

Paolo Monella

paolo.monella@unipa.it

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



University of Palermo, Department *Cultura e Società*
December 17, 2019

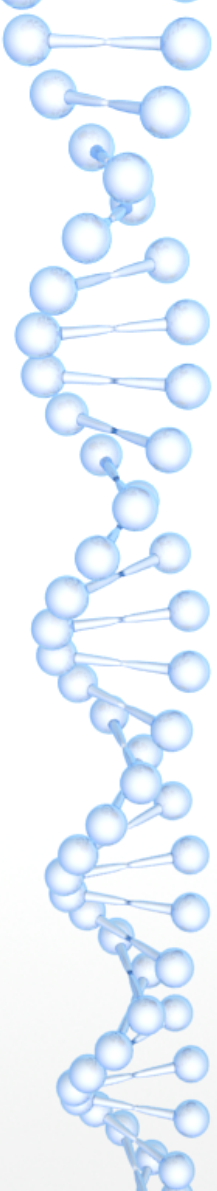




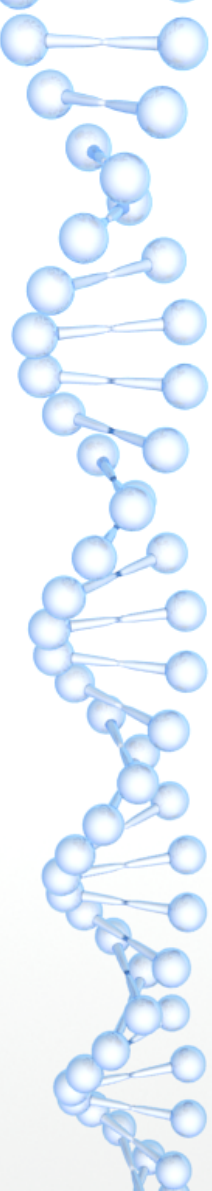
The interoperability issue

The interoperability issue

uenerú



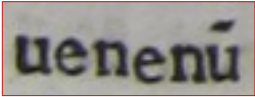
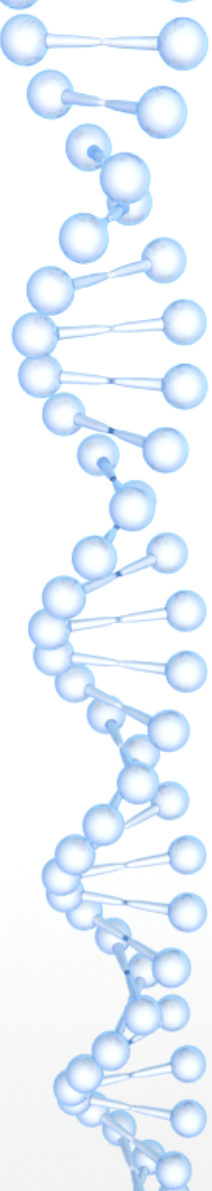
The interoperability issue



uēnēnū

- uēnēnū

The interoperability issue



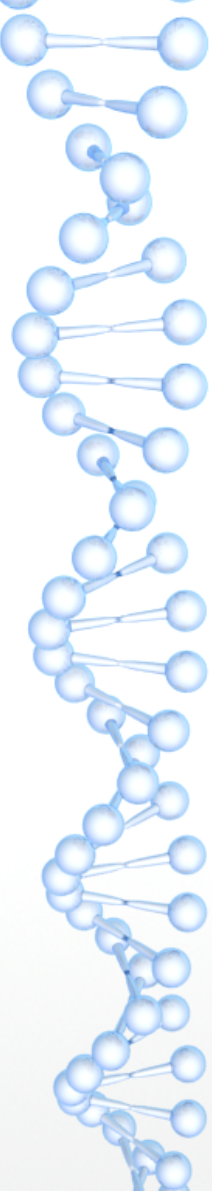
uenenū

- uenenū

Diplomatic

- Historical documentation
- Visualization

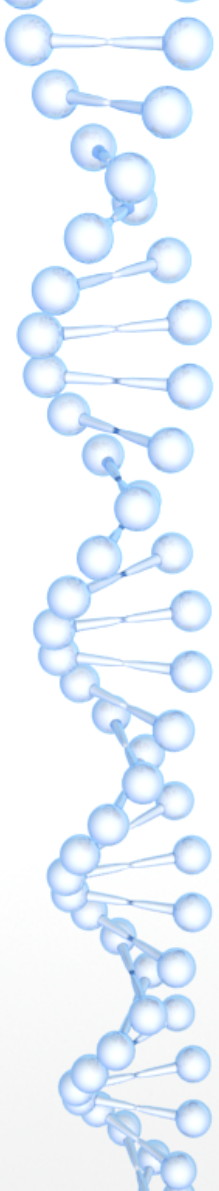
The interoperability issue



uēnenū

- uēnenū

The interoperability issue



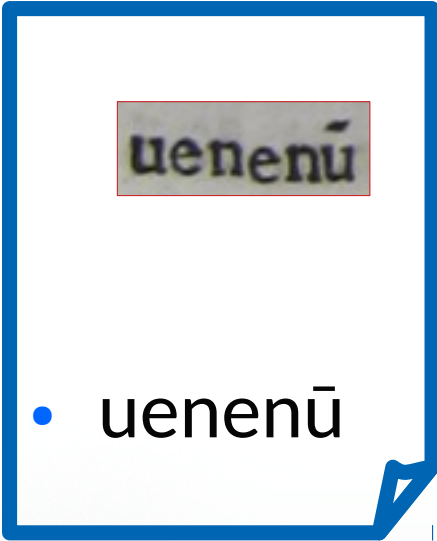
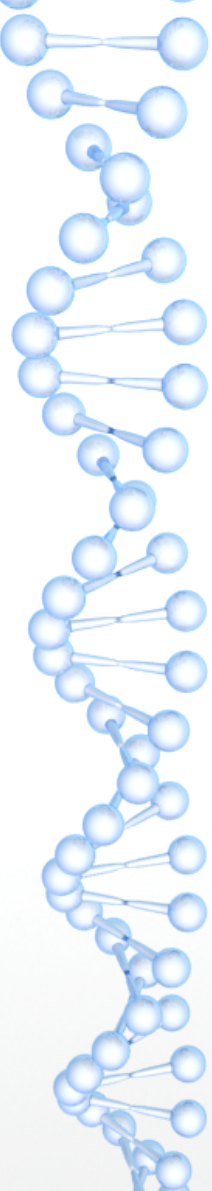
uenenū

- uenenū

- venenum

The interoperability issue

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



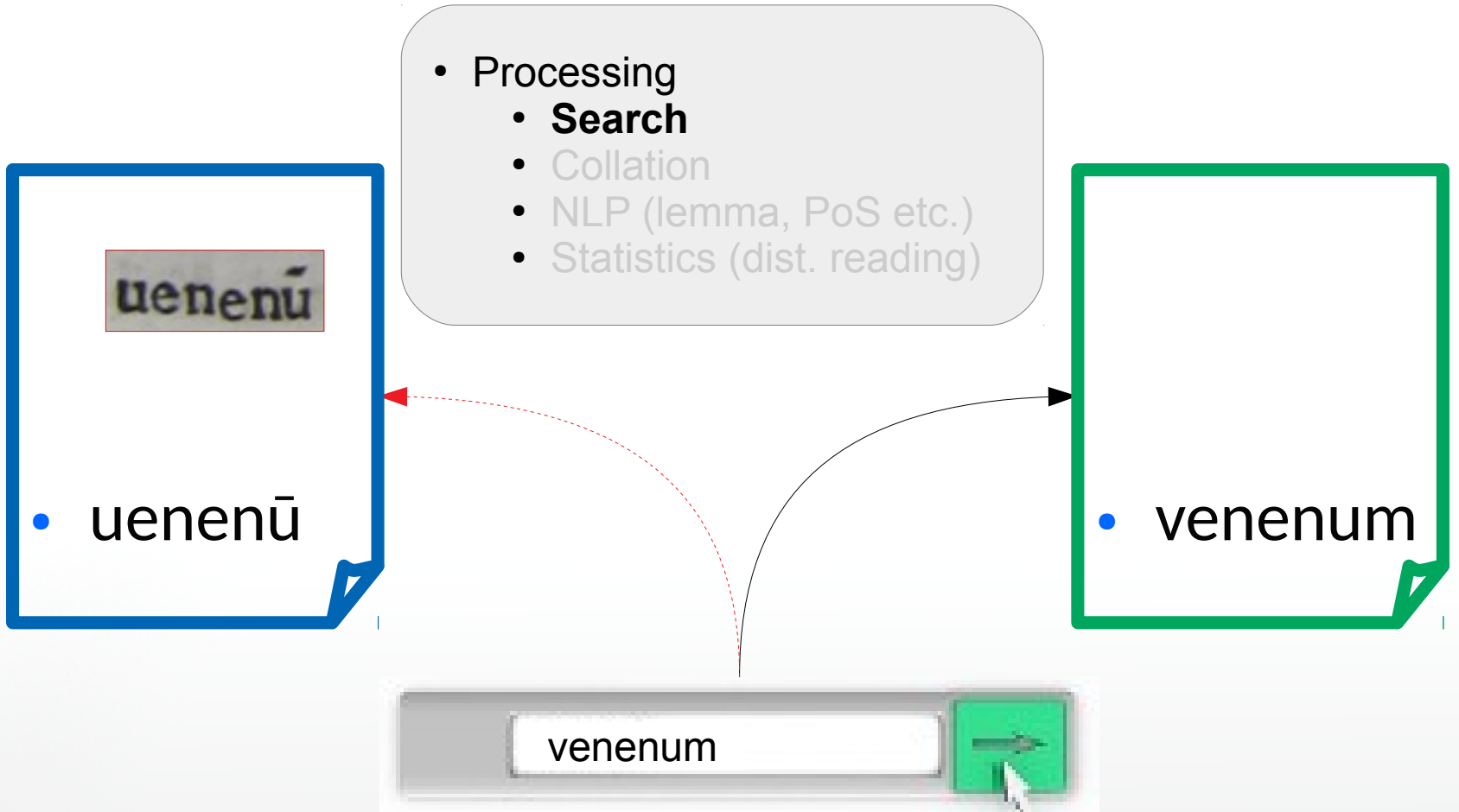
uenenū

• uenenū



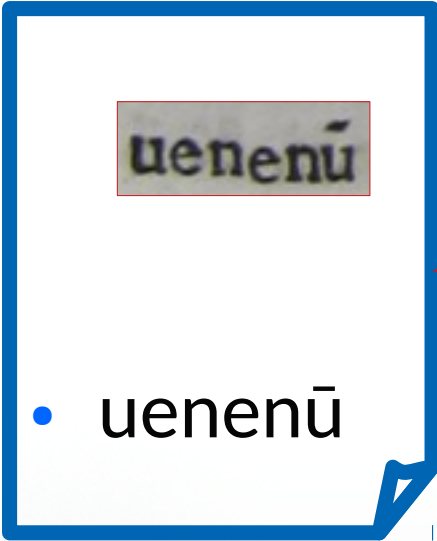
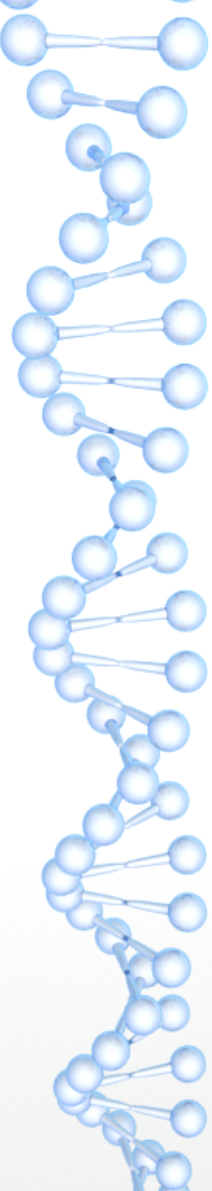
• venenum

The interoperability issue



The interoperability issue

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



uenenū

• uenenū

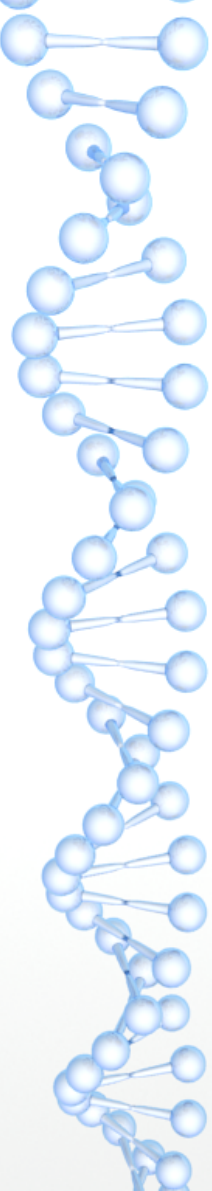


• venenum

112 excedit] excedit corr. ex
exceditis R n70
114 obicitur] obiceretur V S n71
114 sunt] sint S n72

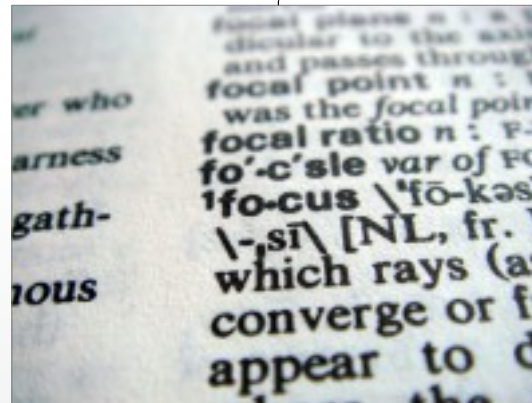
The interoperability issue

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



uenenū

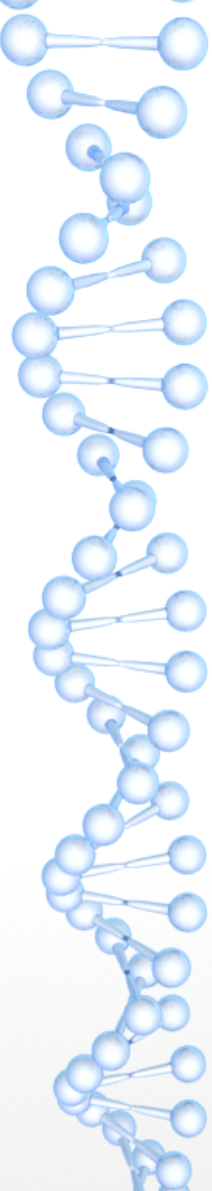
• uenenū



• venenum

The interoperability issue

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



uenenū

• uenenū

• venenum





The interoperability issue

- *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.

