuiusque</w>`

Distinction between

“graphic word” and

“linguistic word”

# Tokenization

## “Graphic word” vs “linguistic word”

`<w>in</w>`

`<w>nomi`

`<lb break="no" n="1r.a.14" />`

`natiuo</w>`

`<pc n=", ">.</pc>`

`<w>et</w>`

`<w>non</w>`

`<pc n="space"> </pc>`

`<w>habet</w>`

`[...]`

`<w>unius`

`<pc n="space"> </pc>`

`cuiusque</w>`

Distinction between  
“graphic word” and  
“linguistic word”

# **Abbreviations**

# Abbreviations

## Abbreviation marks, expansion

<w ana="11C---D1---" lemma="ordo" xml:id="w66">

**ordin**

<choice>

<abbr type="superscription">

**e**<am>**2**</am>

</abbr>

<expansion>**em**</expansion>

</choice>

</w>

Base grapheme

Combining  
abbreviation mark

Expansion

# Abbreviations

## Abbreviation types

**ordin**

ordin<sup>2</sup>

```
<choice>  
  <abbr type="superscription">  
    e<am>2</am>  
  </abbr>  
  <expn>em</expn>  
</choice>
```

**quo**

quoq;

```
<choice>  
  <abbr type="after">  
    q<am>;</am>  
  </abbr>  
  <expn>que</expn>  
</choice>
```

**qbus**

```
<choice>  
  <abbr type="brevigraph">  
    ¢  
  </abbr>  
  <expn>qui</expn>  
</choice>
```

**bus**

**cor**

cor

```
<choice>  
  <abbr type="omission">  
    cor  
  </abbr>  
  <expn>correptum</expn>  
</choice>
```

# **The “e caudatum” conundrum**

# The “e caudatum” conundrum

## Allograph or grapheme/abbreviation?

| <u>Gra-</u><br><u>pheme</u> | <u>Alpha-</u><br><u>beme(s)</u> | <u>Grapheme</u><br><u>visualiza-</u><br><u>tion</u> | <u>Type</u>       |
|-----------------------------|---------------------------------|-----------------------------------------------------|-------------------|
| a                           | a                               | a                                                   | <u>Alphabetic</u> |
| b                           | b                               | b                                                   | <u>Alphabetic</u> |
| c                           | c                               | c                                                   | <u>Alphabetic</u> |
| d                           | d                               | d                                                   | <u>Alphabetic</u> |
| e                           | e                               | e                                                   | <u>Alphabetic</u> |
| <u>æ</u>                    | <u>ae</u>                       | <u>ē</u>                                            | <u>Alphabetic</u> |
| f                           | f                               | f                                                   | <u>Alphabetic</u> |
| g                           | g                               | g                                                   | <u>Alphabetic</u> |
| h                           | h                               | h                                                   | <u>Alphabetic</u> |
| i                           | i                               | i                                                   | <u>Alphabetic</u> |

Medieval Latin:  
“cesar”, not “caesar”

But:  
“cesar” **and** “cęsar”

Question:  
(A) allograph of “e” or  
(B) abbreviation of “ae”?



# *Ursus Beneventanus – Data pitch*

Graphic markup (punctuation, space),  
abbreviations  
and the “e caudatum” conundrum

## **Paolo Monella**

*Die (hyper-)diplomatistische Transkription und ihre Erkenntnispotentiale / Hackathon*  
Bergische Universität Wuppertal (BUW), 7. Februar 2020