

Paolo Monella  
paolo.monella@unipa.it

# Fuori dalla torre di Babele: interoperabilità e sistemi grafici pre-moderni



*Out of the Tower of Babel :  
interoperability and pre-modern  
writing systems*

December 4, 2019

[ve]dph

Venice Centre for  
Digital and Public  
Humanities





**In a nutshell**



# In a nutshell

- Digital scholarly editions: diplomatic/normalized
- Hjelmslev's "analysis" as digital modelling for pre-modern writing systems
- Open issues
- Interoperability through modelling



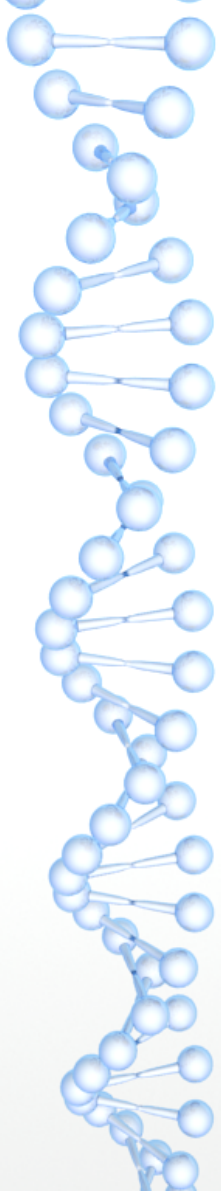
# My focus

- *European Medieval handwriting*
- Pre-Gutenberg
  - Handwriting
  - Early print → imitating handwriting
- Alphabetic writing systems
  - **Latin** script (Italian, English...), Greek, Cyrillic...
  - No Cuneiform, Arabic, Chinese etc.

**Diplomatic/normalized**

# Diplomatic/normalized

ueneñ

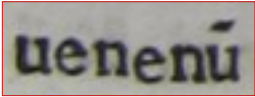
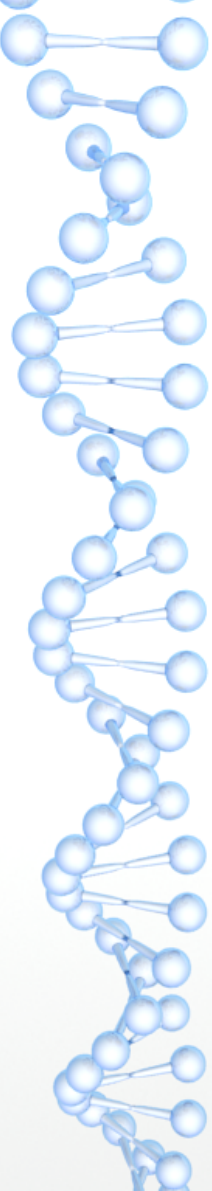


# Diplomatic/normalized

uenenū

- uenenū

# Diplomatic/normalized



uenenū

- uenenū

Diplomatic

- Historical documentation
- Visualization



# Diplomatic/normalized

- venenum

uenenū

- uenenū

Diplomatic

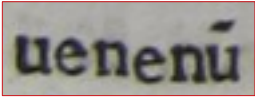
- Historical documentation
- Visualization

# Diplomatic/normalized

- venenum

Normalized

- Processing
  - Search
  - Collation
  - NLP (lemma, PoS etc.)
  - Statistics (distant reading)...



uenenū

- uenenū

Diplomatic

- Historical documentation
- Visualization

# Diplomatic/normalized

The screenshot shows the 'The Digital Vercelli Book beta' interface. On the left, a thumbnail of a manuscript page is displayed. On the right, the text of the manuscript is shown in a diplomatic transcription. The interface includes a search bar, navigation controls, and a zoom level of 27%.

Magnifier 🔍 HotSpot ○ TextLink ℹ️

DOTR 104v Diplomatic 🔍

7 **h**Wæt ic swefna cýrt secgan wylle hæf mege mætte  
8 to midre nihte syðþan reord berend reste wunedon.  
9 þuhte me þæt ic ge sawe syllicre treow onlyft  
10 lædan leohte be wunden beama beorhtost call þæt  
11 beacen wæf be goten mid golde gimmar stodon figere  
12 æt foldan fceatum . swylce þær fife wæron uppe  
13 onþam eaxle ge sþanne be heoldon þær engel dryht  
14 nef ealle figere þurh forð ge sceaft newæf ðærhuru  
15 fracodef gealga . achine þær beheoldon halige galfar  
16 men ofer moldan 7eall þeof mære ge sceaft ⁊  
17 Syllic wæf se fige beam 7ic rynnū fah for wunded  
18 mid wommū ge seah ic wuldref treow. wædum geweor  
19 ðode wynnū scinan ge gyred mid golde gimmar hæfdon  
20 be wrigene weorðlice wealdes treow ⁊ hwæðre ic  
21 þurh þæt gold ongytan meahthe eamra ærgewin  
22 hæf hit arrest ongan swatan on þa fwiðran healf

1:1 | 27%

7 **h**Wæt ic swefna

<http://vbd.humnet.unipi.it/beta/#104v>

# Diplomatic/normalized

The Digital Vercelli Book beta

Magnifier 🔍 HotSpot ○ TextLink ⌘

DOTR 104v Interpretative

Diplomatic

Interpretative

7 Hwæt ic swefna eyst seggan wylle hæt me gemætte  
8 to midre nihte syðan reordberend reste wunedon.  
9 þuhte me hæt ic gesawe syllicre treow on lyft  
10 lædan leohte bewunden beama beorhtost eall þæt  
11 beacen wæs begoten mid golde gimmas stodon fægere  
12 æt foldan sceatum . swylce þær fife wæron uppe  
13 on þam eaxlegespanne beheoldon þær engel dryht  
14 nes ealle fægere þurh forðgescaft ne wæs ðær huru  
15 fracodes gealga . ac hinc þær beheoldon halige gastas  
16 men ofer moldan and eall þeos mære gescaft .  
17 Syllic wæs se sigebeam and ic synnum fah forwunded  
18 mid wommum geseah ic wuldres treow . wædum geweor  
19 ðode wynnnum scinan gegyred mid golde gimmas hæfdon  
20 bewrigene weorðlice wealdes treow . hwæðre ic  
21 þurh hæt gold ongytan meahthe earmra ærgewin  
22 hæt hit ærest ongan swætan on þa swiðran healf

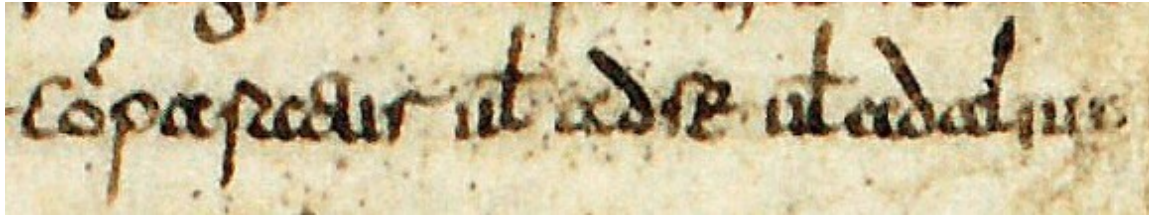
1:1 27%

7 Hwæt ic swefna

<http://vbd.humnet.unipi.it/beta/#104v>

## Diplomatic/normalized

- Comparatur vel ad se vel ad alium



- cōparaEur ut adfe utadaliu

**Hjelmslev's "analysis"  
as digital modelling for  
pre-modern writing systems**

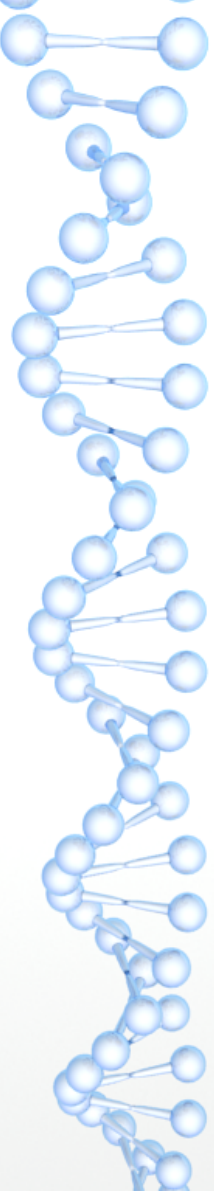
# System / text

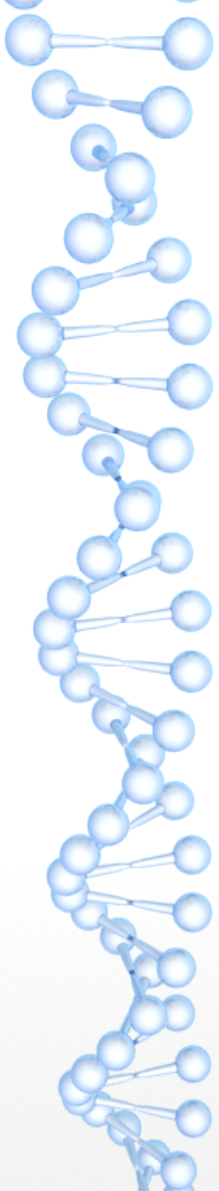
Paradigmatic

Syntagmatic (text, process)

• còparaεur uł adfe uładaliu

(langue, system)