The interoperability issue

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 . , ; : !

The interoperability issue

OCR from Teubner

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

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 . , ; : !



Ceci n'est pas un linguiste.





Out of the Tower



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>?
 -

Project-specific solutions

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

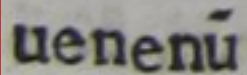


Diplomatic/normalized: the surrender?

- venenum

Normalized

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



uenenū

- uenenū

Diplomatic

- Historical documentation
- Visualization

Modelling: the (long) way out



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



In a nutshell

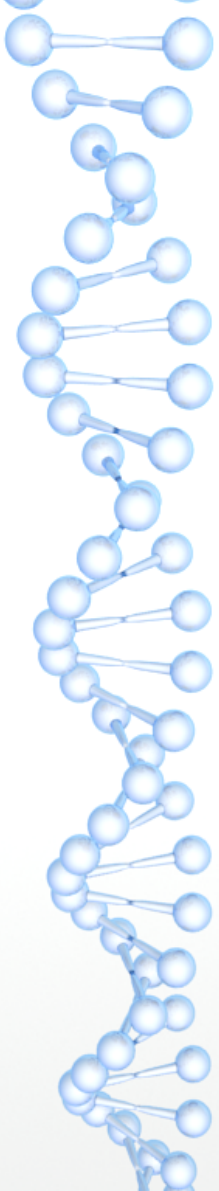
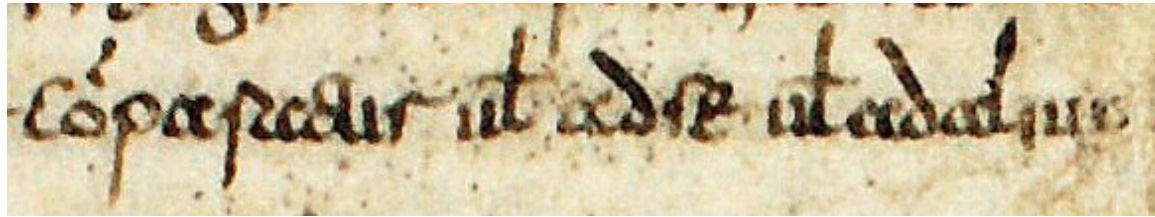


In a nutshell

- The interoperability issue
- Interoperability through modelling
- Open issues

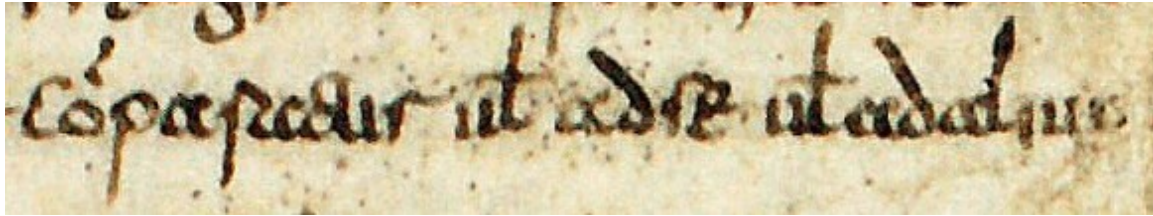
**A structuralist approach to
digital modelling for
pre-modern writing systems**

Modelling



Modelling

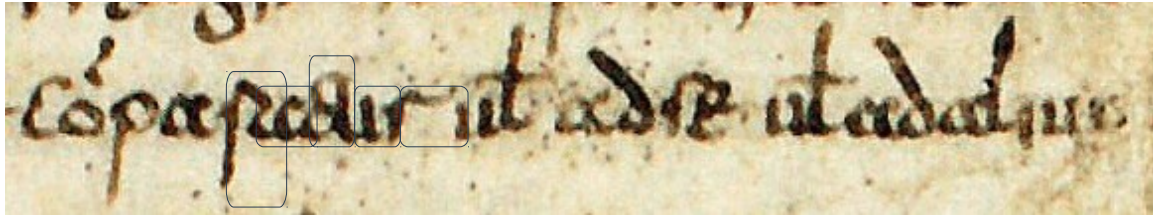
- Comparatur vel ad se vel ad alium



- cōparaEur ut adfe utadaliū

Modelling

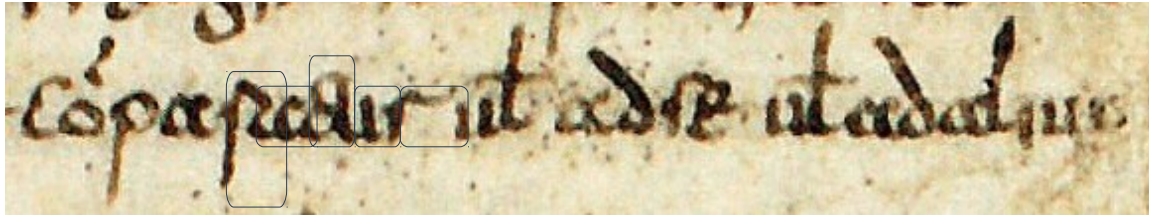
- Comparatur vel ad se vel ad alium



- cōparaetur ut adfe ut ad alium

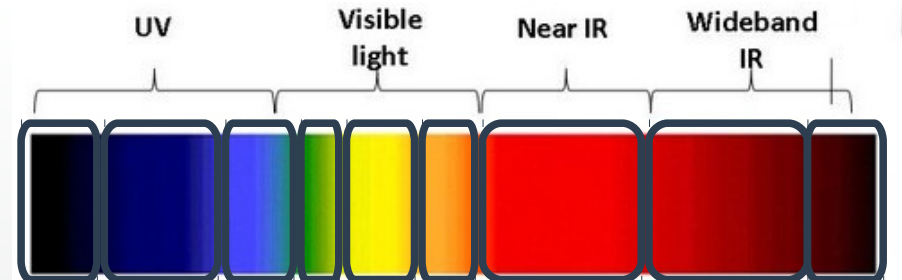
Modelling

- Comparatur vel ad se vel ad alium



- cōparaEur ut adfe utadaliu

Digital modelling



Modelling

- còparaEur ut adfe utadaliu

Modelling

- còparaEur

u† adfe u†adaliu

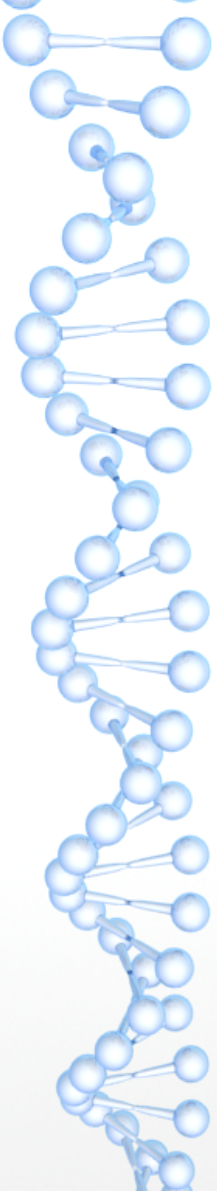
System / text

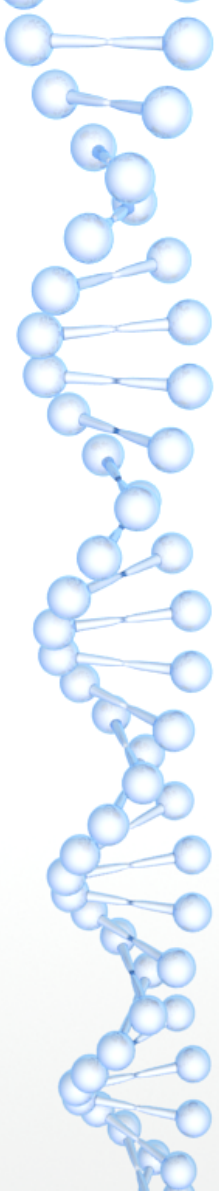
Paradigmatic

Syntagmatic (text, process)

• còparaεur uł adfe uładaliu

(langue, system)





System / text

System

<s>
<t>

Text

- còparaEur

uł

adfe

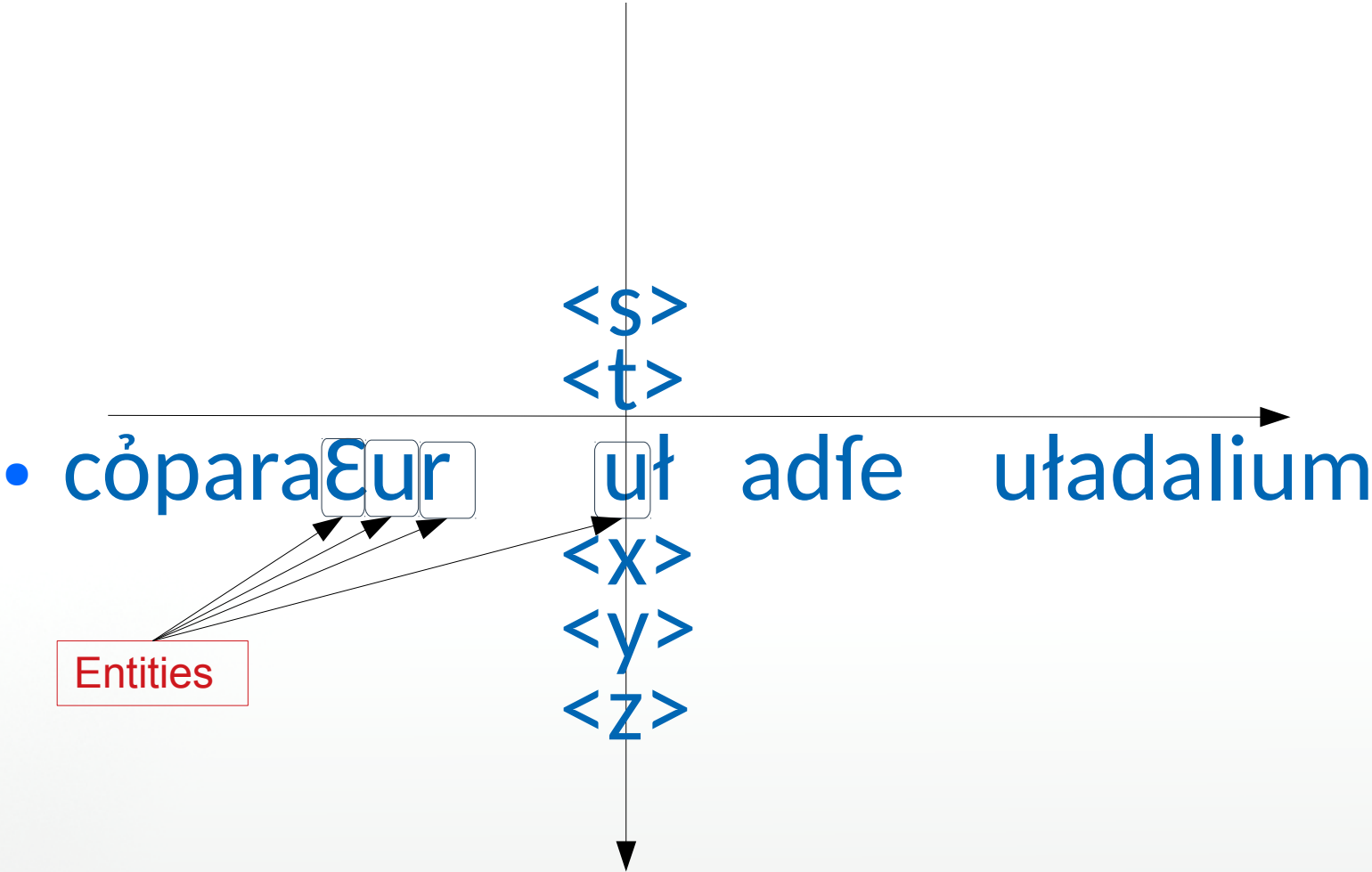
uładaliu

<x>

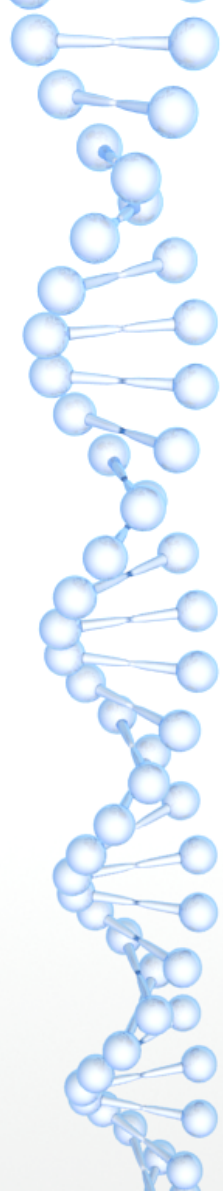
<y>

<z>

Modelling/“analysis”: entities



Modelling/“analysis”: entities

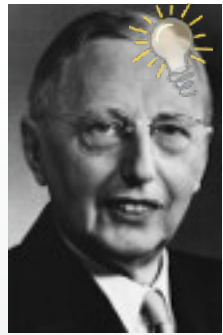


• còparaξur

Entities

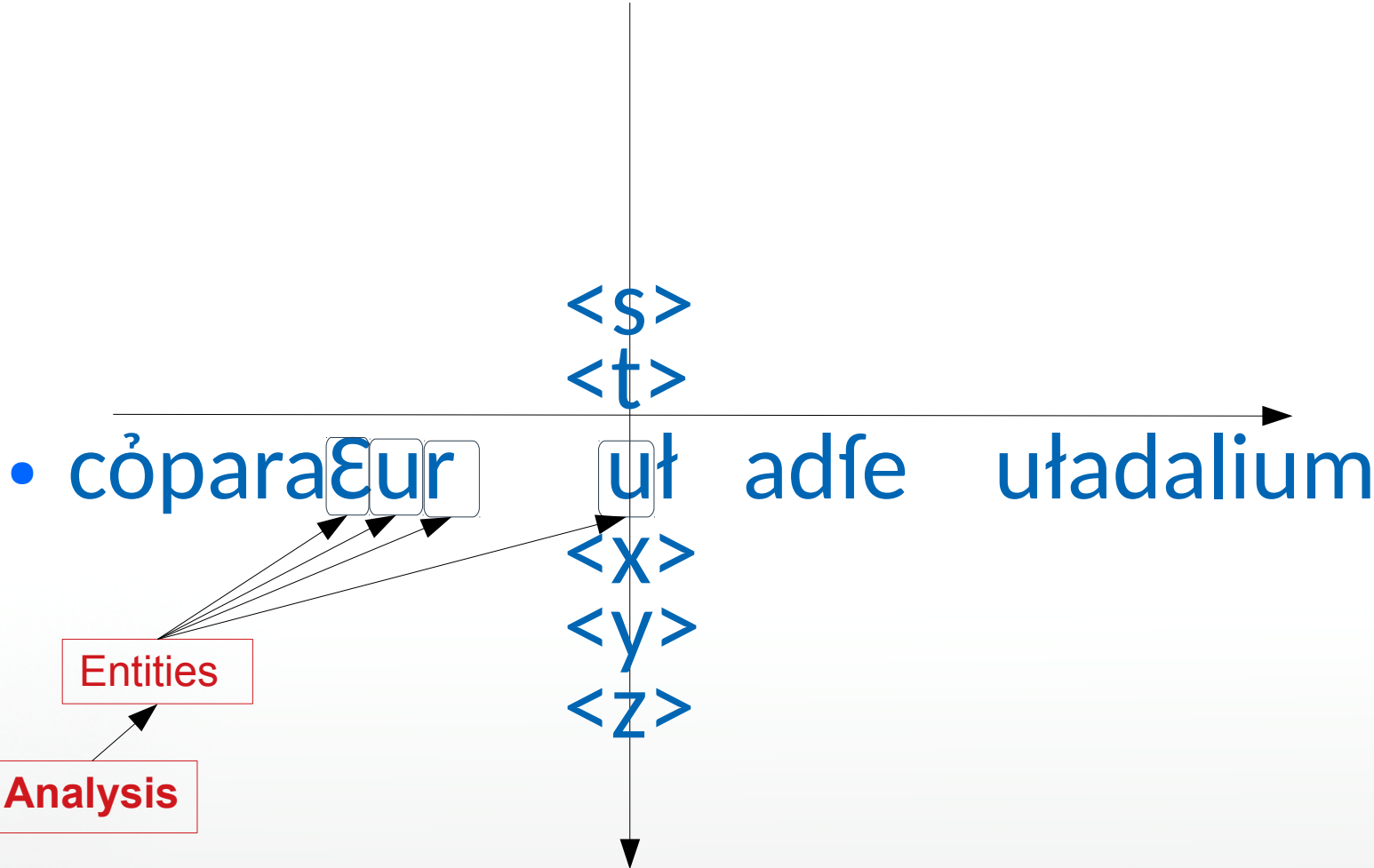
<s>
<t>
<x>
<y>
<z>

u† adfe u†adaliu

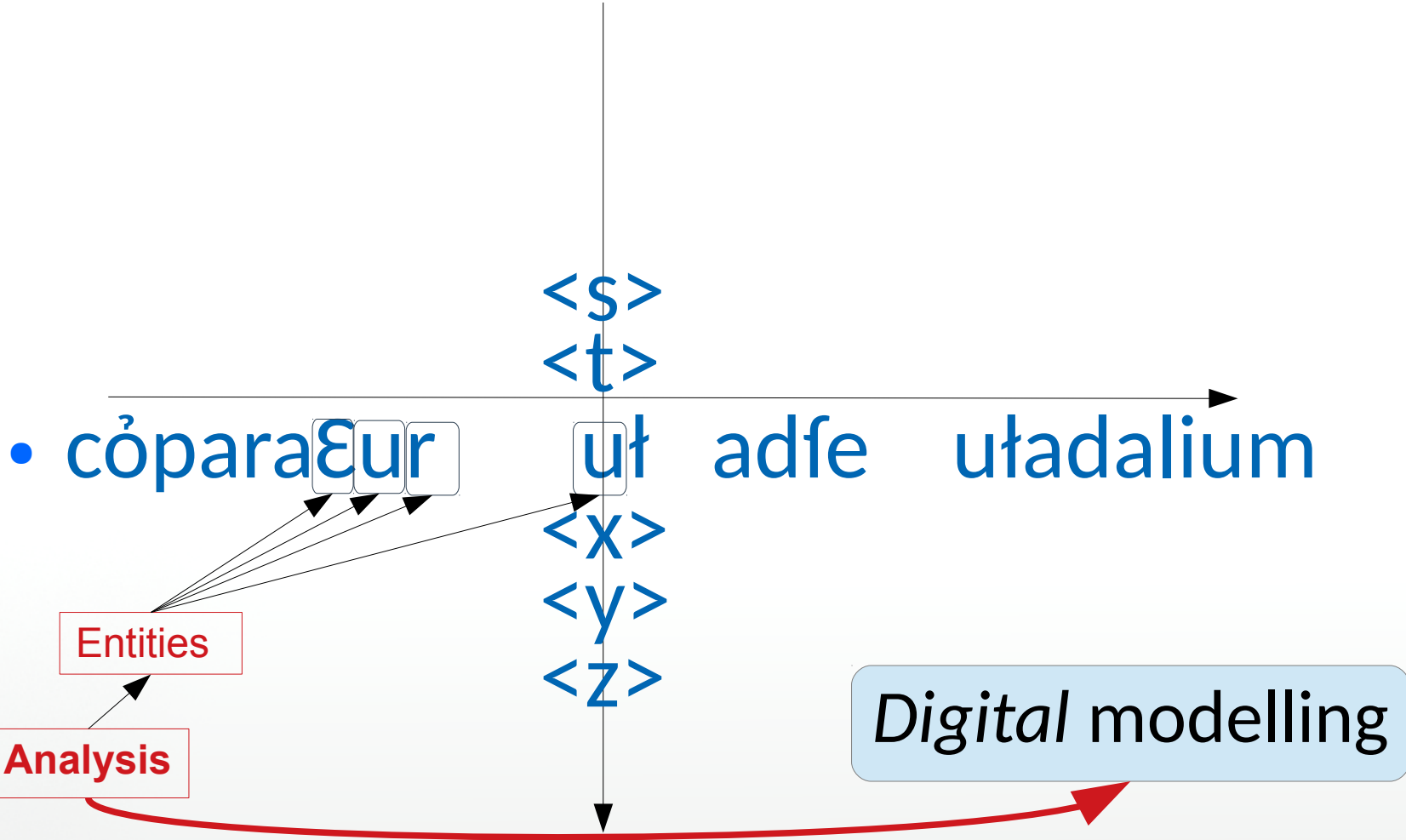


“If [we are] given anything..., it is the as yet unanalyzed *text* in its undivided and absolute integrity” (“deduction”, *Prolog.* Ch. 4)