# System / text

System

<s>  
<t>

Text

- cōparaεur

uł

adfe

uładaliu

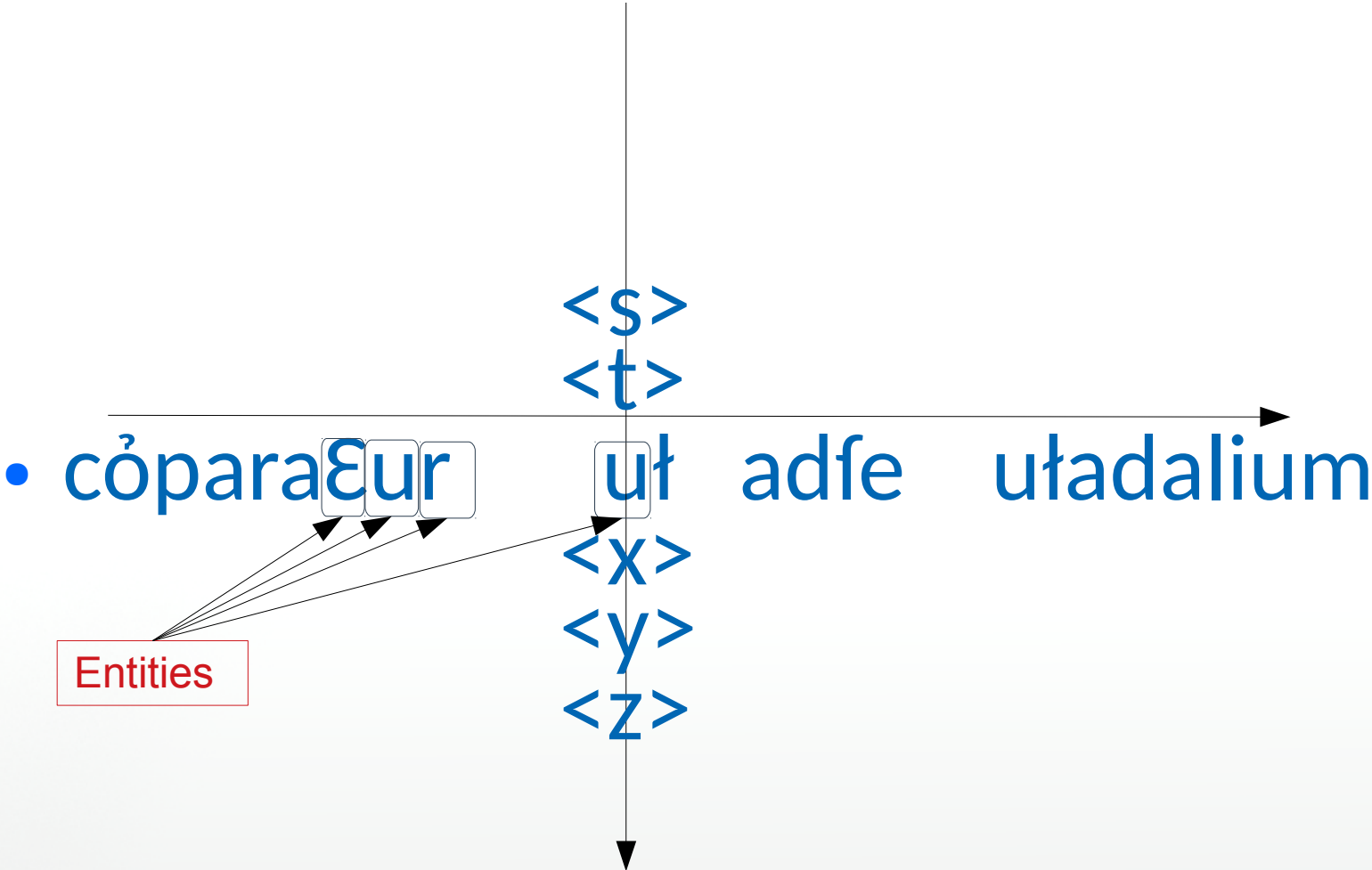
<x>

<y>

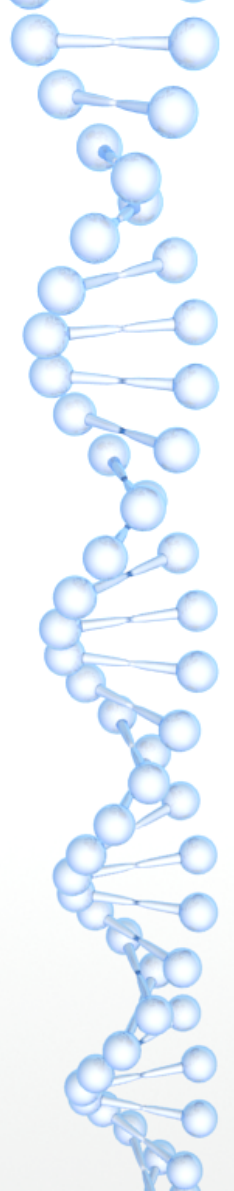
<z>



# Modelling/“analysis”: entities



# Modelling/“analysis”: entities

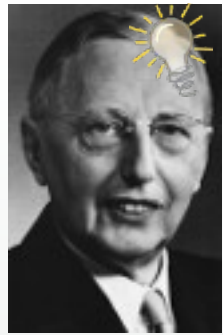


• còparaξur

Entities

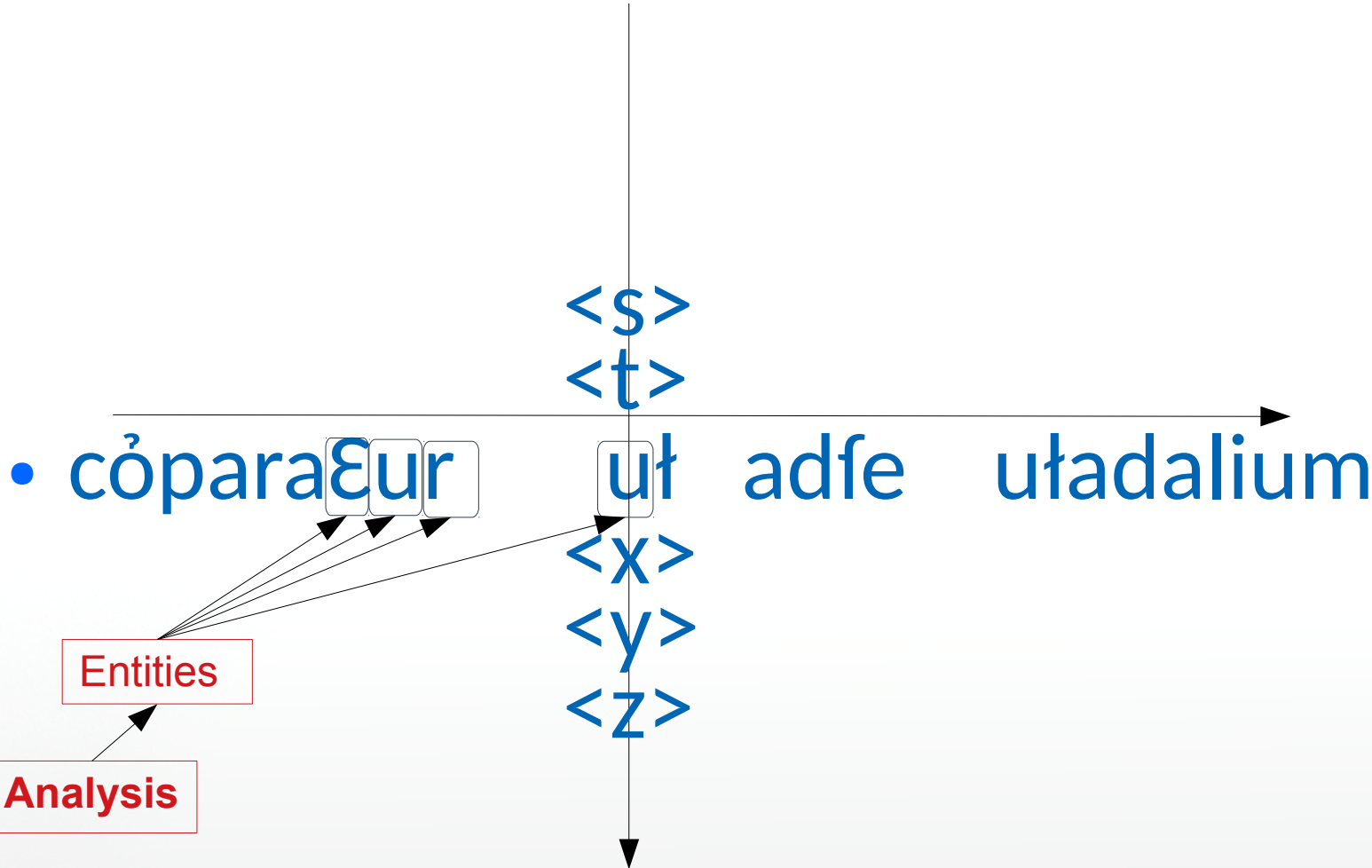
<s>  
<t>  
<x>  
<y>  
<z>

ut adfe utadaliu

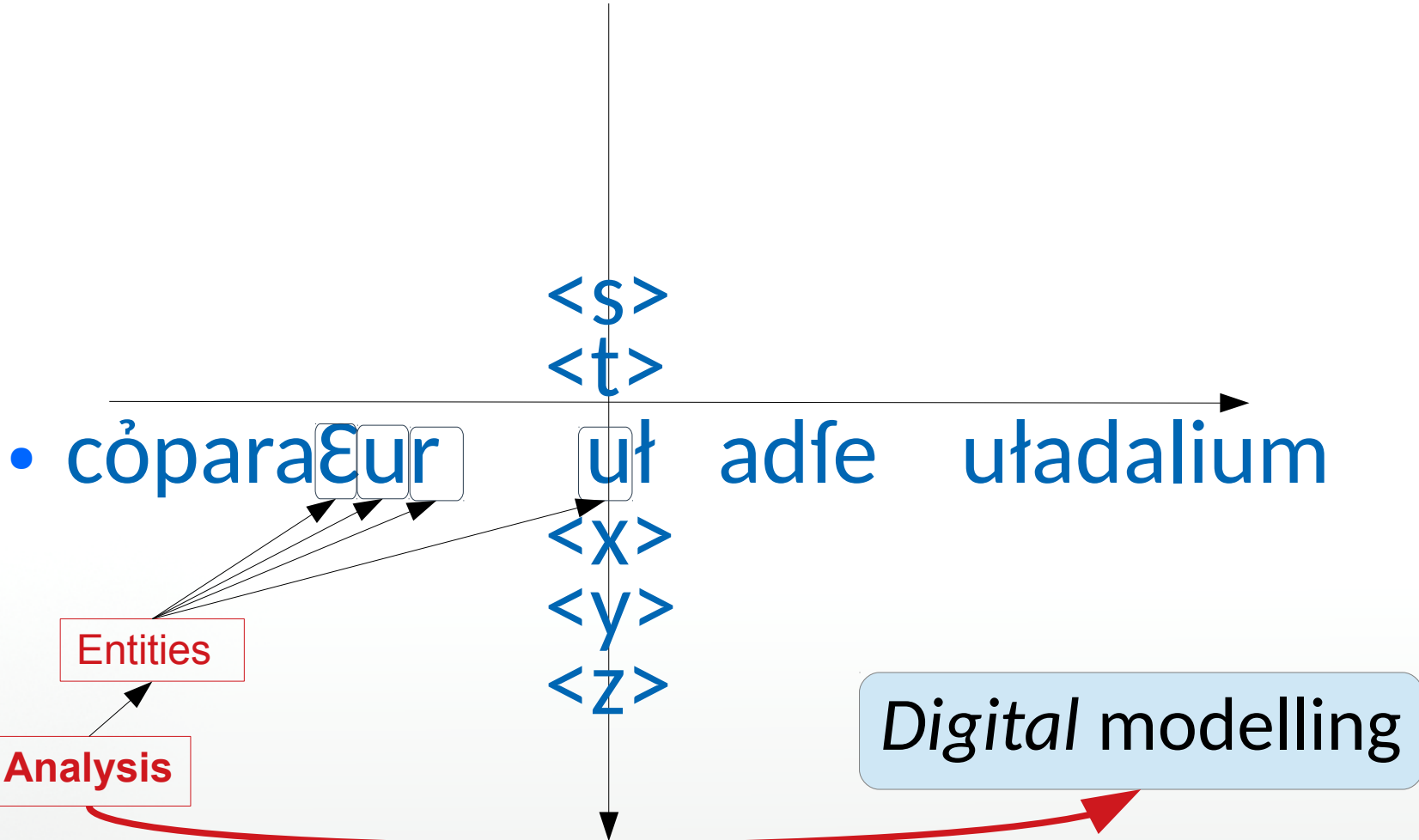


“If [we are] given anything..., it is the as yet unanalyzed *text* in its undivided and absolute integrity” (“**deduction**”, *Prolog. Ch. 4*)