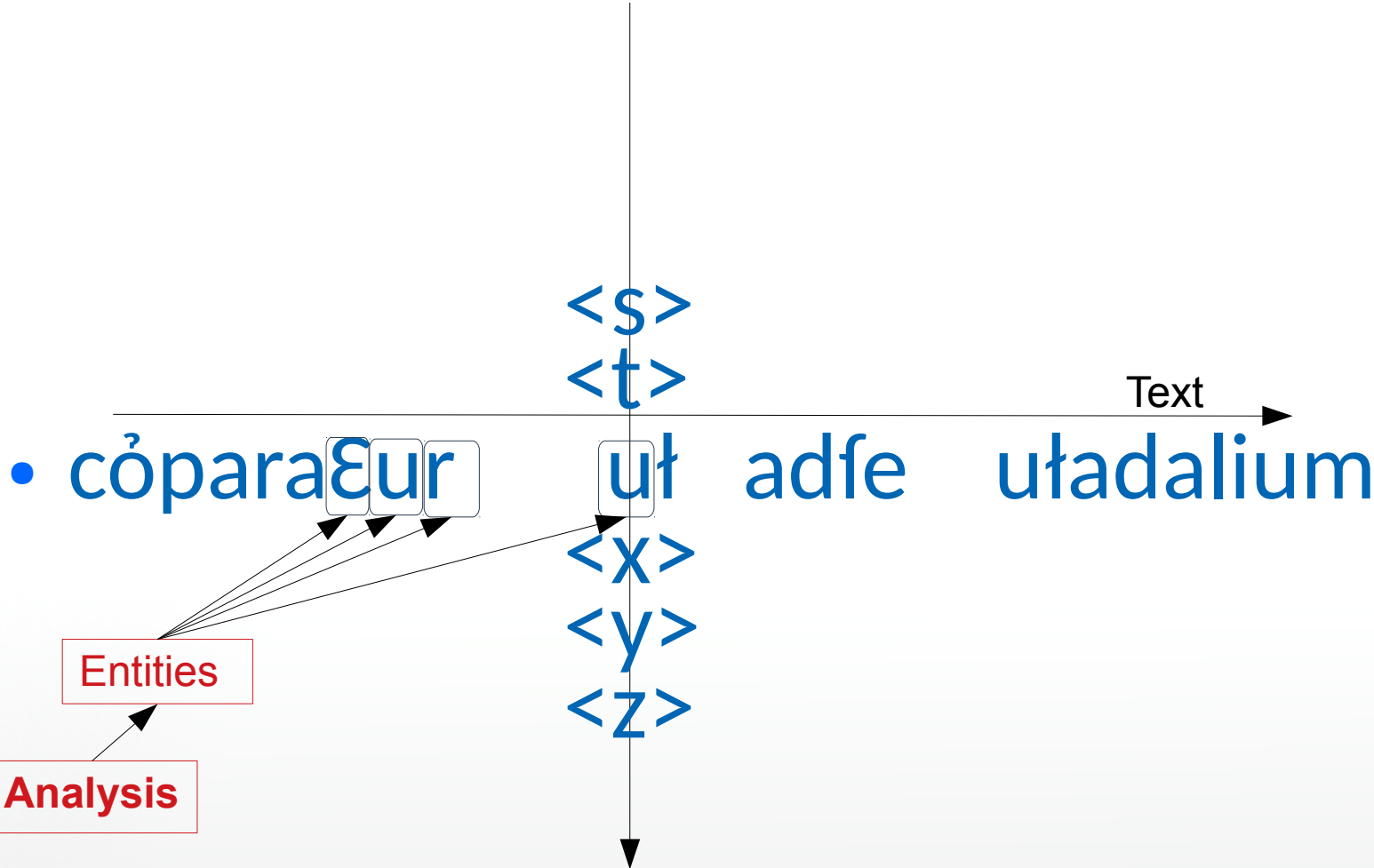
Modelling/“analysis”: entities



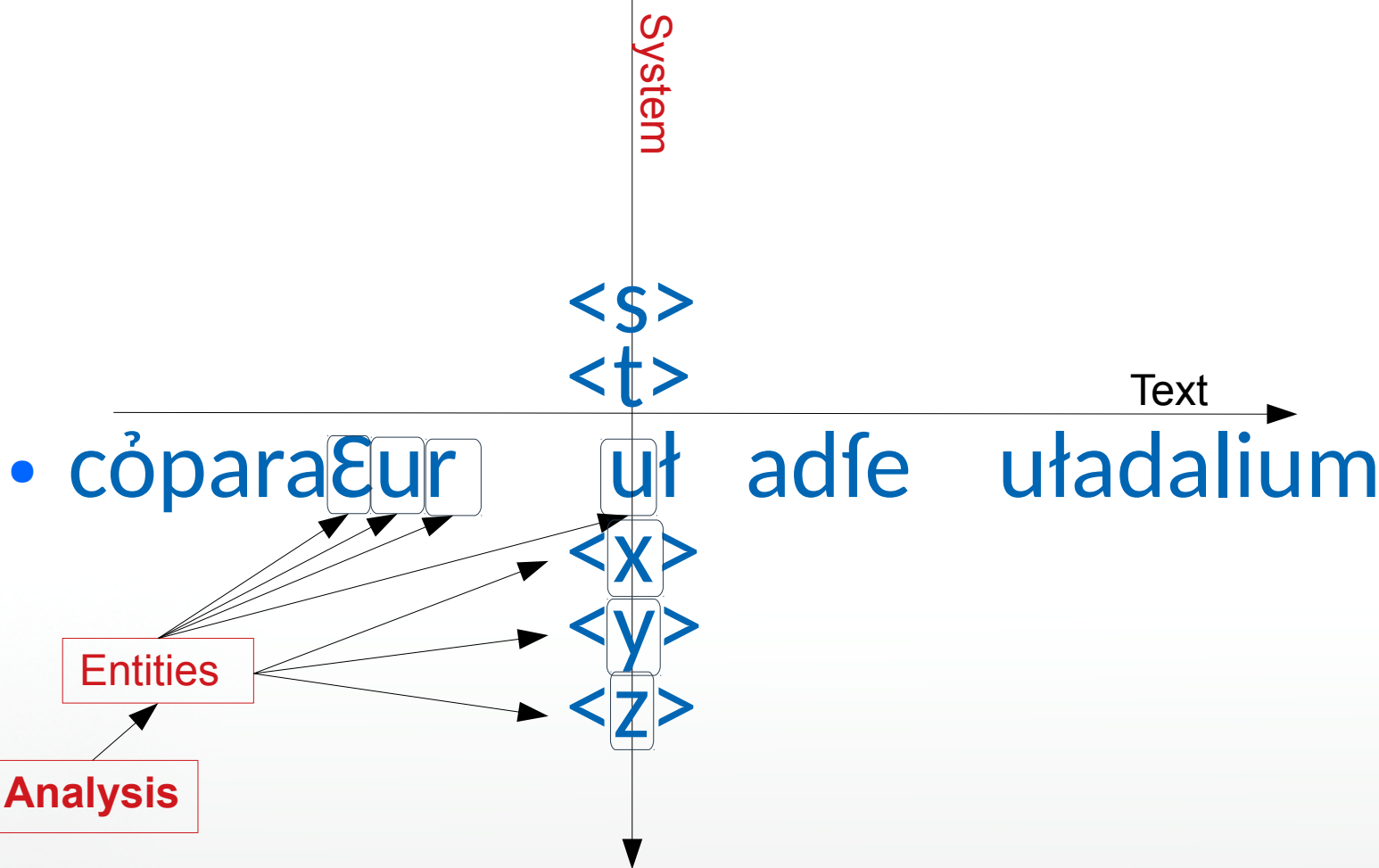
Modelling/“analysis”: entities



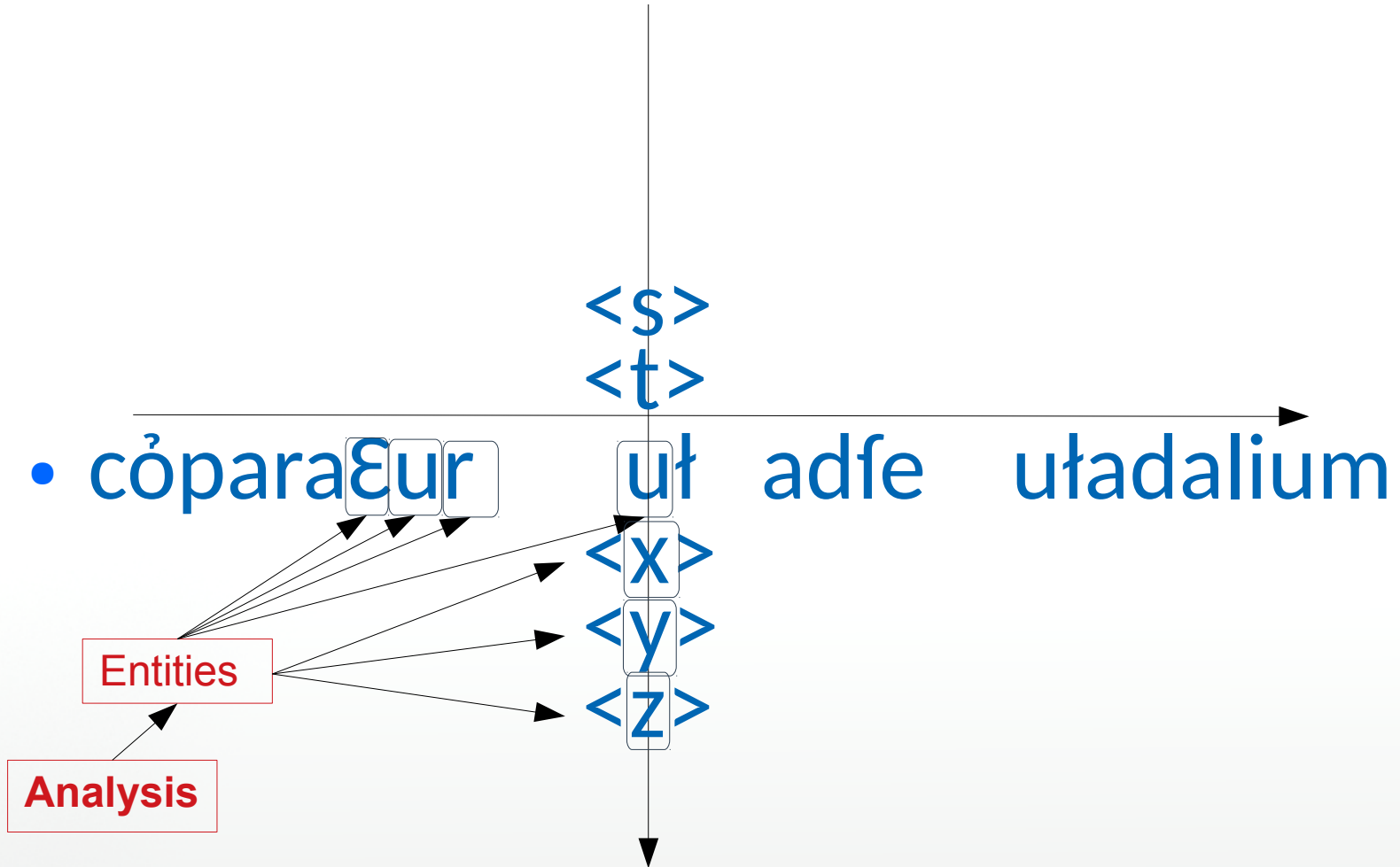
Modelling/“analysis”: entities



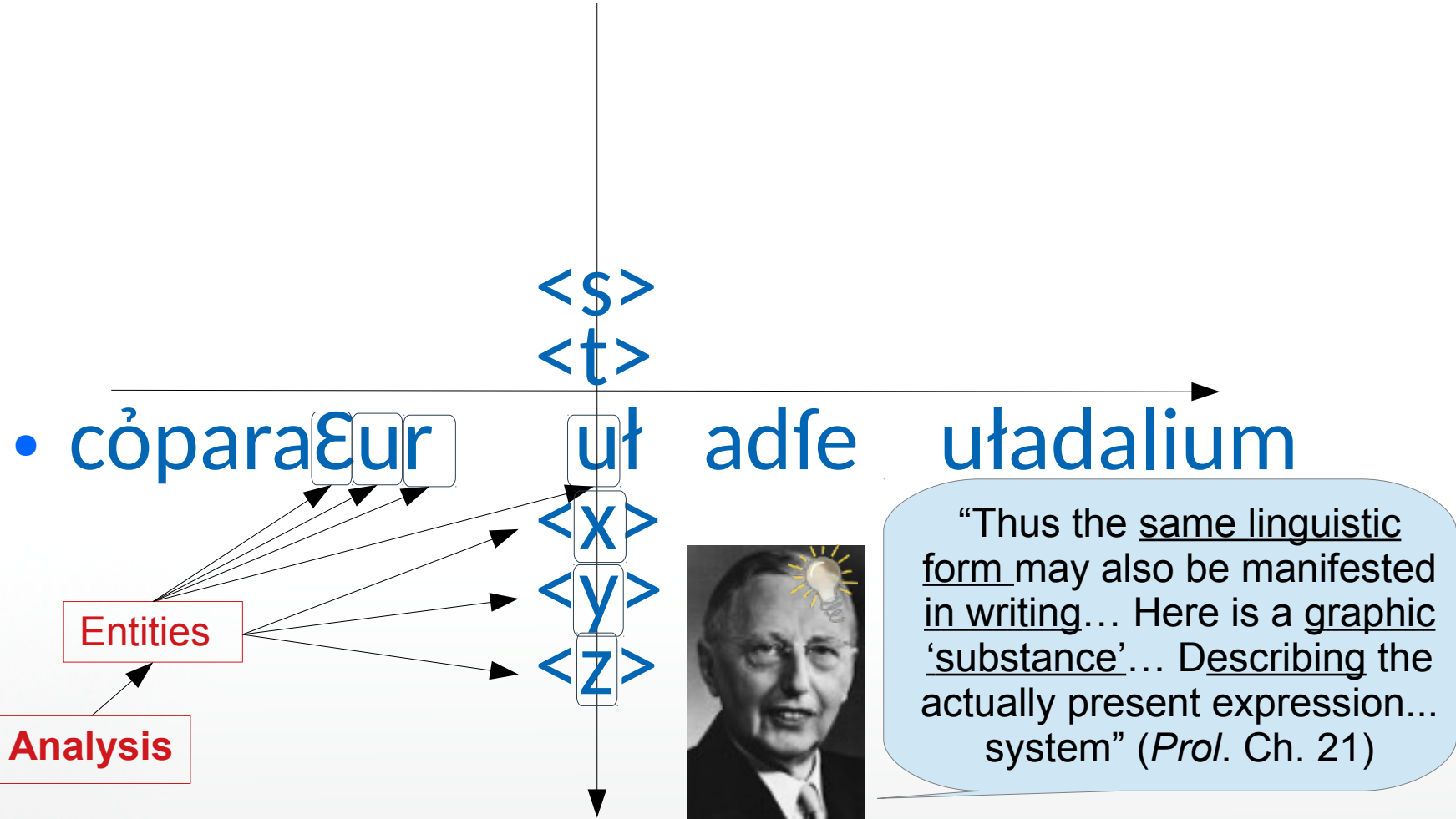
Modelling/“analysis”: entities



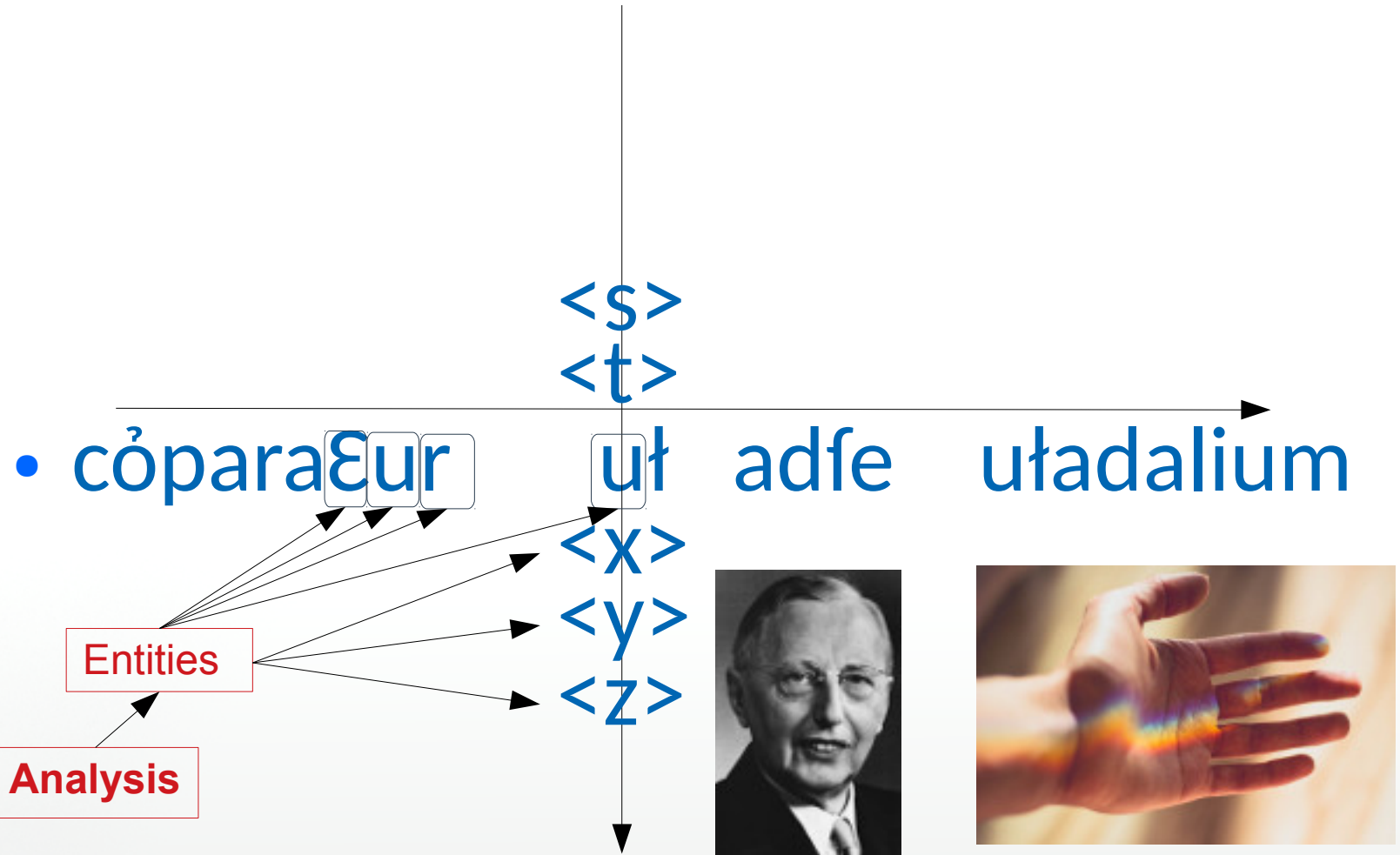
Graphemes as entities?



Graphemes as entities?



Graphemes as entities?



Graphemes as entities?

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

• còparaξur

<s>
<t>

ut

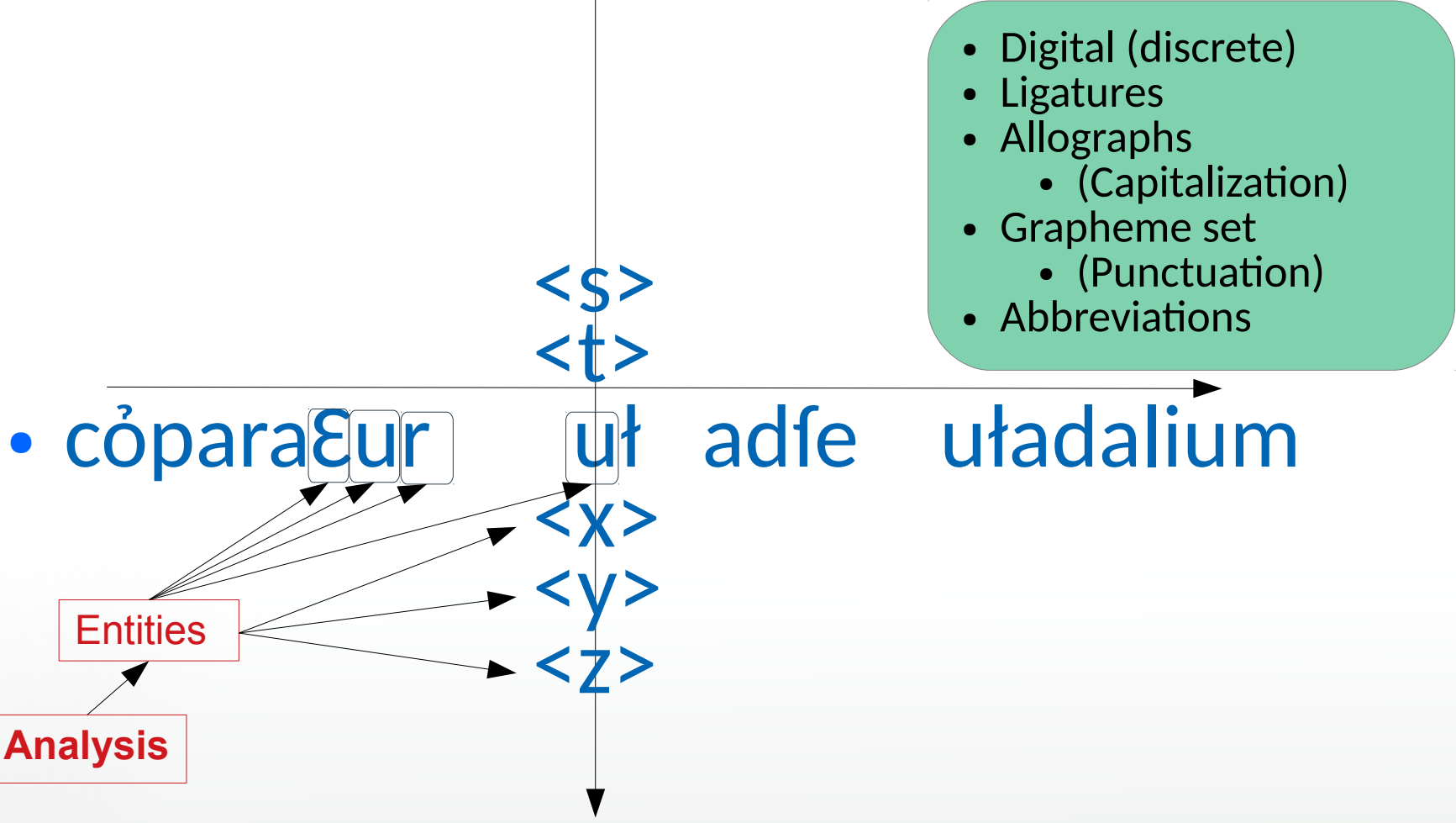
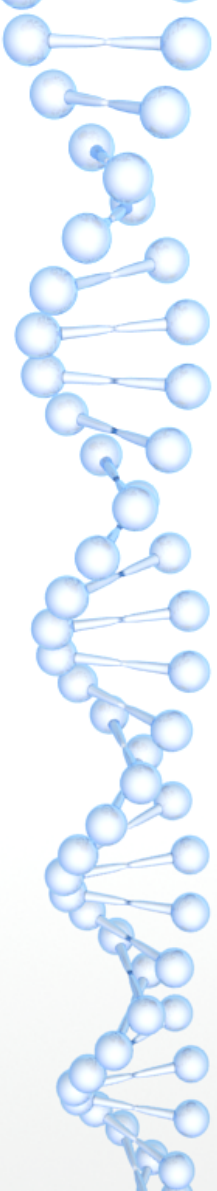
adfe

utadalium

Analysis

Entities

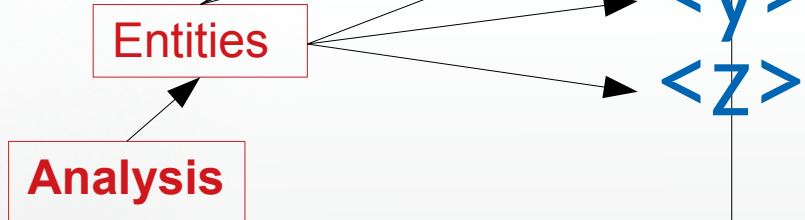
<x>
<y>
<z>



Graphemes as entities?

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

• còparaEUR ut adfe utadalium



<s>

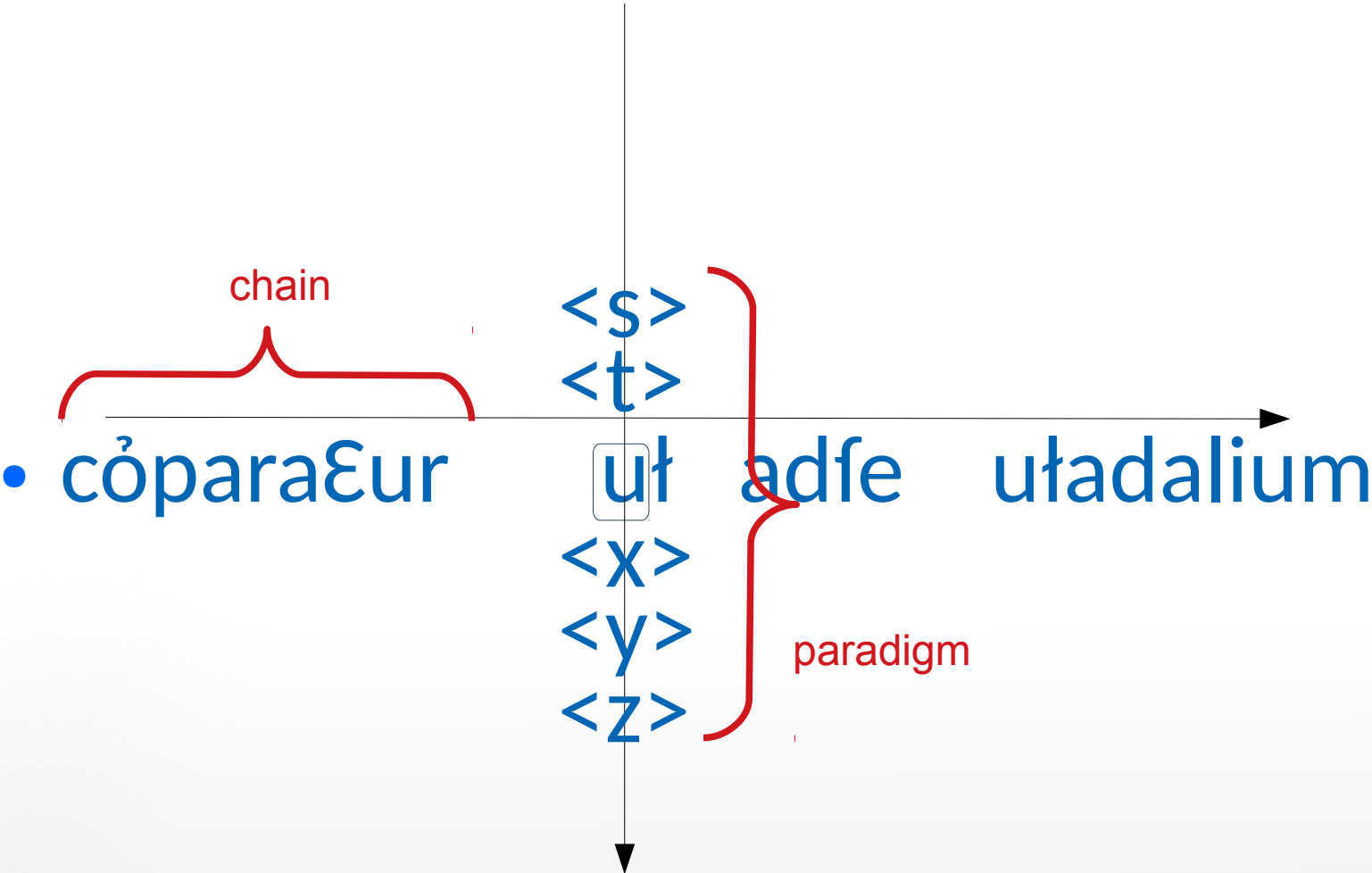
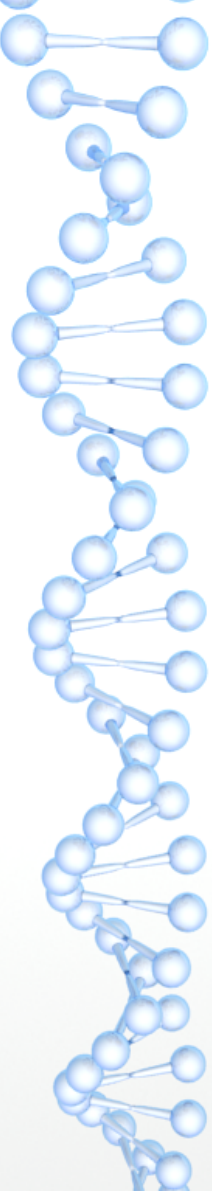
<t>

<x>

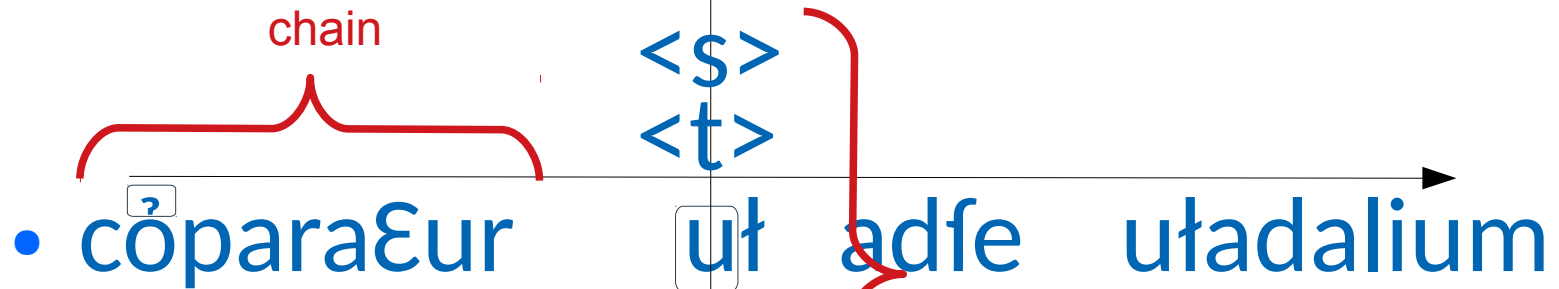
<y>

<z>

Chains and paradigms



Chains and paradigms



<S>

<t>

<x>

<y>

<z>

paradigm

“chain” → sequence?