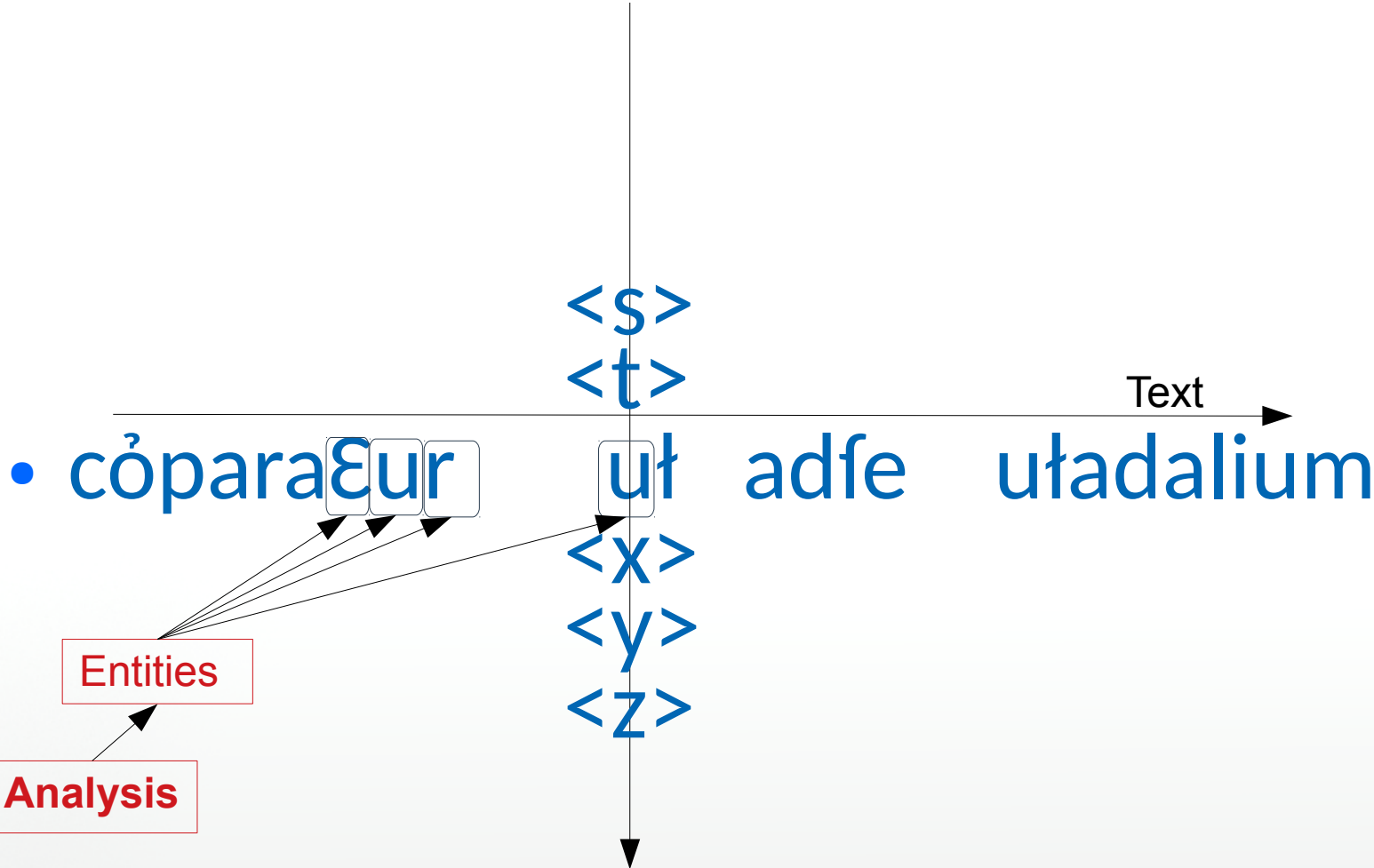
# Modelling/“analysis”: entities



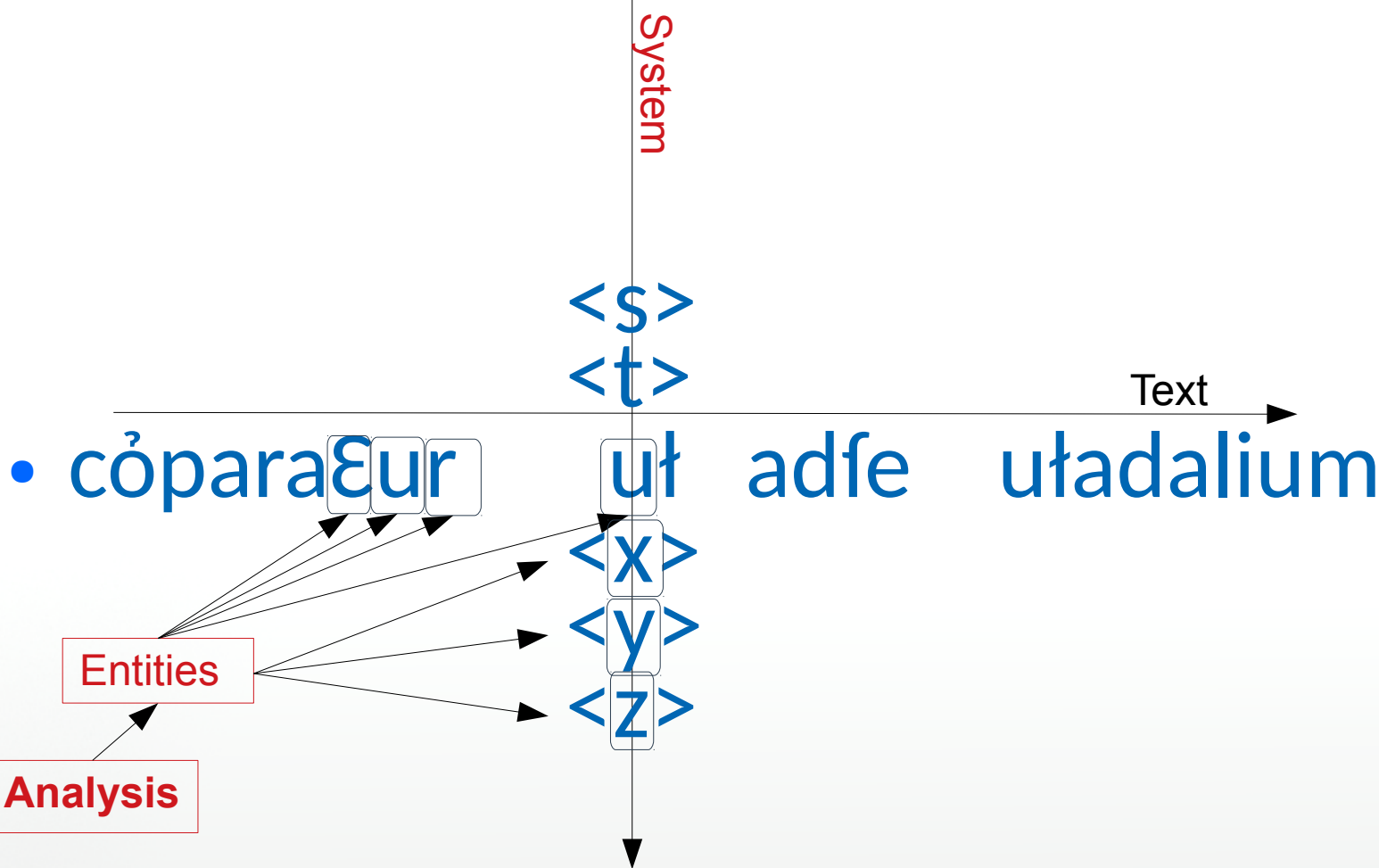
# Modelling/“analysis”: entities



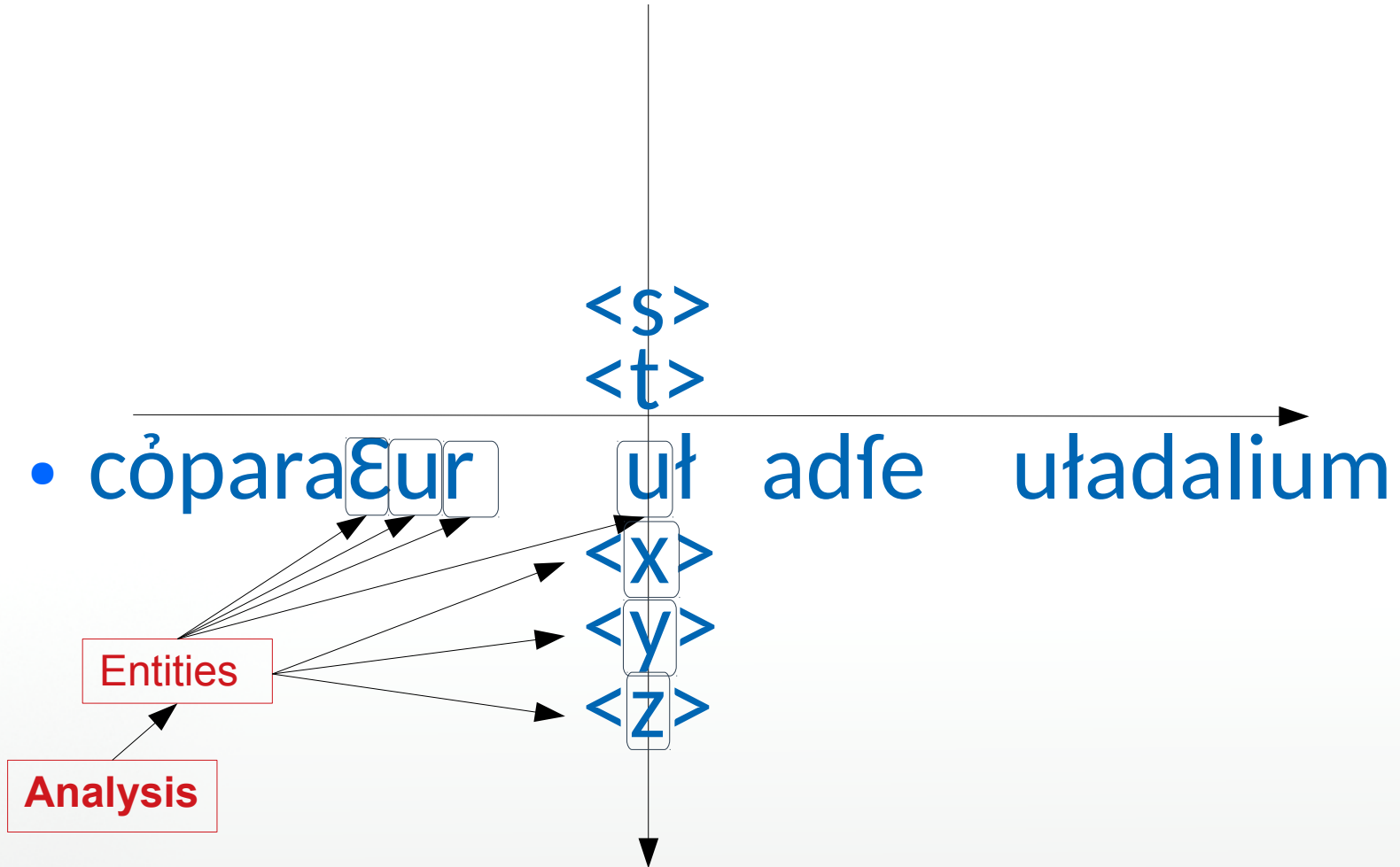
# Modelling/“analysis”: entities



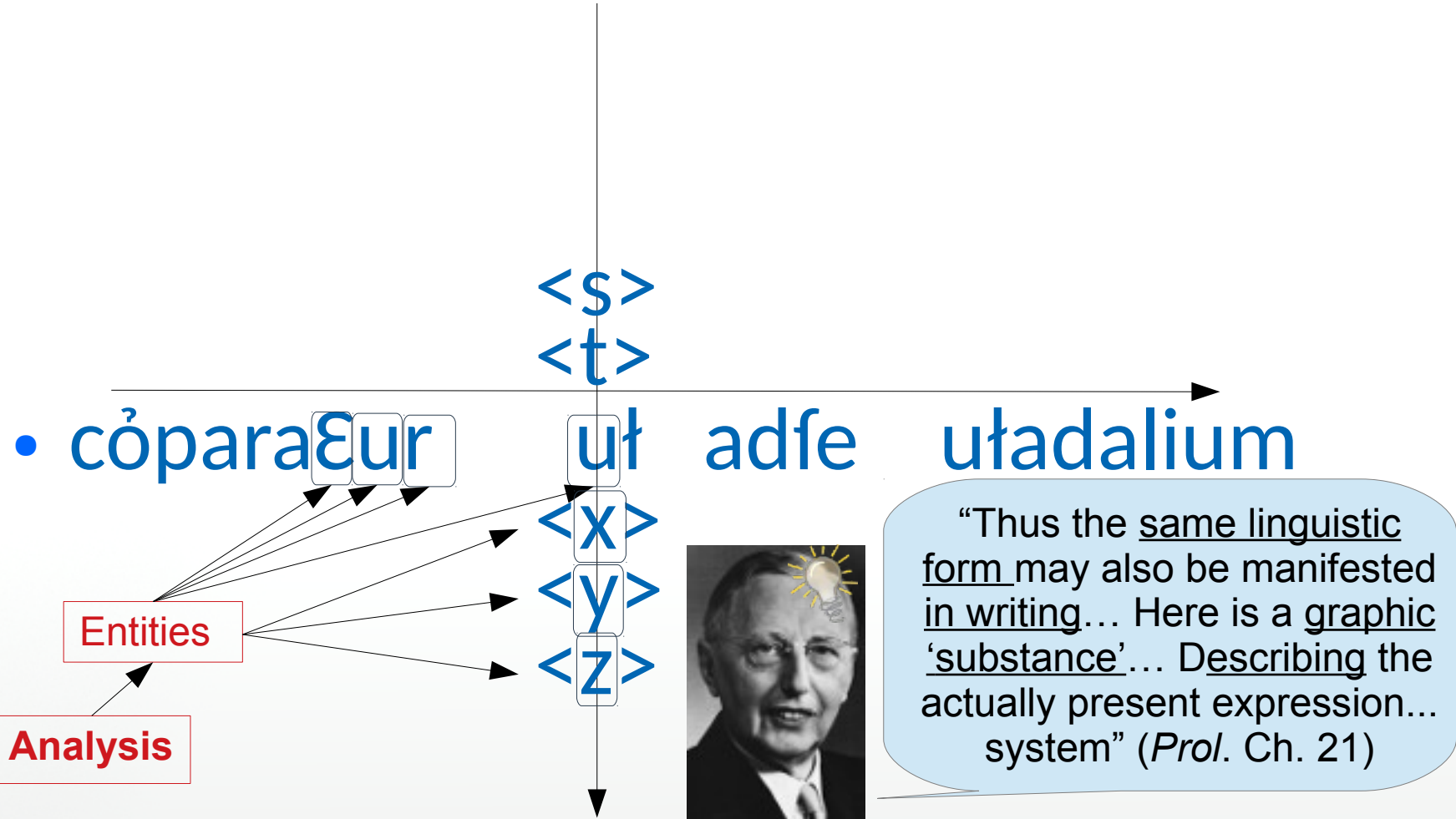
# Modelling/“analysis”: entities



# Graphemes as entities?

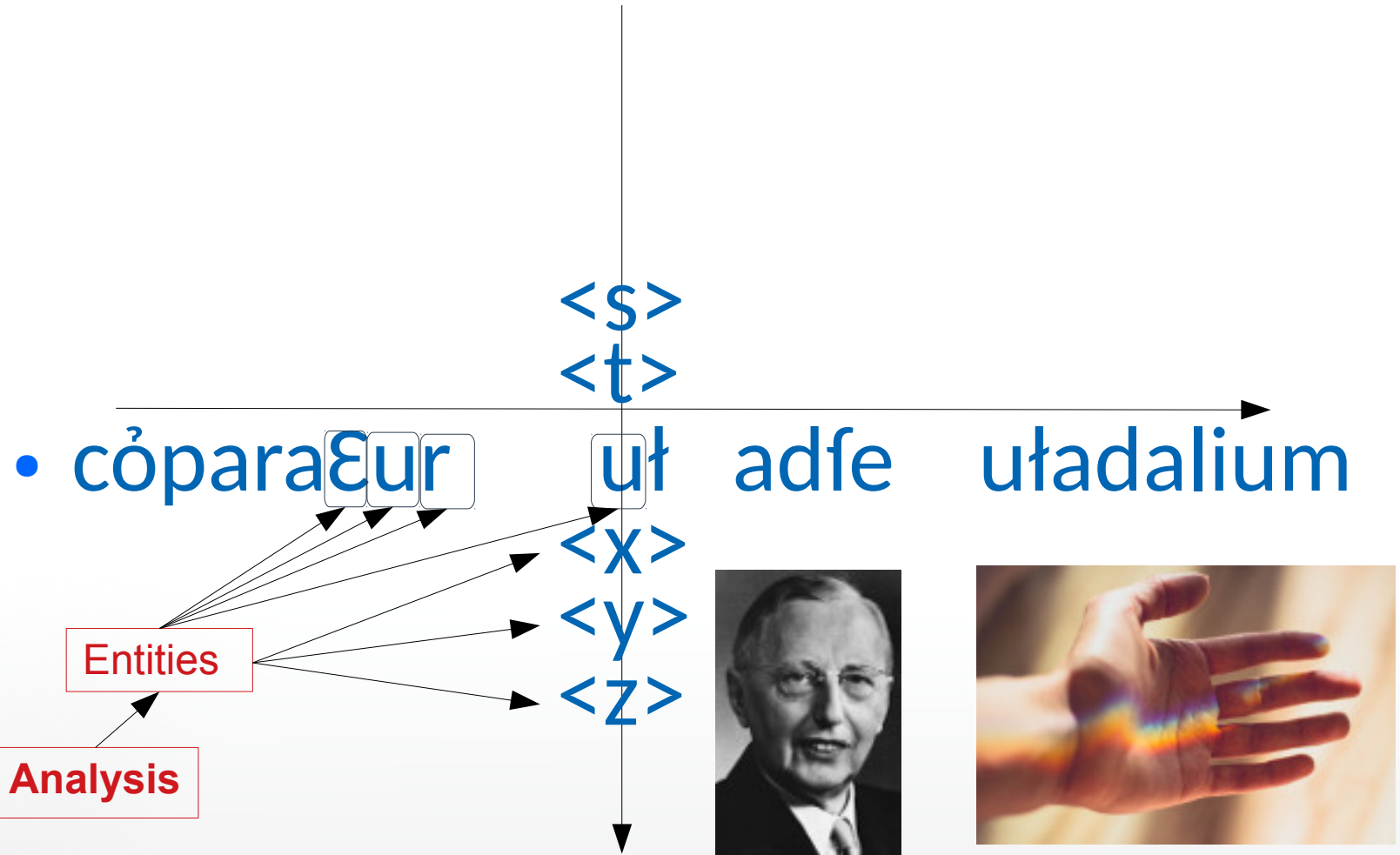


# Graphemes as entities?





# Graphemes as entities?



# Graphemes as entities?

- Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- Abbreviations

• còparaξur

<s>  
<t>

ut

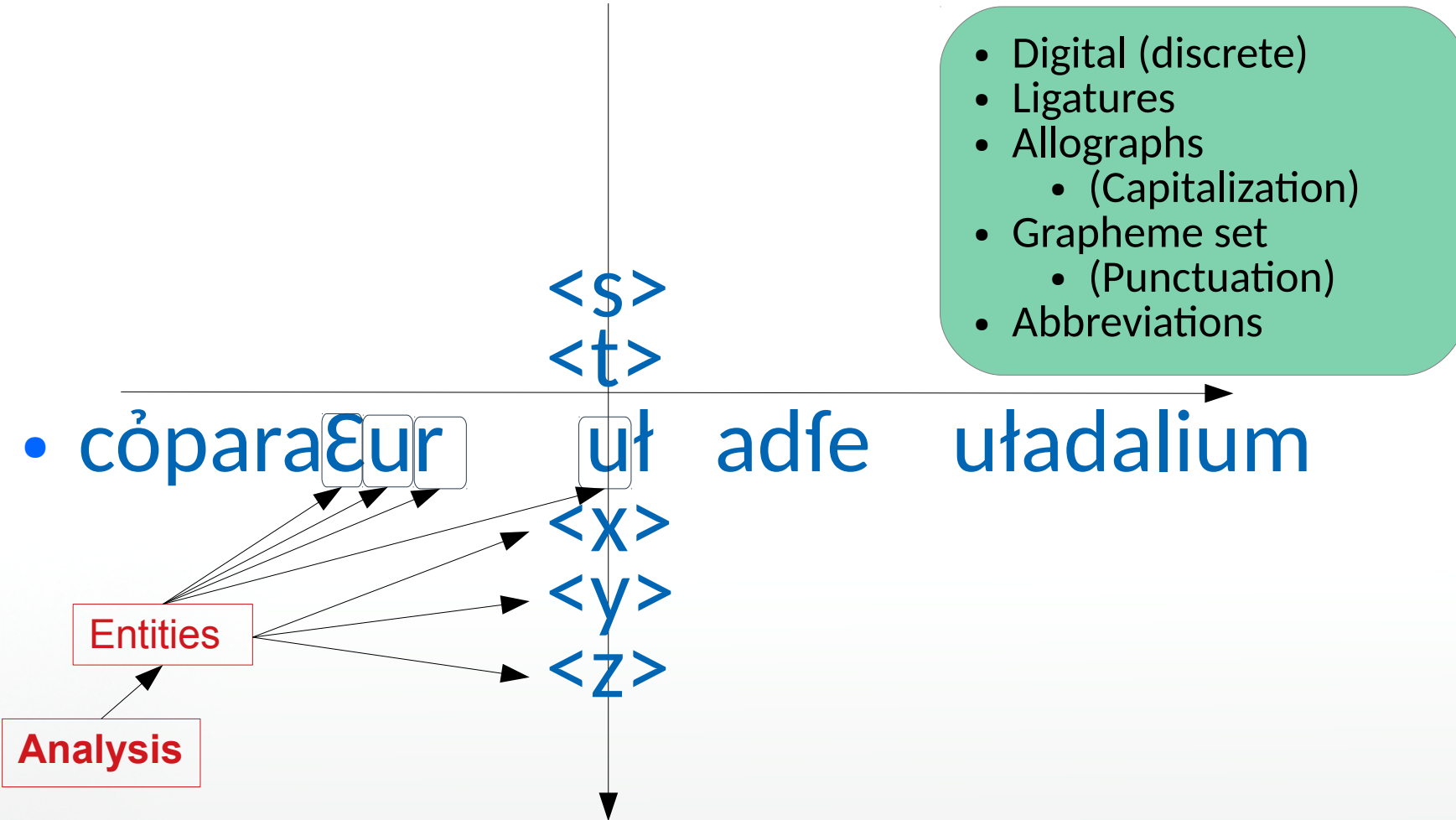
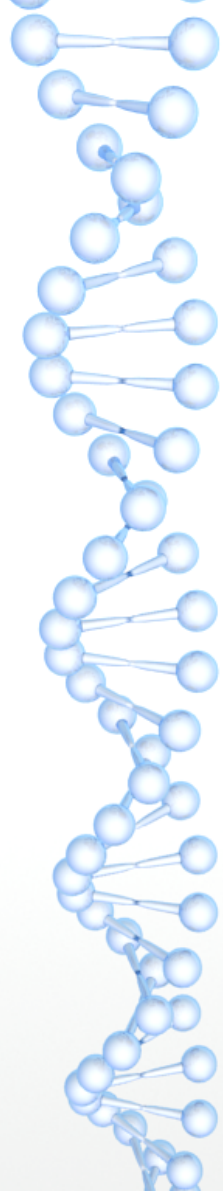
adfe

utadalium

Entities

Analysis

<x>  
<y>  
<z>



# Graphemes as entities?

- ✓ Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- Abbreviations

• còparaEUR    ut    adfe    utadalium

<s>

<t>

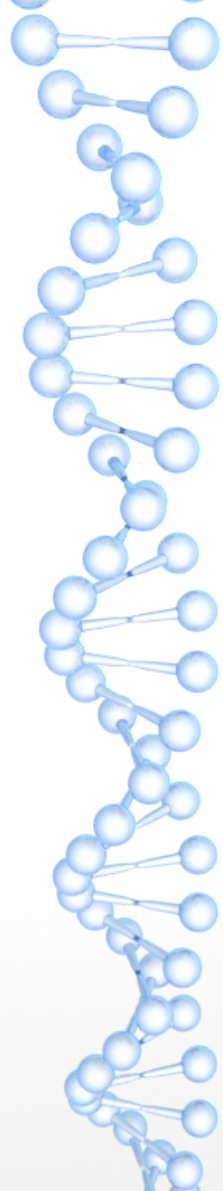
<x>

<y>