Functions



• còparaEur



Function

<s>

<t>

ut

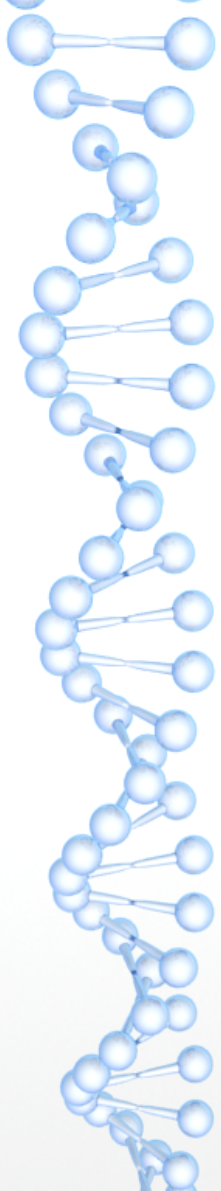
adfe

utadaliu

<x>

<y>

<z>



Functions

<s>
<t>

• còparaEur

u

adfe

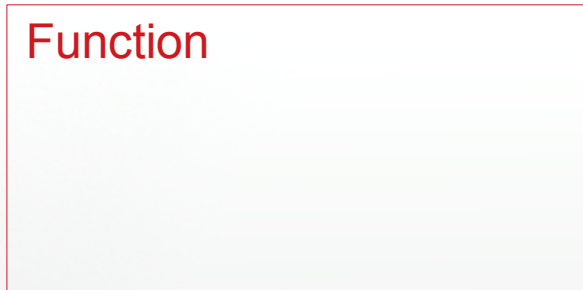
uładaliu

<x>

<y>

<z>

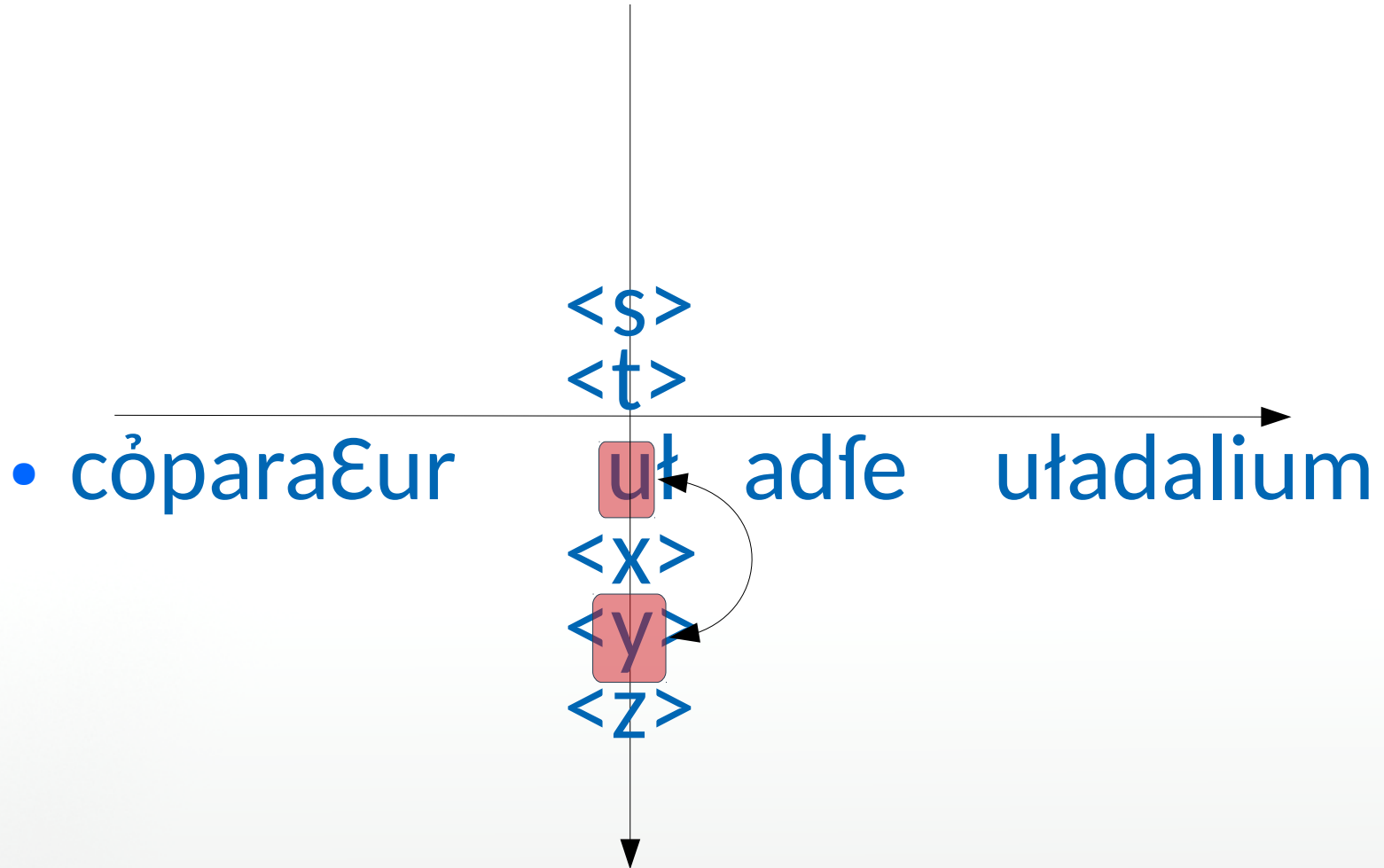
Function



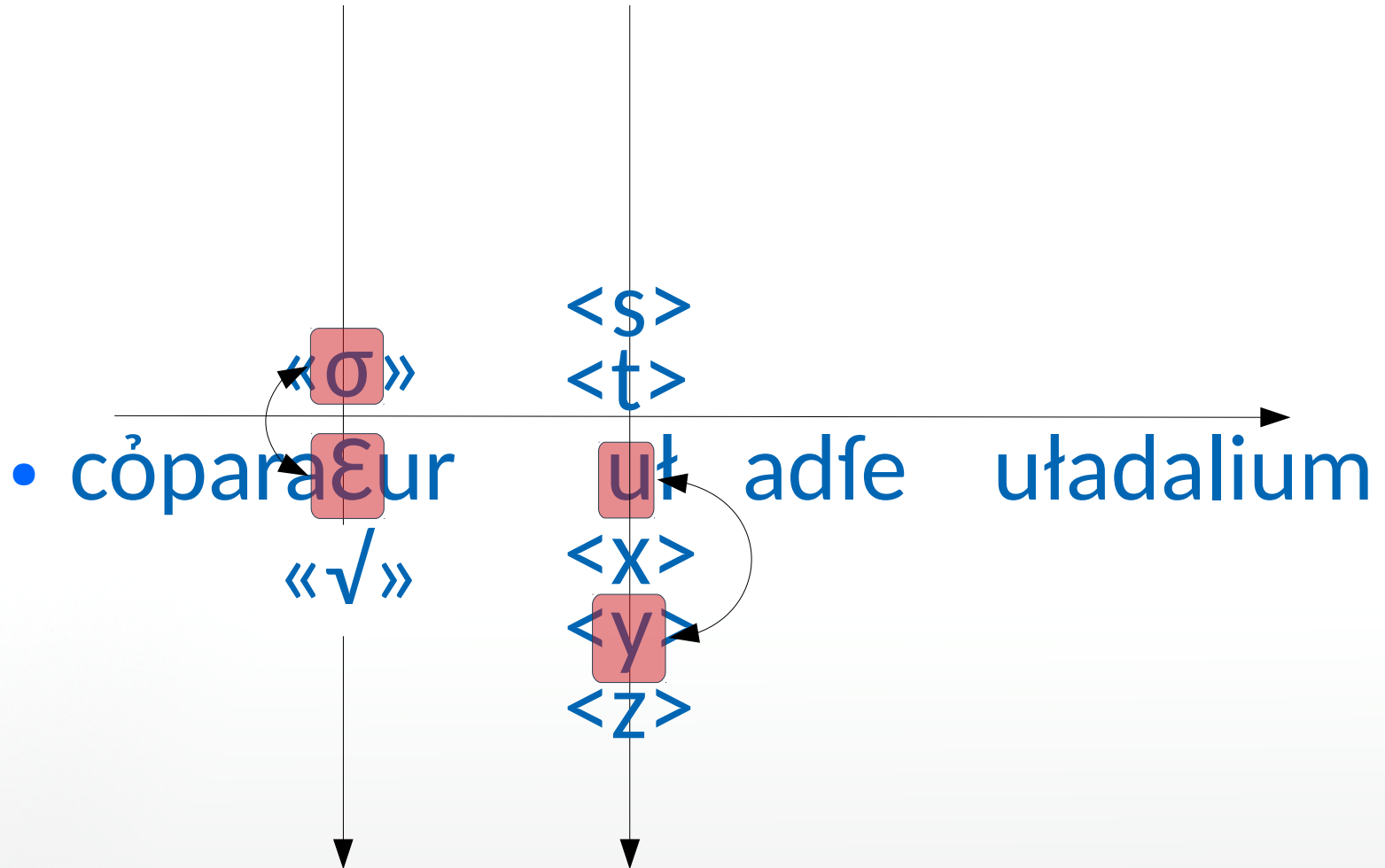


Graphemes/allographs

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>

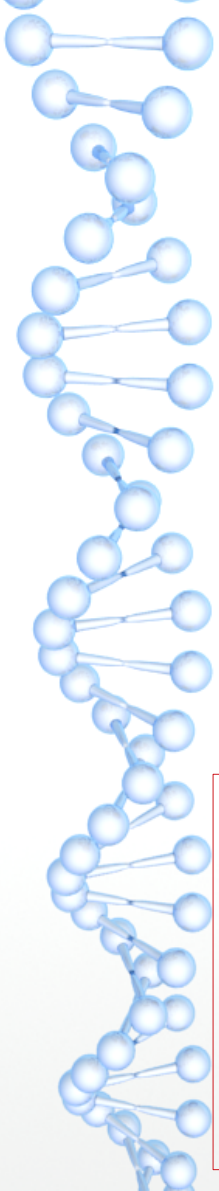
adfe

utadaliu

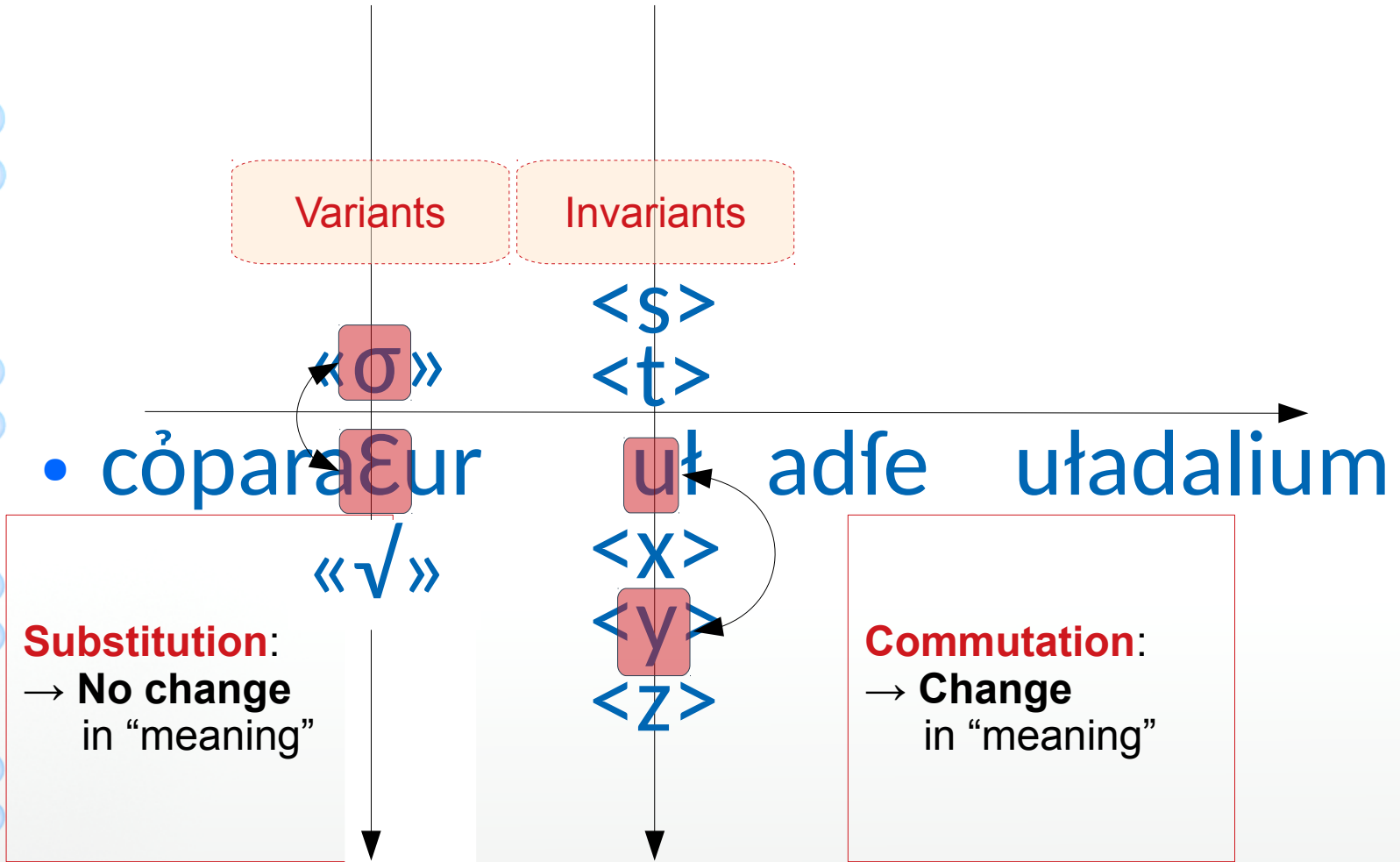
Function

Commutation:
→ Change
in “meaning”

annus / annys

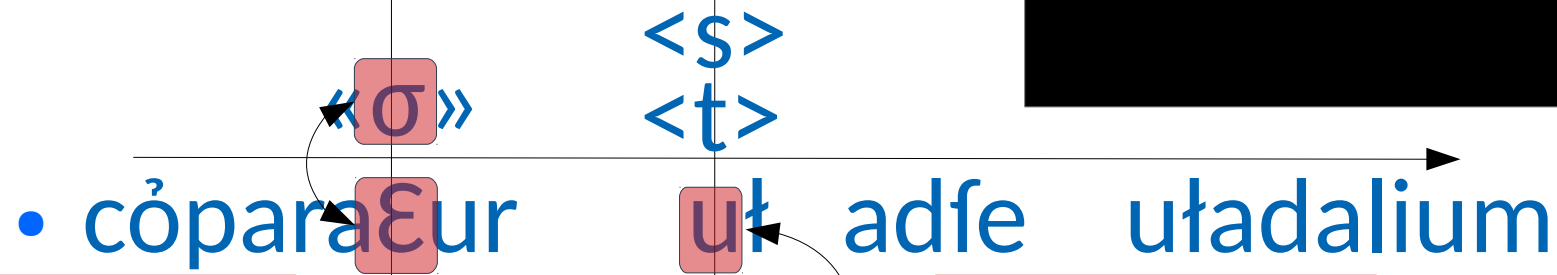


The commutation test



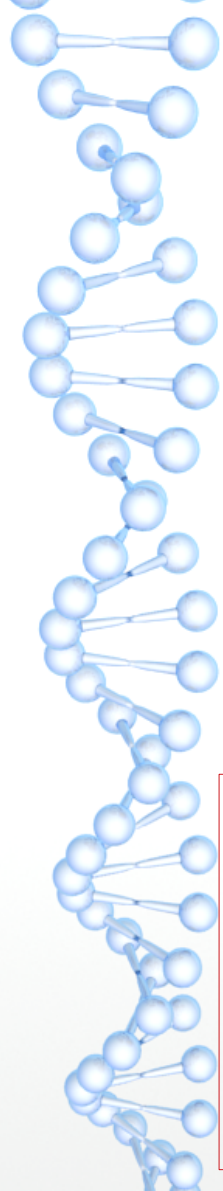
Graphemes / allographs

Gr	Allogr
t:	σ ε $\sqrt{}$
u:	u v
z:	z

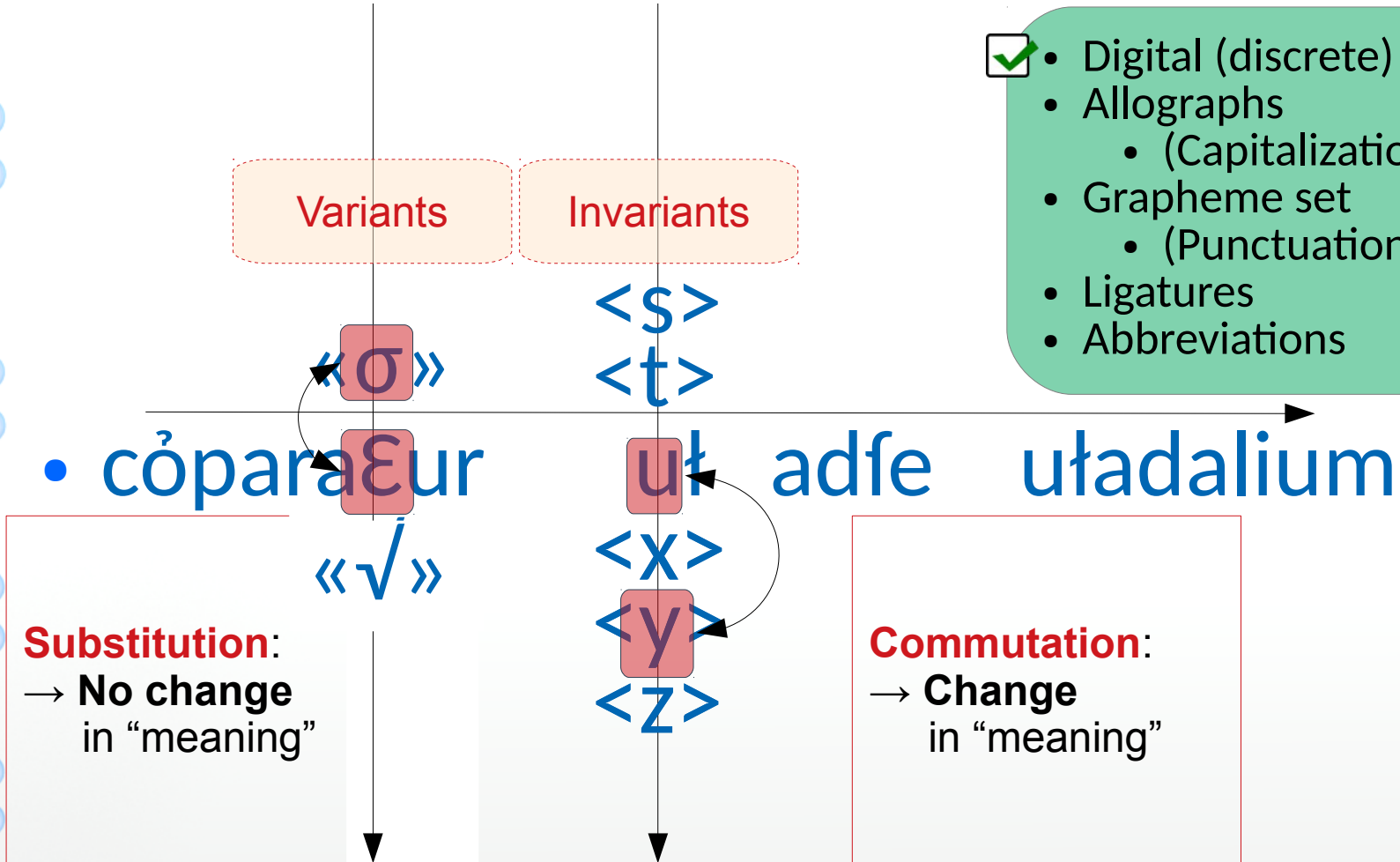


Substitution:
 → No change
 in "meaning"

Commutation:
 → Change
 in "meaning"



Graphemes / allographs



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

• còparaEur

<S>
<t>
u
<x>
<y>
<z>

adfe

utadaliu

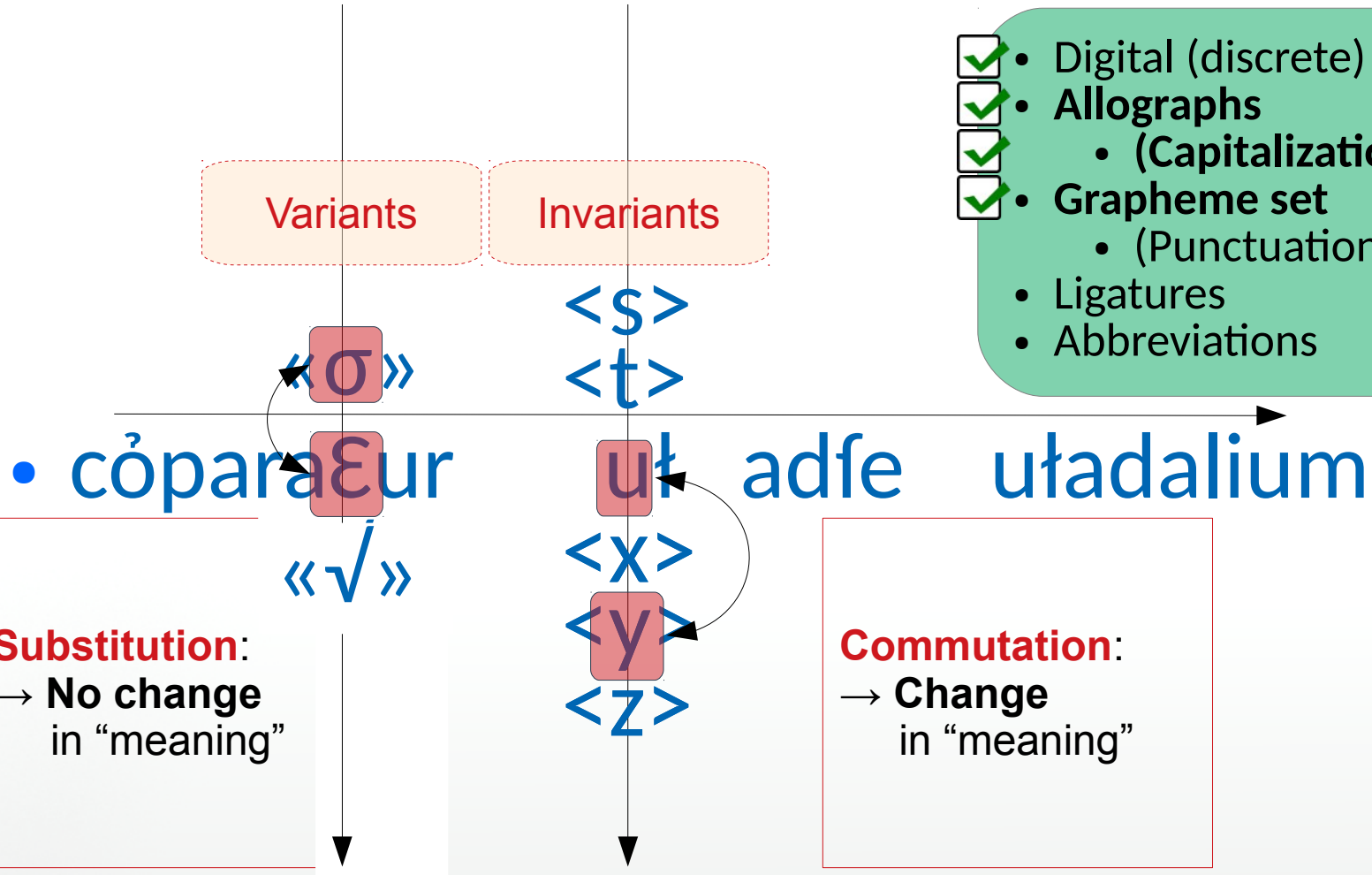
«σ»
«√»