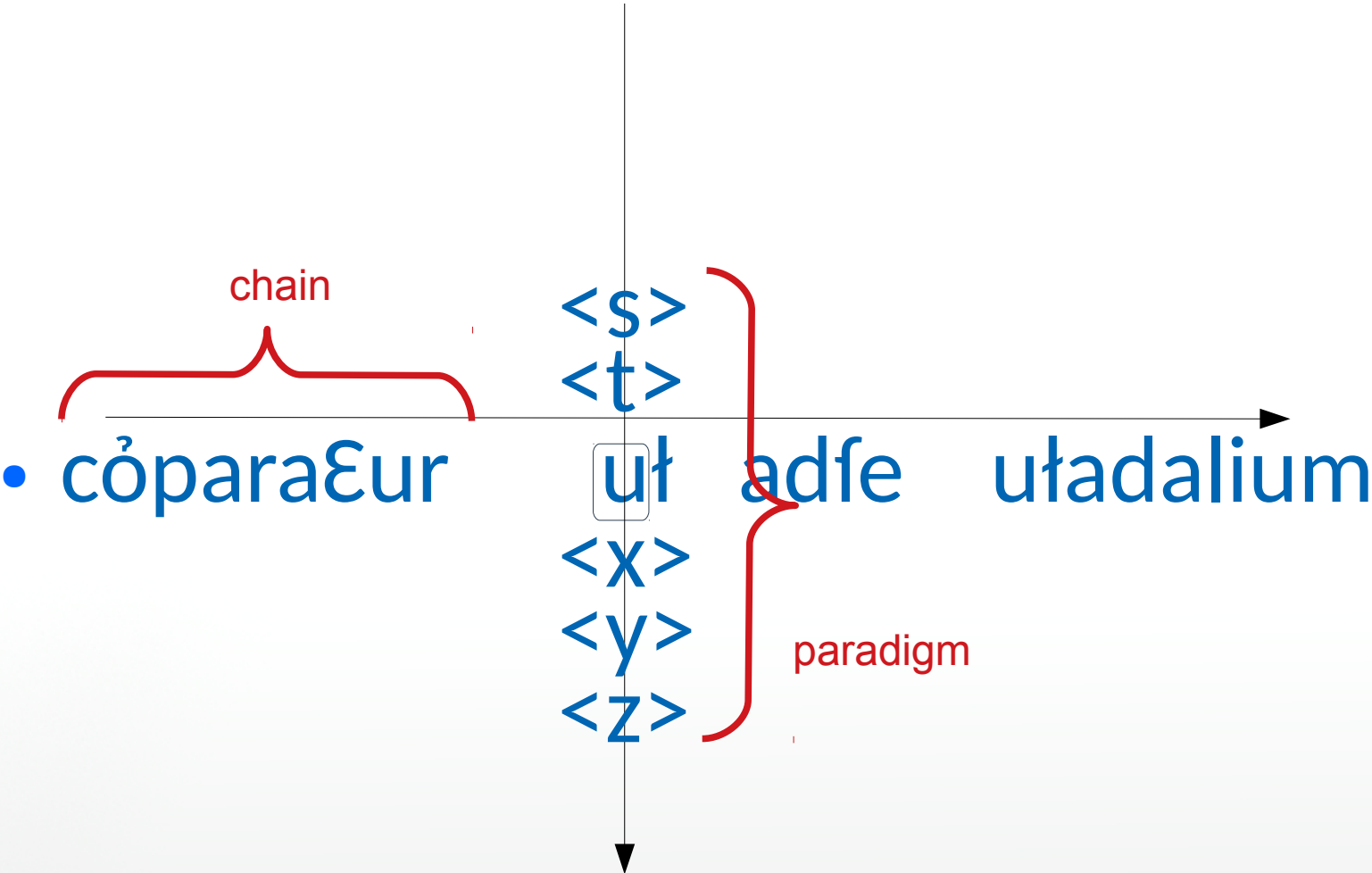
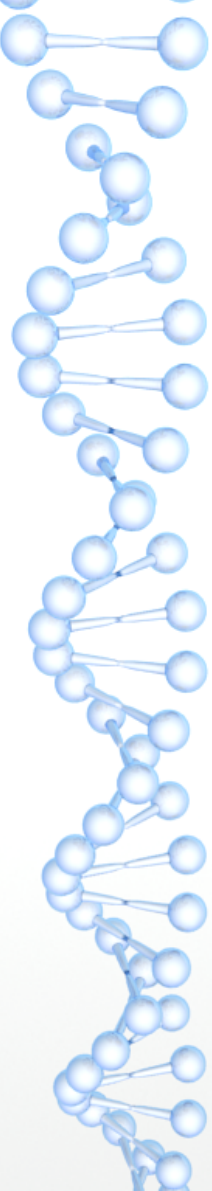
<z>

Entities

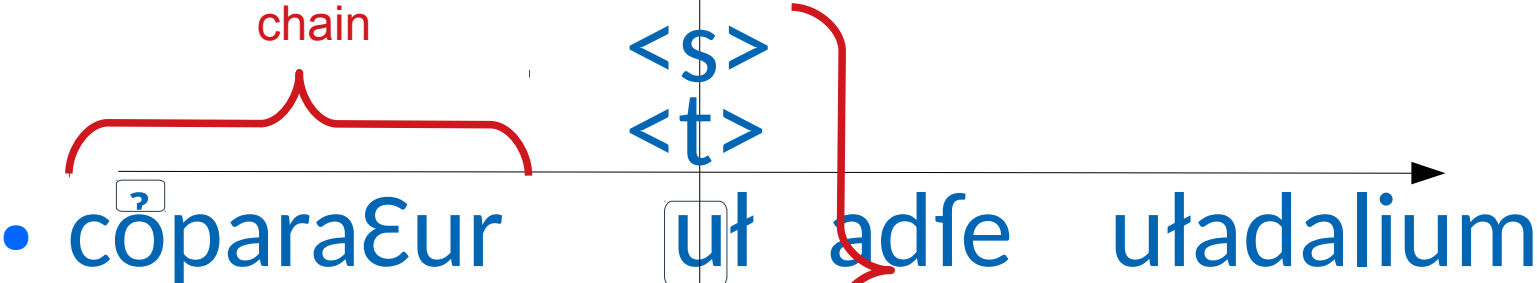
Analysis



# Chains and paradigms



# Chains and paradigms



“chain” → sequence?



# Functions



• còparaEur



Function

<s>

<t>

u†

adfe

u†adaliu

<x>

<y>

<z>

A white speech bubble with a dark blue background. The bubble has a rectangular top and a pointed bottom. The word "Ligatures" is centered inside the bubble in a bold, dark blue font.

**Ligatures**

# Ligatures



• còparaEur



Function

“Ligature”  
(entities, parts of a chain)

<s>

<t>

uł

adfe

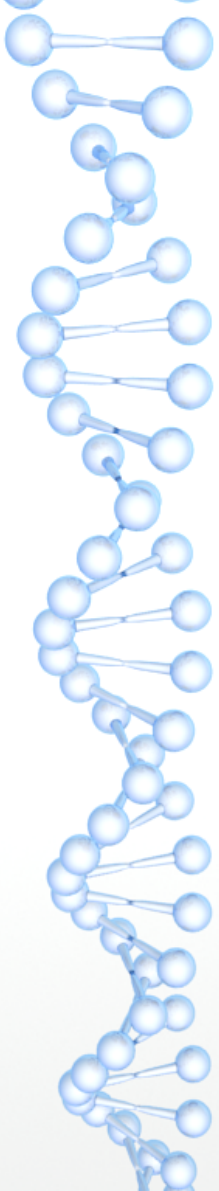
uładaliu

<x>

<y>

<z>





# Ligatures

Varieties

«σ»

<s>

<t>

• còparaεur

ut

adfe

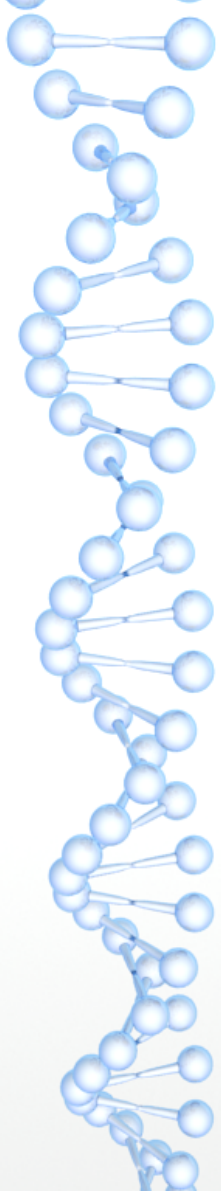
utadaliu

«√»

<x>

<y>

<z>



# Ligatures

Varieties:  
Solidal  
variants

«σ»

<s>

<t>

• còparaur

uł

adfe

uładalium

«√»

<x>

<y>

<z>

# Ligatures

Varieties:  
Solidal  
variants

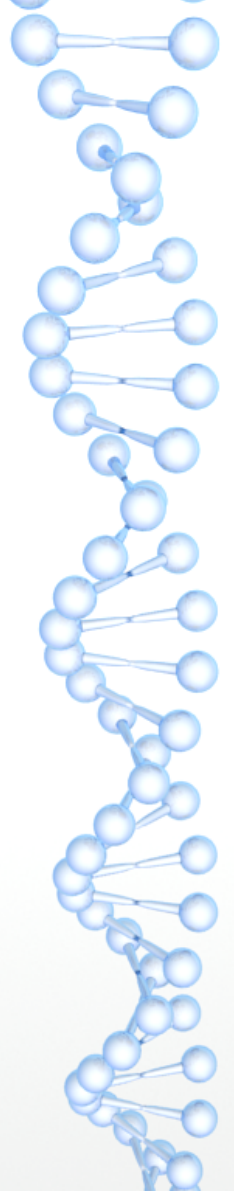
- ✓ Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- Abbreviations

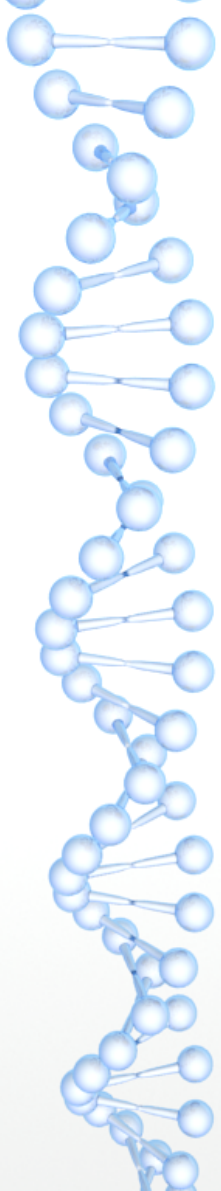
• còparaur

<s>  
<t>

ut adfe utadalium

<x>  
<y>  
<z>





# Ligatures

Varieties:  
Solidal  
variants

- ✓ Digital (discrete)
- ✓ **Ligatures**
  - Allographs
    - (Capitalization)
  - Grapheme set
    - (Punctuation)
  - Abbreviations

• còparaur

<s>

<t>

ut

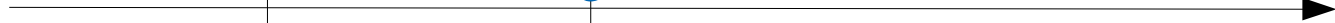
adfe

utadalium

<x>

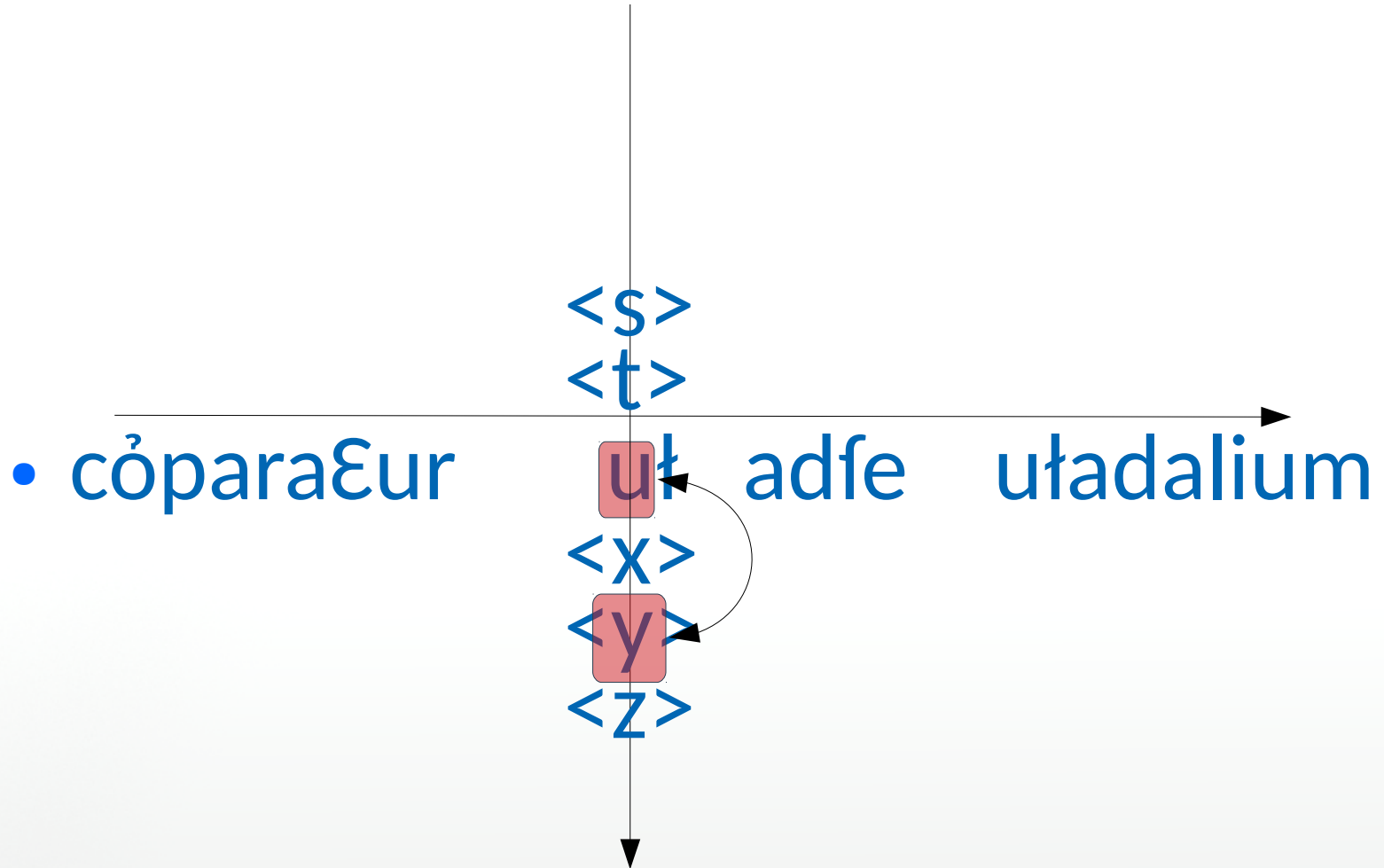
<y>

<z>

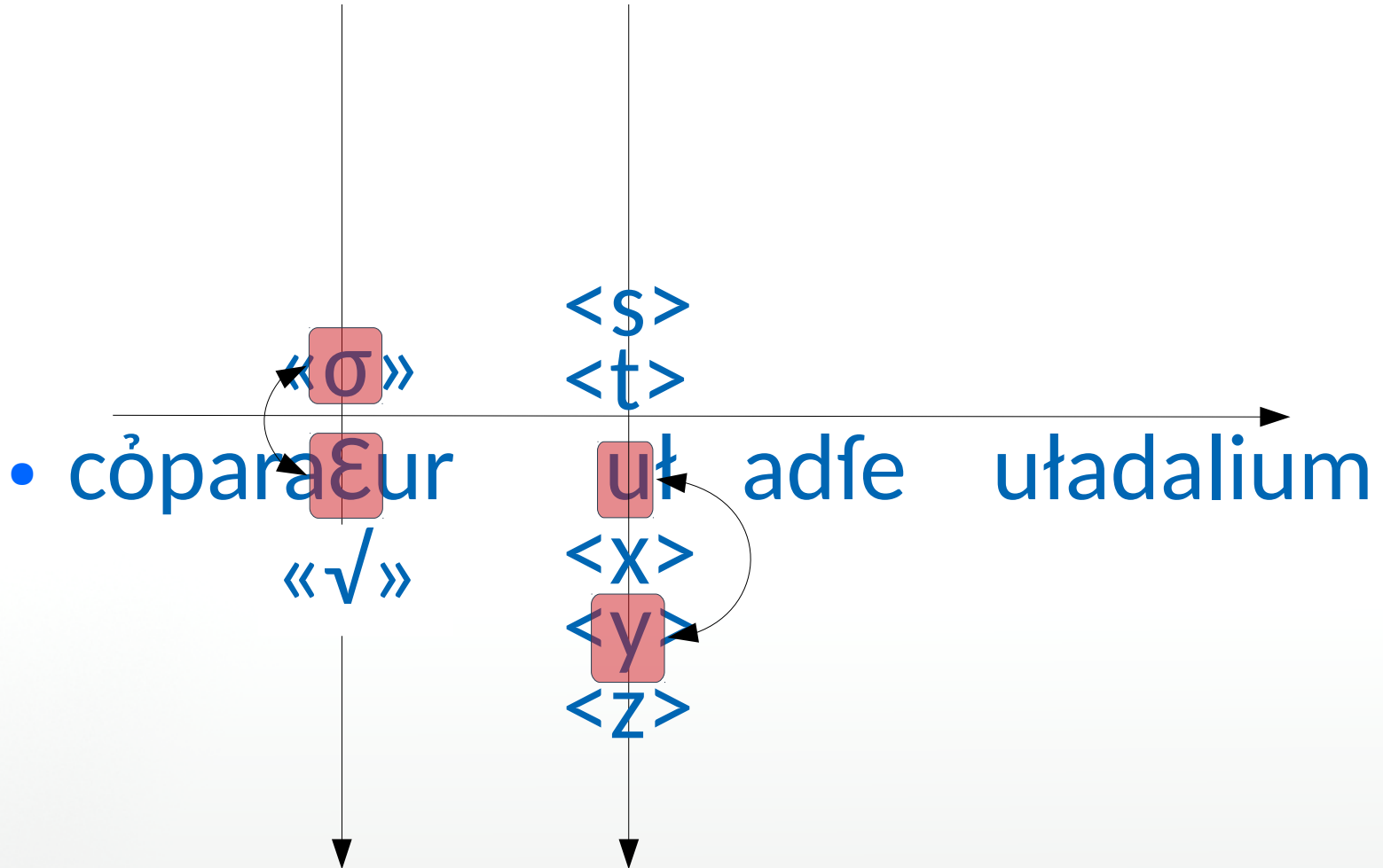


**Graphemes/allographs**

# The commutation test



# The commutation test



# The commutation test

• còparaEur

Function

**Substitution:**  
→ No change  
in “meaning”

daEur / daEur

«σ»

«√»

<S>  
<t>

ut

<x>  
<y>

<z>

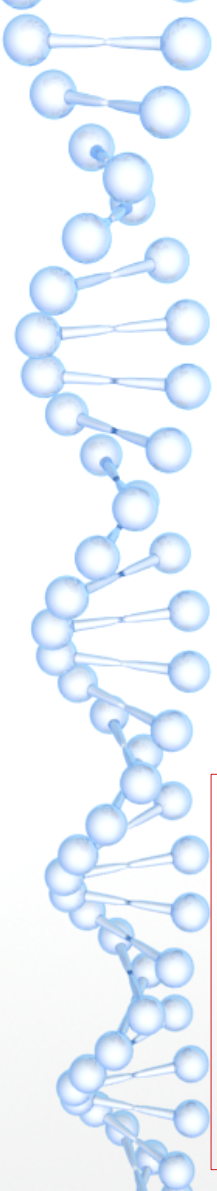
adfe

utadaliu

Function

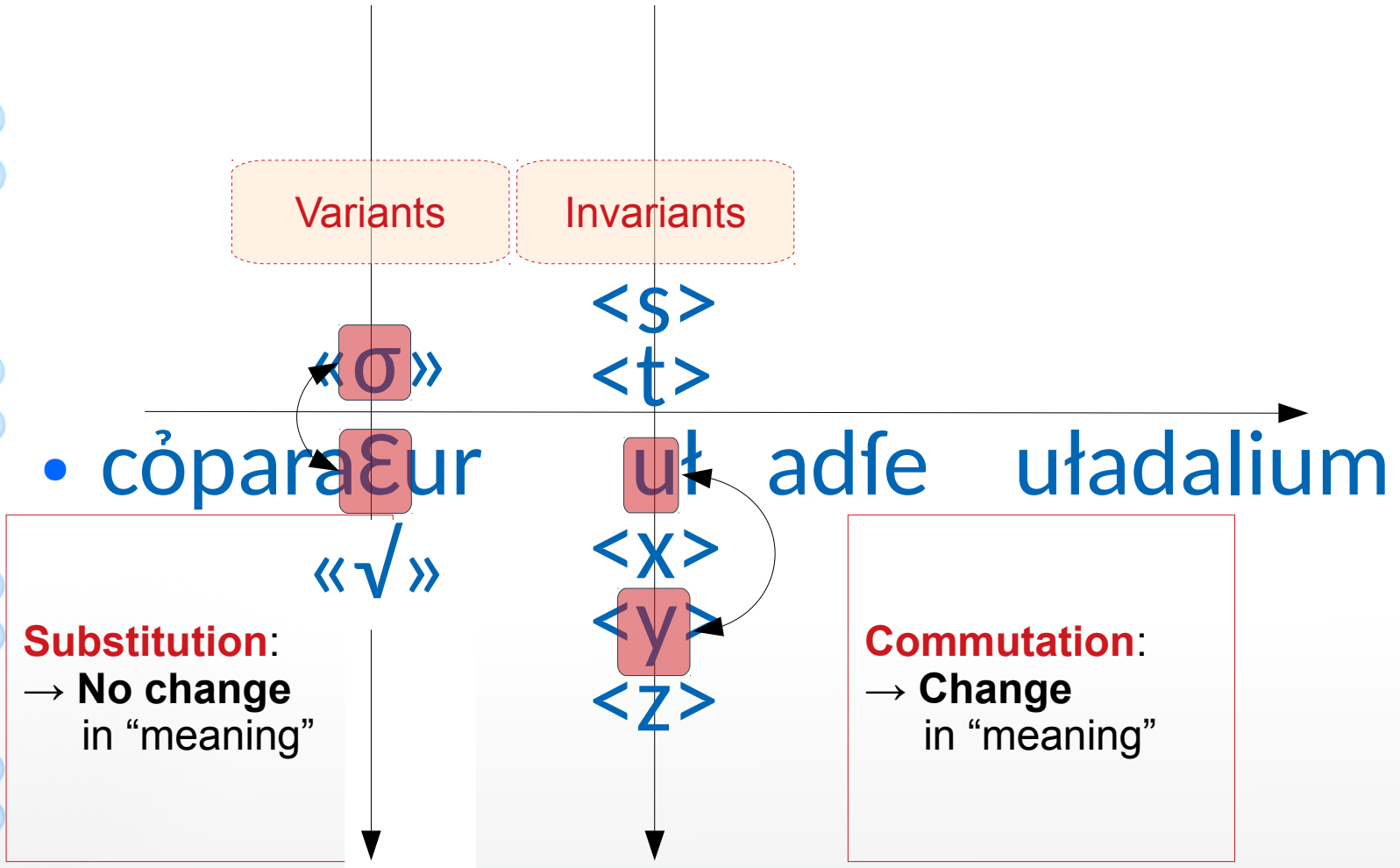
**Commutation:**  
→ Change  
in “meaning”

annus / annys

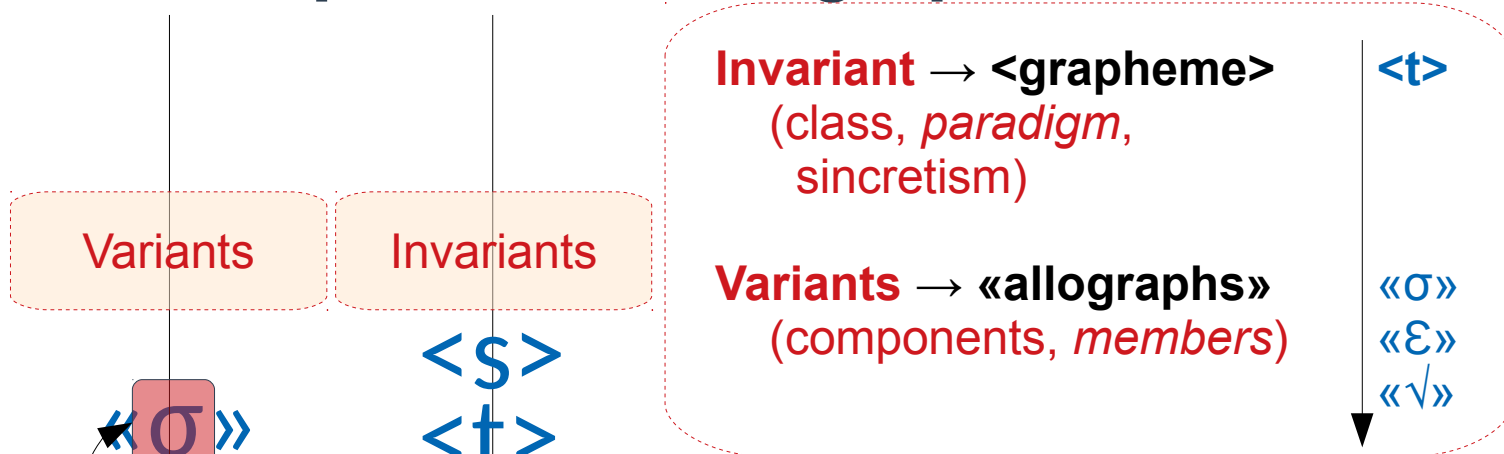
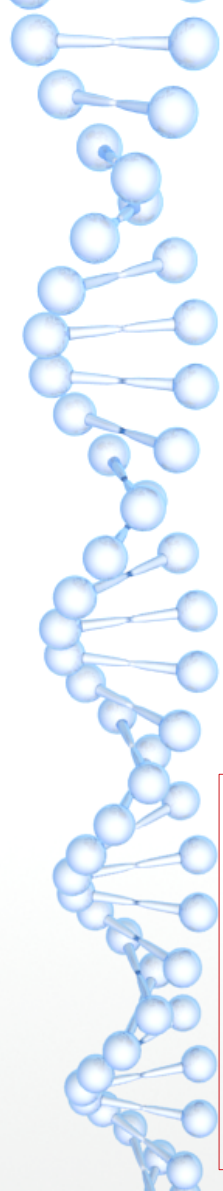




# The commutation test



# Graphemes / allographs



• còparaEur

<s>  
<t>

adfe

utadaliu

«σ»  
«√»

**Substitution:**  
→ No change  
in “meaning”

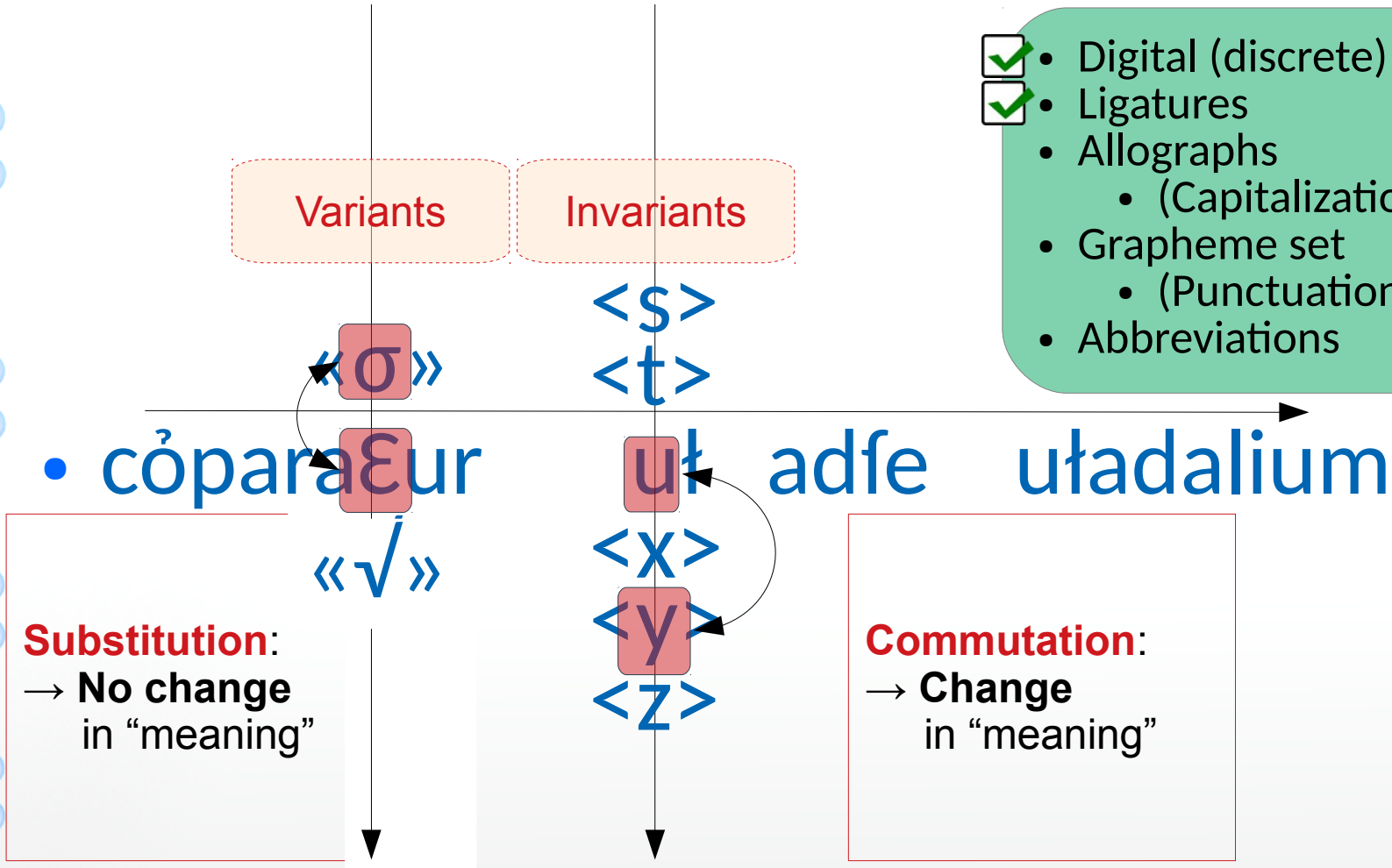
<x>  
<y>  
<z>

**Commutation:**  
→ Change  
in “meaning”

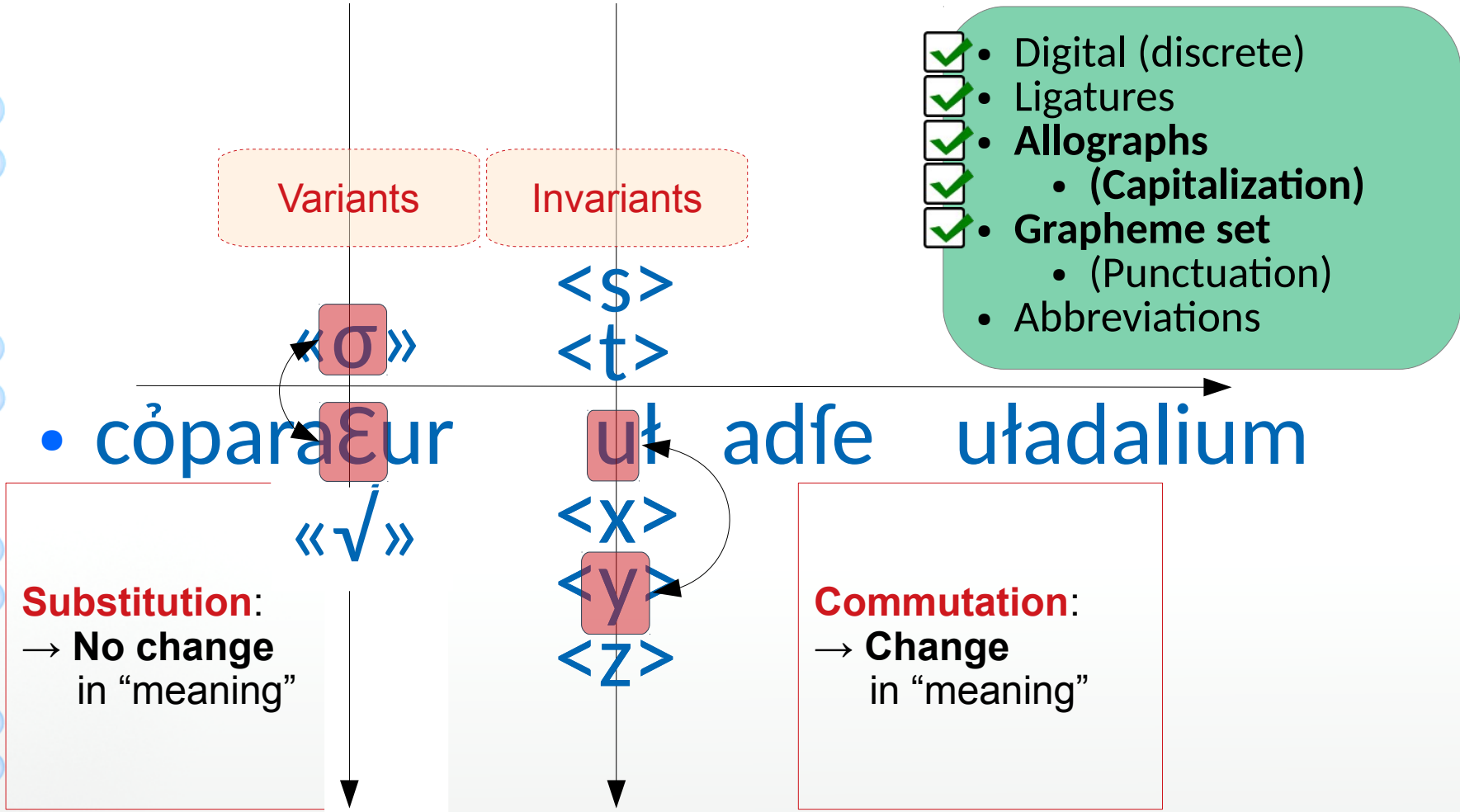
<t>  
«σ»  
«ε»  
«√»

# Graphemes / allographs

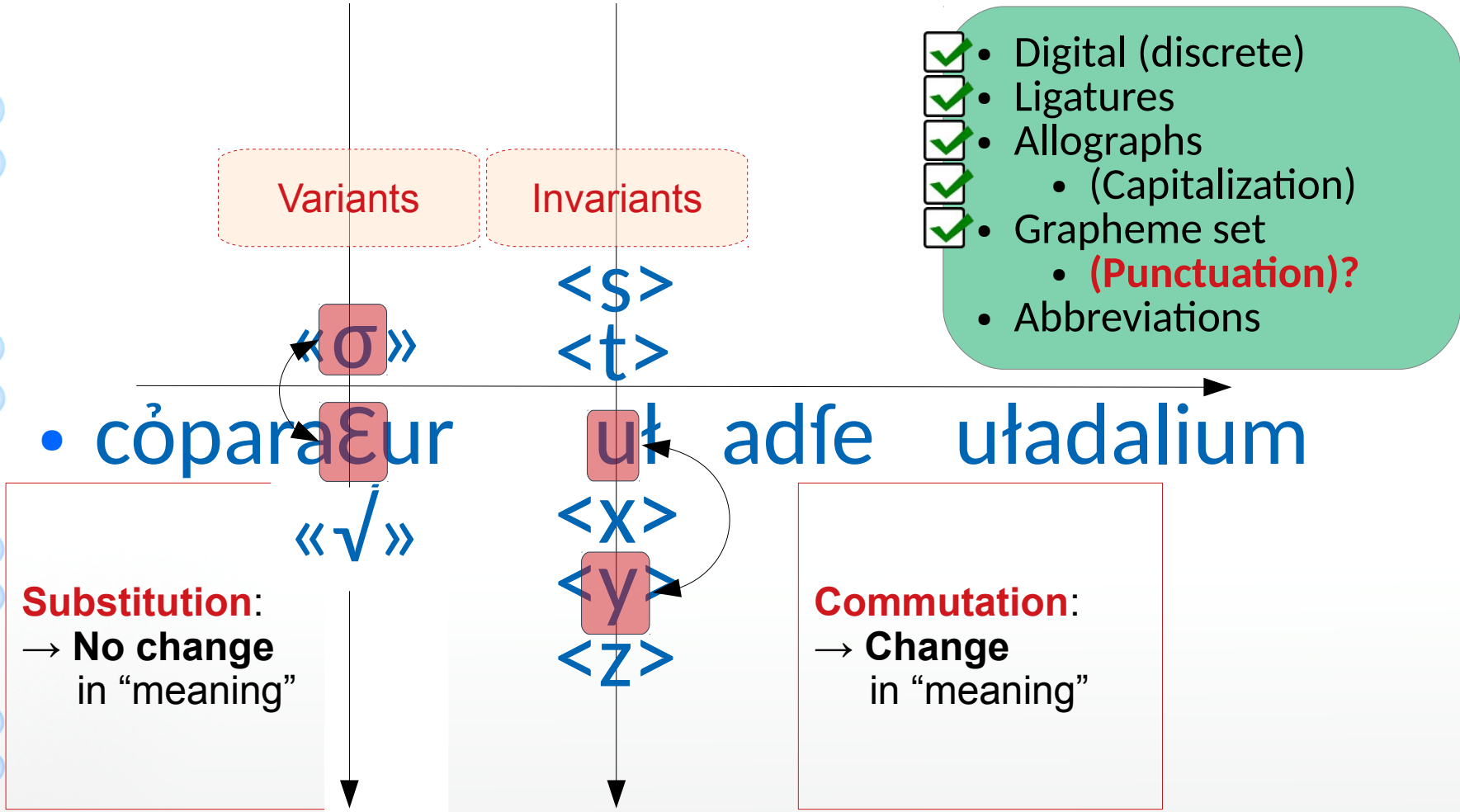
- ✓ Digital (discrete)
- ✓ Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- Abbreviations



# Graphemes / allographs



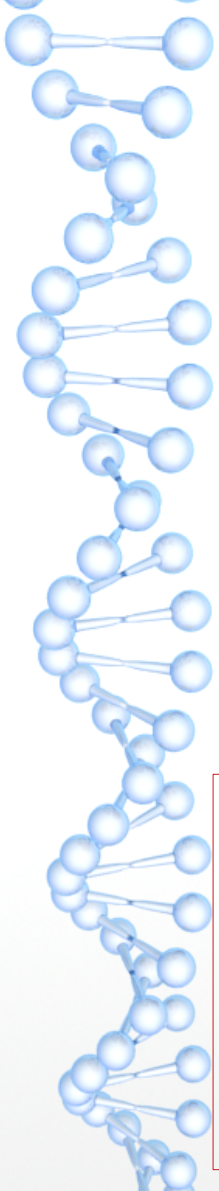
# Graphemes / allographs





**Punctuation**

# Punctuation



• còparaEur

«σ»

«√»

**Substitution:**  
→ No change  
in “meaning”

<S>  
<t>

uł

<x>

<y>

<z>

**Commutation:**  
→ Change  
in “meaning”

adfe

uładaliu



“content-form”  
(*Prol. Ch. 13*)

# Punctuation



“content-form”  
(*Prolog* Ch. 13)

Including larger  
units, such as  
sentences

• còparaEur

<S>  
<t>

ut adfe utadaliu

«σ»

u

«√»

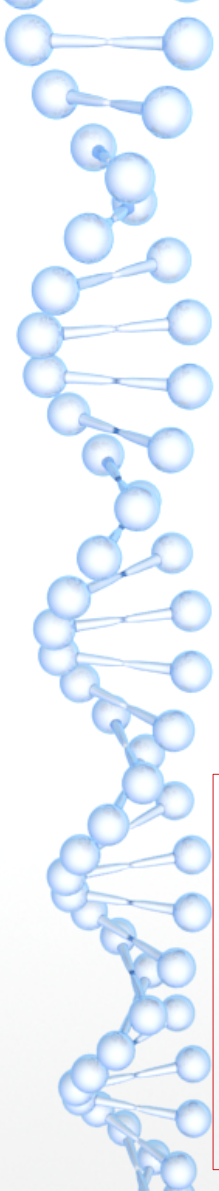
<x>

<y>

<z>

**Substitution:**  
→ No change  
in “meaning”

**Commutation:**  
→ Change  
in “meaning”





# Punctuation



“content-form”  
(*Prolog. Ch. 13*)

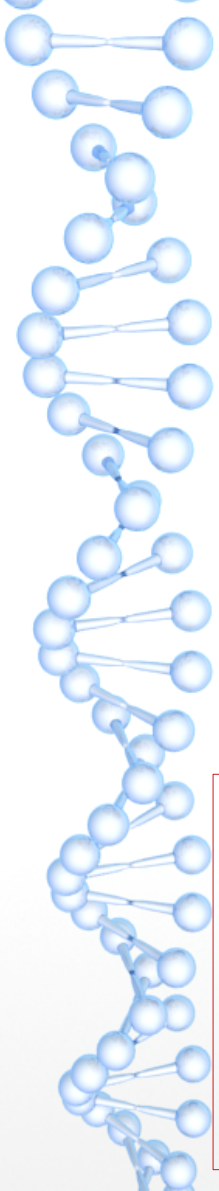
Including larger  
units, such as  
sentences

Truly I tell you, today you will be with me in paradise

Truly I tell you today, you will be with me in paradise

**Substitution:**  
→ No change  
in “meaning”

**Commutation:**  
→ Change  
in “meaning”



# Punctuation



“content-form”  
(*Prol. Ch. 13*)

Including larger  
units, such as  
sentences

Truly I tell you, today you will be with me in paradise

Truly I tell you today, you will be with me in paradise

**Substitution:**  
→ No change  
in “meaning”

**Commutation:**  
→ Change  
in “meaning”

# Punctuation

• còparaEur

«σ»

«√»

**Substitution:**  
→ No change  
in “meaning”

<s>  
<t>

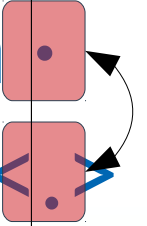
ut

<x>  
<y>  
<z>

adfe

utadaliu

**Commutation:**  
→ Change  
in “meaning”



# Punctuation

- Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- Abbreviations

«σ»

<s>  
<t>

• cōparaεur

ut

adfe

utadaliu

«√»

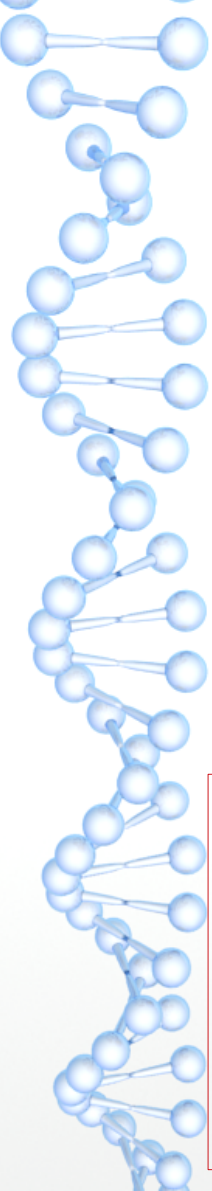
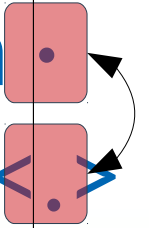
<x>

<y>

<z>

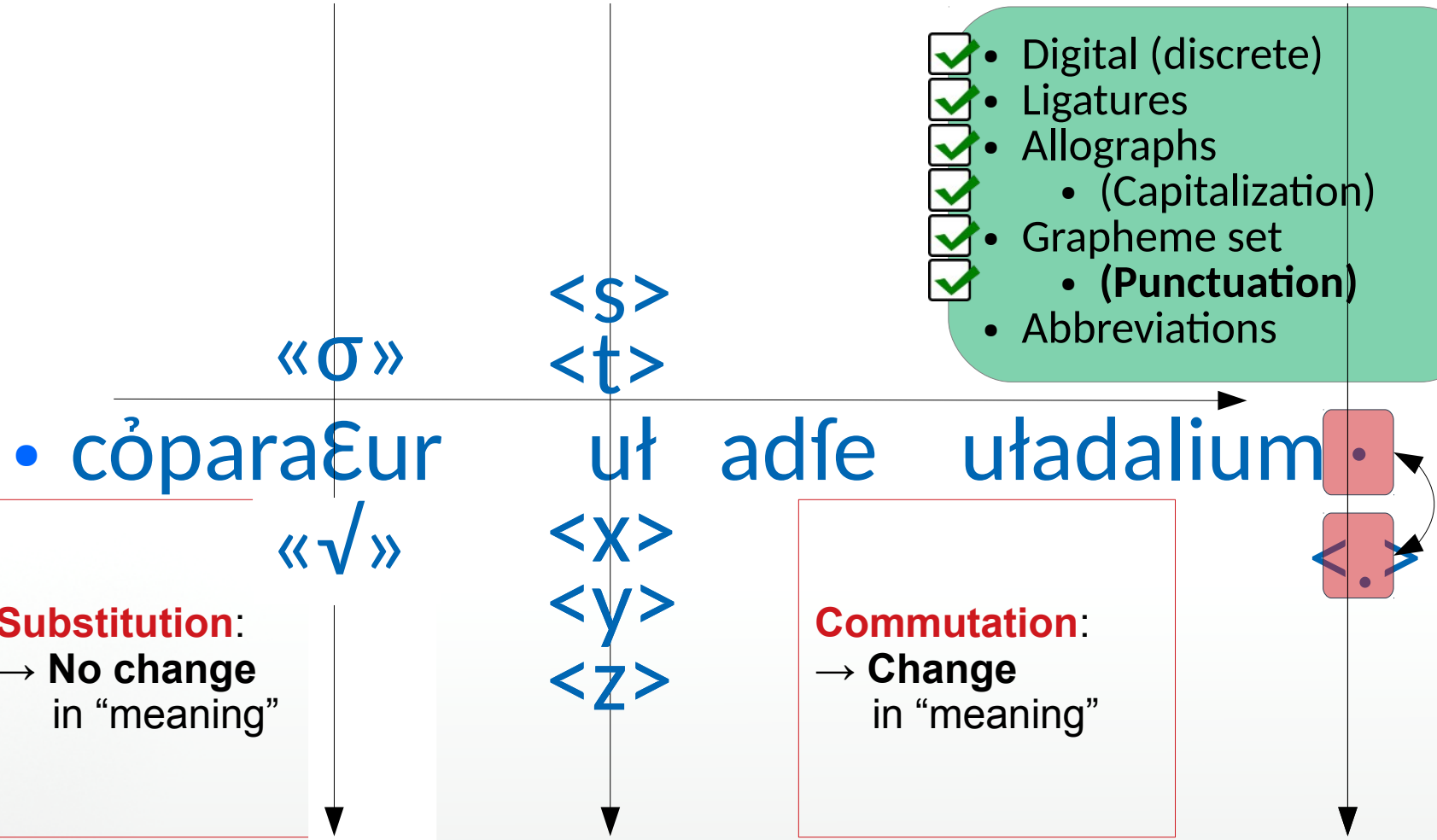
**Substitution:**  
→ No change  
in "meaning"

**Commutation:**  
→ Change  
in "meaning"



# Punctuation

- Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - **(Punctuation)**
- Abbreviations





# Abbreviations

# Abbreviations

-    comparatur ut adfe utadaliu

# Abbreviations: one grapheme?

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)

<O>

<Ô>

<Ō>

• cōparaEUR

ut adfe

utadadium

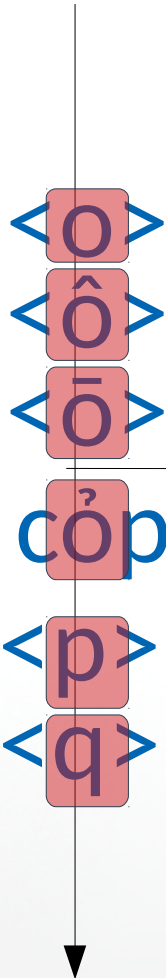
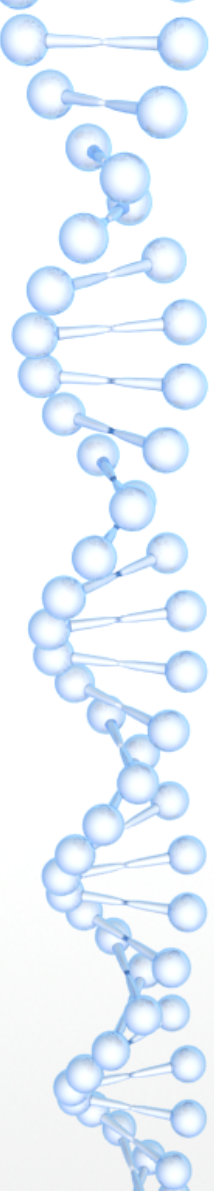
<p>

<q>





# Abbreviations: one grapheme?



• cōparaεur

ut adfe

utadaliu



Principles: **simplicity**,  
**economy**, **reduction**

(*Prolog*, Ch. 6)

“lowest possible number  
of elements” (Ch. 13)

# Abbreviations: function

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)

•  comparatur ut adfe utadaliu

Function

(entities)

# Abbreviations: function → ligature?

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)

•  comparatur      ut   adfe      utadalium

Function

“Ligature”  
(**Soidal** variants)



# Abbreviations: function → ligature?

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)

•  comparatur ut adfe utadaliu

Function

“Ligature”  
(**Soidal** variants)



**Soidal**: both mandatory

# Abbreviations: function → selection?

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)
  - 2B. **Selection** (one governs the other)

•  comparaEUR      ut   adfe      utadadium

**Function**

**Selection:**

- one is optional
- one governs the other



# Abbreviations: function → selection?

•  **comparatur**      **ut adfe**      **ut adalium**

**Function**

**Selection:**

- one is optional
- one governs the other



“chain” → sequence?

# Abbreviations: function → complementarity?

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)
  - 2B. **Selection** (one governs the other)
  - 2C. **Complementarity** (**interdependence**)

•  **comparaEur**      **ut adfe utadaliu**

**Function**

**Complementarity**  
(**interdependence** in a **system**)

# Abbreviations: function → complementarity?

•  comparatur      ut      adfe      ut adalium

Function

Complementarity  
(interdependence in a system)

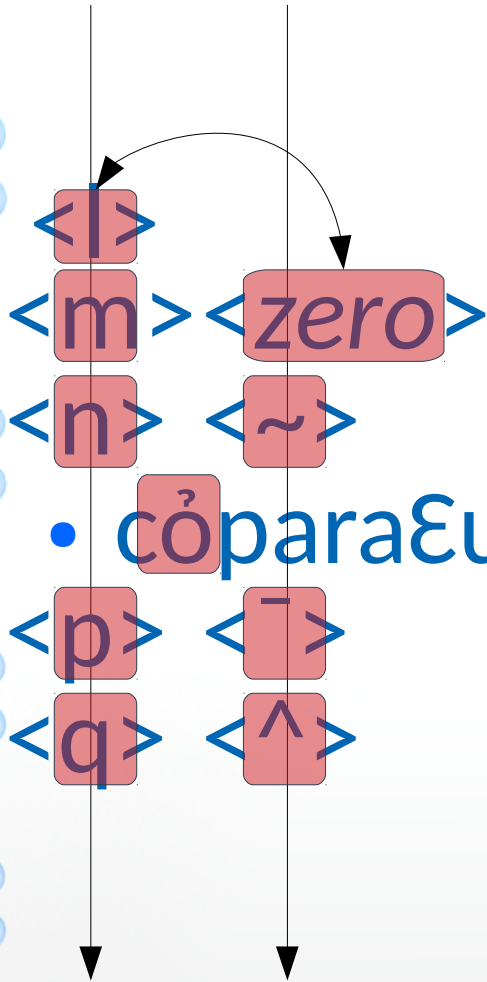
Example of  
complementarity

[case + gender + number]

Alt	-	<b>us</b>	nom+masc+sing
Alt	-	<b>orum</b>	gen+m/neu+plur
Alt	-	<b>arum</b>	gen+ fem + plur



# Abbreviations: function $\rightarrow$ complementarity?



•  $\langle \overset{?}{c} \rangle$  paraEur

ut adfe utadaliu



# Abbreviations

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)
  - 2B. **Selection** (one governs the other)
  - 2C. **Complementarity** (**interdependence**)

• **c**ōparaεur      uł    adfe      uładaliu

# Abbreviations

- Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- **Abbreviations**

• ċōparaεur

uł adfe

uładaliu

# Abbreviations

- Digital (discrete)
- Ligatures
- Allographs
  - (Capitalization)
- Grapheme set
  - (Punctuation)
- **Abbreviations**

• c?ōparaεur      uł   adfe      uładaliu



**Open issues**

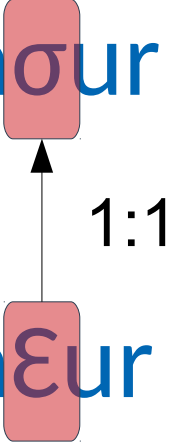


# Issues: 1. Abbreviations → functions

1. One **entity**
  - Whole abbreviation = grapheme (**invariant**)
2. Two **entities** (**function**)
  - 2A. Ligature (**solidal variants** in a **chain**)
  - 2B. **Selection** (one governs the other)
  - 2C. **Complementarity** (**interdependence**)

• **c**ōparaεur      uł   adfe      uładaliu

## Issues: 2. Abbreviations → one:many

- còparaσur
  - còparaϵur
- 



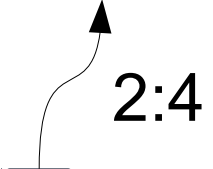
## Issues: 2. Abbreviations → one:many

- **com**paraσur
  - **c<sup>?</sup>o**paraεur
- 2:2 (2 ≠ 2)

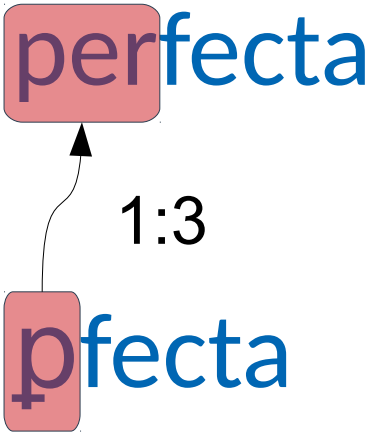




## Issues: 2. Abbreviations → one:many

- **prae**positio
  - **p̄**positio
- 2:4
- 

## Issues: 2. Abbreviations → one:many

- **perfecta**
  - **p<sub>♀</sub>fecta**
- 1:3
- 
- A diagram illustrating a one-to-many relationship between two abbreviations. The top item is 'perfecta', where the 'p' is enclosed in a red rounded rectangle. The bottom item is 'p
- <sub>♀</sub>
- fecta', where the 'p' with a female symbol is enclosed in a red rounded rectangle. A curved arrow points from the 'p
- <sub>♀</sub>
- ' box to the 'p' box, with the text '1:3' positioned next to the arrow.

## Issues: 2. Abbreviations → one:many



- **perfecta**

Alphabemes  
(alphabetical letters)

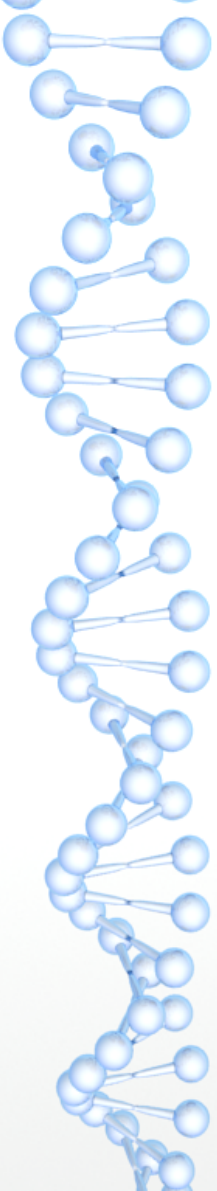
1:3

- **p**fecta

Graphemes

# Issues: 3. Ligatures

(syntagmatic)



# Issues: 3. Ligatures



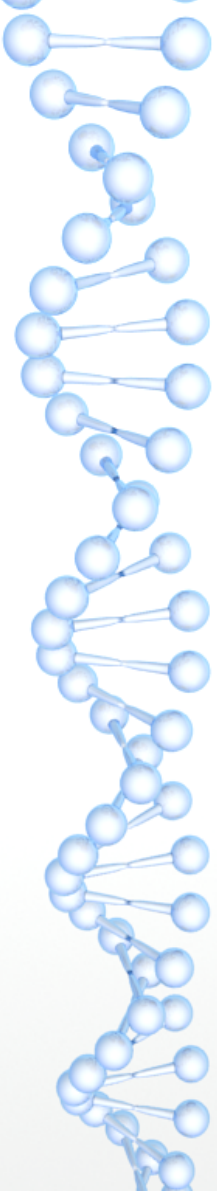
& (U+0026;  
ASCII 38)

# Issues: 3. Ligatures



& (U+0026;  
ASCII 38)

Historical/“etymological”  
considerations



# Issues: 3. Ligatures



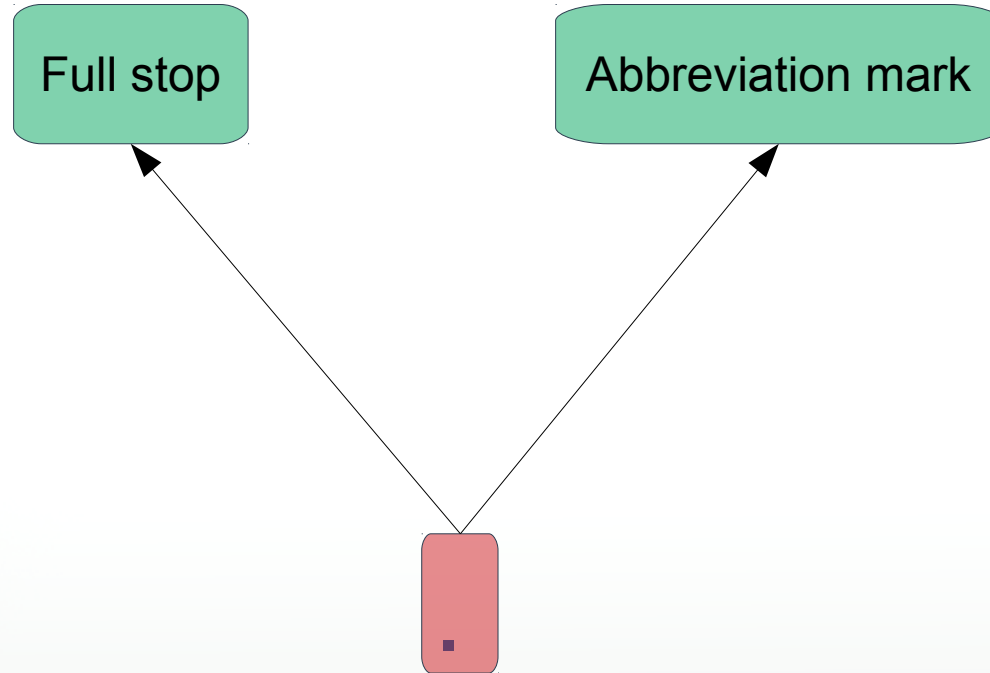
& (U+0026;  
ASCII 38)

Historical/“etymological”  
considerations



# Issues: 4. Grapheme definition

(paradigmatic)

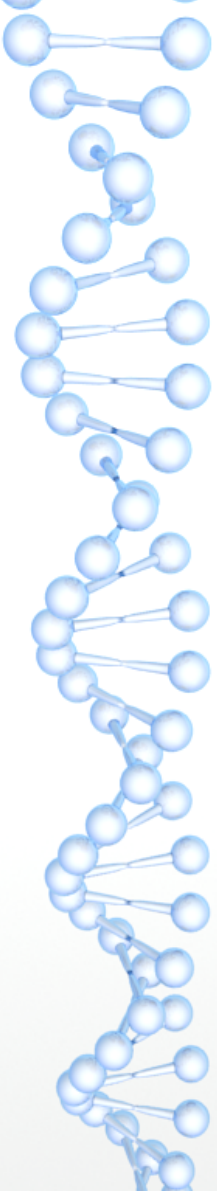
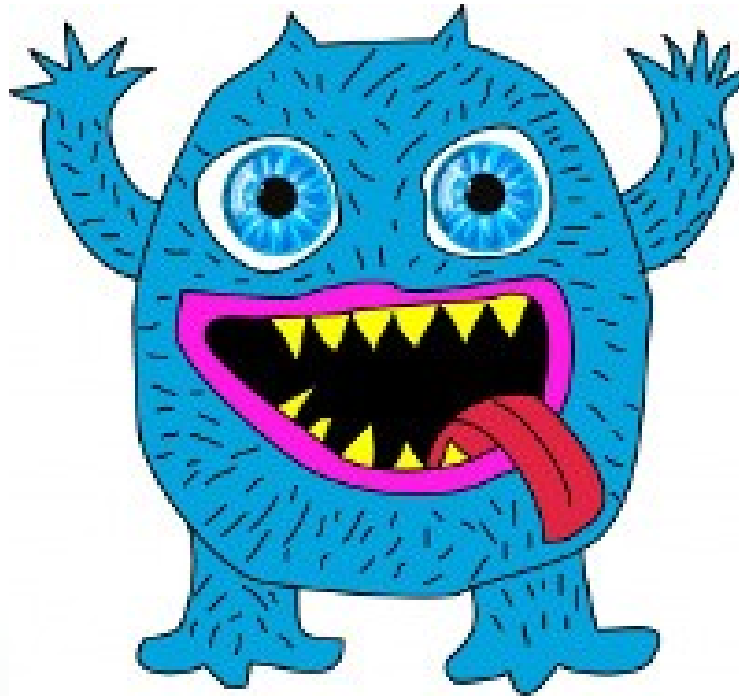






## Issues: 5. Allograph definition (metasemiology)

# Issues: 5. Allograph definition (metasemiology)



# Issues: 5. Allograph definition (metasemiology)

Allographs  
(variants)

« $\sigma$ »

• còparaEur

uł

adfe

uładaliu

« $\sqrt{\quad}$ »

# Issues: 5. Allograph definition (metasemiology)

« $\sigma$ »

« $\sigma$ »

« $\sigma$ »

« $\sigma$ »

« $\sigma$ »

« $\sigma$ »

• còparaEur

ut adfe

utadaliu

« $\sqrt{\quad}$ »

« $\sqrt{\quad}$ »

« $\sqrt{\quad}$ »

« $\sqrt{\quad}$ »



# Open issues

1. Abbreviations → functions
2. Abbreviations → one:many
3. Ligatures
4. Grapheme definition
5. Allograph definition (metasemiology)

**In the Tower of Babel**

# In the Tower of Babel: “diplomatic”

MS A

a b c d e f g h **i** l m n o p q r s t **u** z . ;



*Ceci n'est pas un linguiste.*

MS B

a b c d e f g h **i j** l m n o p q r s t **u v** z . , ; : !

# In the Tower of Babel: “normalized”

OCR from Teubner

a b c d e f g h **i** l m n o p q r s t **u** z . ;



*Ceci n'est pas un linguiste.*

OCR from Loeb

a b c d e f g h **i j** l m n o p q r s t **u v** z . , ; : !





# In the Tower of Babel

- Disposable home-made solutions (project-specific)
  - TEI: theory-agnostic
  - ***Diplomatic***: Modelling pre-modern writing systems
  - ***Normalized***: Normalization models / software





**Out of the Tower**

# The (long) way out of the Tower



**WANTED**

**DEAD OR ALIVE  
REWARD  
\$4,000**

- Scholarly discussion on **modelling**
- Documentation on project-specific modelling
  - formal (data models, software code, tables)
  - prose
- Shared models
- Reusable **software** libraries



# Is Unicode the short way out (as TEI says)?

- Solution for new digital texts
- Not enough for pre-modern writing systems
  - Ligatures
    - **&** (U+0026; ASCII 38)
    - Have I encoded that it is equivalent to “e + t” in that MS?
  - Allographs
    - **ſ** (U+017F) / **s** (U+0073; ASCII 115)
    - Have I encoded that they are variants of grapheme <s>?
  - Grapheme set
    - **u** (U+0075; ASCII 117)
    - Have I encoded whether it “covers” (or not) <u> *and* <v>?
  -



**In a nutshell**



# In a nutshell

- Digital scholarly editions: diplomatic/normalized
- Hjelmslev's "analysis" as digital modelling for pre-modern writing systems
- Open issues
- Interoperability through modelling