Substitution:
→ No change
in "meaning"

Commutation:
→ Change
in "meaning"

Graphemes / allographs

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



• còparaEur

<S>
<t>
ut
<x>
<y>
<z>

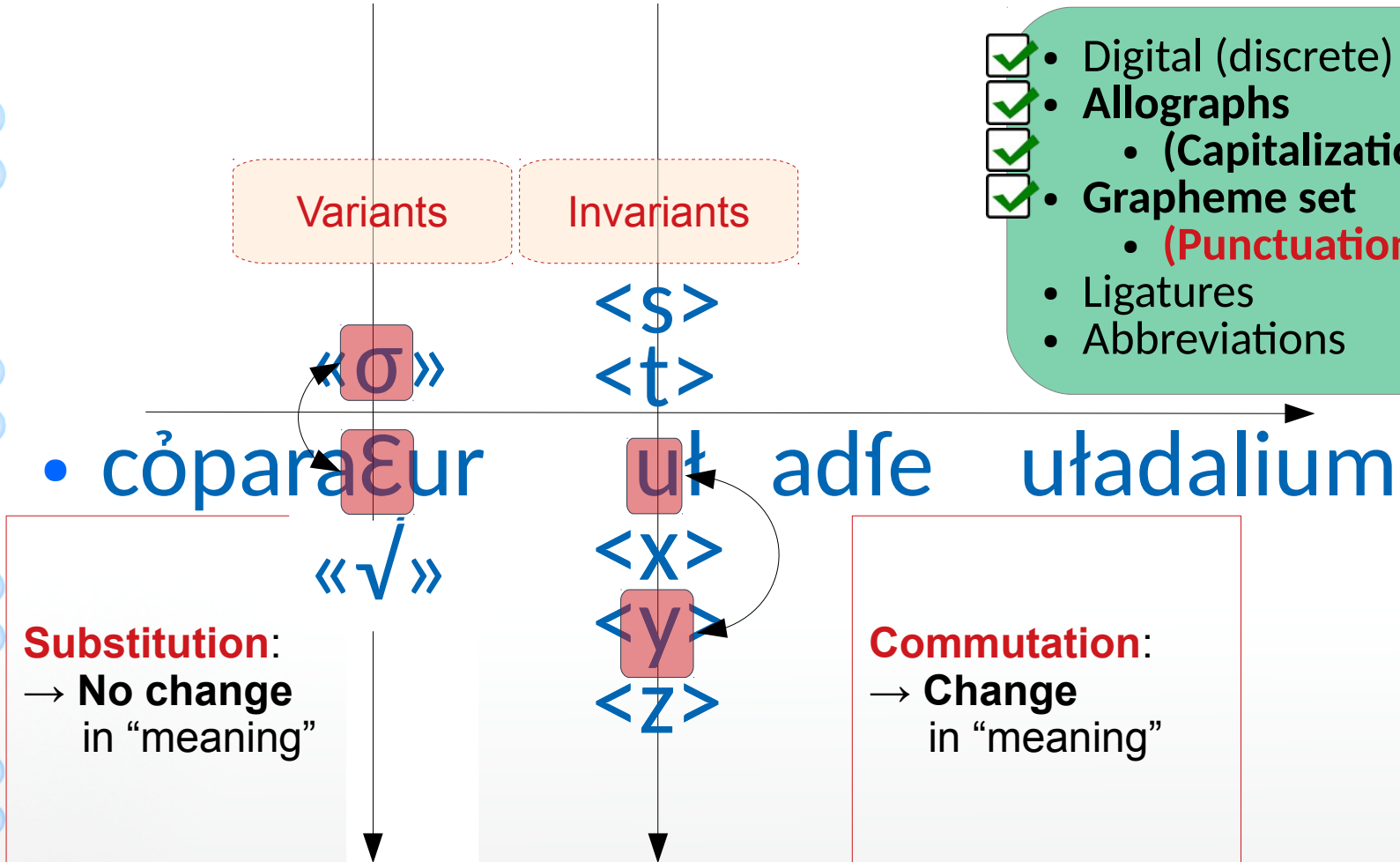
adfe utadaliu

Substitution:
→ No change
in "meaning"

Commutation:
→ Change
in "meaning"

Graphemes / allographs

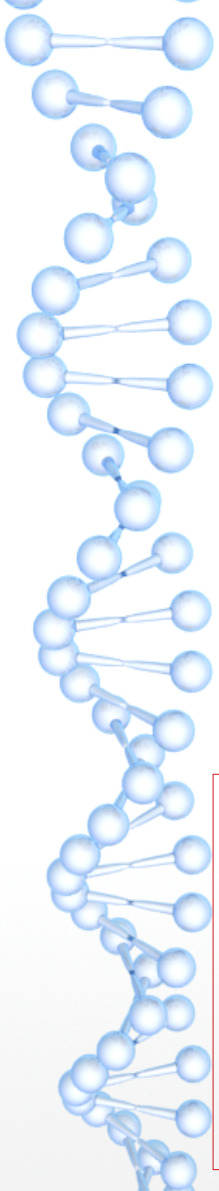
- ✓ Digital (discrete)
- ✓ Allographs
 - (Capitalization)
- ✓ Grapheme set
 - (Punctuation)
 - Ligatures
 - Abbreviations





Punctuation

Punctuation



• còparaEur

«σ»

«√»

Substitution:
→ No change
in “meaning”

<S>
<t>

uł

<x>

<y>

<z>

adfe

uładaliu



“content-form”
(*Prol. Ch. 13*)

Commutation:
→ Change
in “meaning”

Punctuation



“content-form”
(*Prolog* Ch. 13)

Including larger
units, such as
sentences

• còparaEur

<S>
<t>

adfe

utadaliu

«σ»

u

«√»

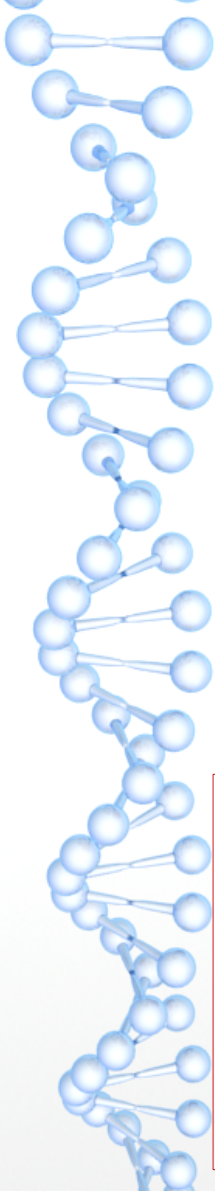
<x>

<y>

<z>

Substitution:
→ No change
in “meaning”

Commutation:
→ Change
in “meaning”



Punctuation



“content-form”
(*Prolog. Ch. 13*)

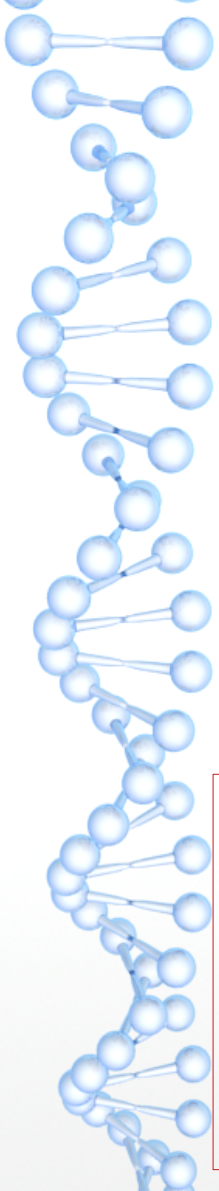
Including larger
units, such as
sentences

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”
(*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

• còparaEur

«σ»

«√»

Substitution:
→ No change
in "meaning"

<s>
<t>

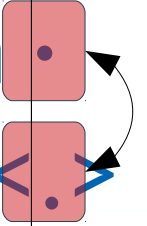
ut

<x>
<y>
<z>

adfe

utadaliu

Commutation:
→ Change
in "meaning"



Punctuation

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

«σ»

<s>
<t>

• còparaEur

ut

adfe

utadaliu

«√»

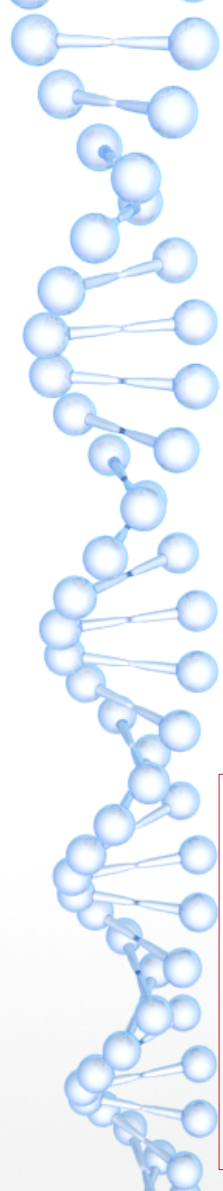
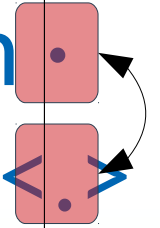
<x>

<y>

<z>

Substitution:
→ No change
in “meaning”

Commutation:
→ Change
in “meaning”



Punctuation

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

«σ»

<s>
<t>

• còparaEur

ut

adfe

utadaliu

«√»

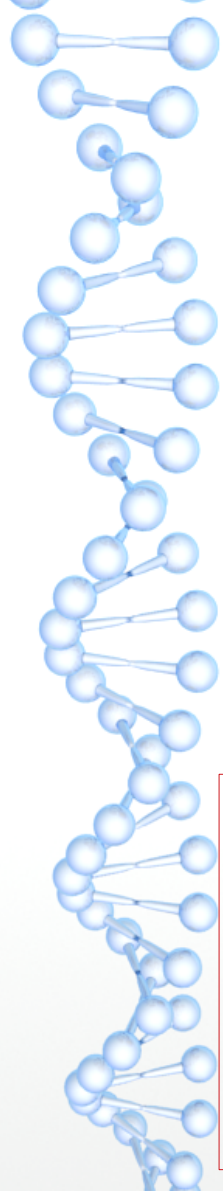
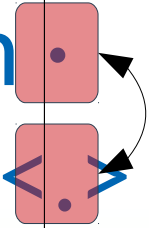
<x>

<y>

<z>

Substitution:
→ No change
in "meaning"

Commutation:
→ Change
in "meaning"





Ligatures

Ligatures



• còparaEur



Function

“Ligature”
(entities, parts of a chain)

<s>

<t>

uł

adfe

uładaliu

<x>

<y>

<z>

Ligatures

Varieties



«σ»

<s>
<t>

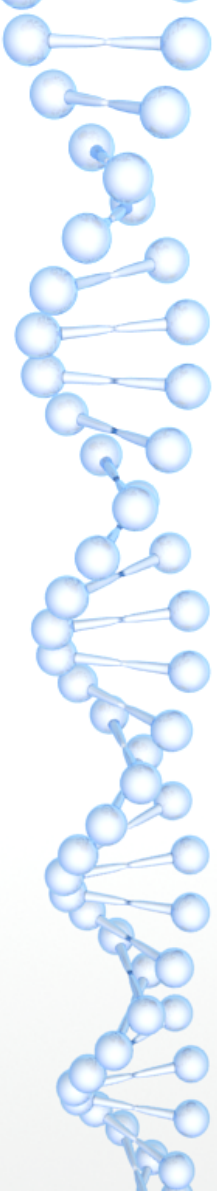
• còparaεur

«√»

ut
<x>
<y>
<z>

adfe

utadaliu



Ligatures

Varieties:
Solidal
variants



«σ»

<s>

<t>

• còparaur

ut

adfe

utadalium

«√»

<x>

<y>

<z>

Ligatures

Varieties:
Solidal
variants



«σ»

<s>
<t>

```
<seg type="lig">tu</seg>  
"lig"/tu → ε+ϰ → t+u  
  
<g ref="ligTU"/>  
"ligTU" → ε+ϰ → t+u
```

• còparaεur

«√»

ut

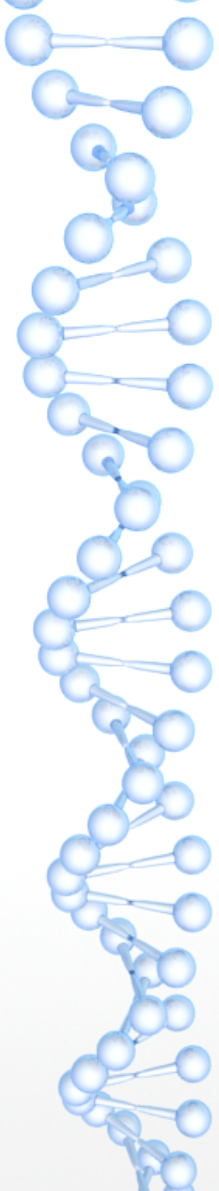
<x>

<y>

<z>

adfe

utadaliu



Ligatures

Varieties:
Solidal
variants



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

• còparaur

<s>
<t>

ut

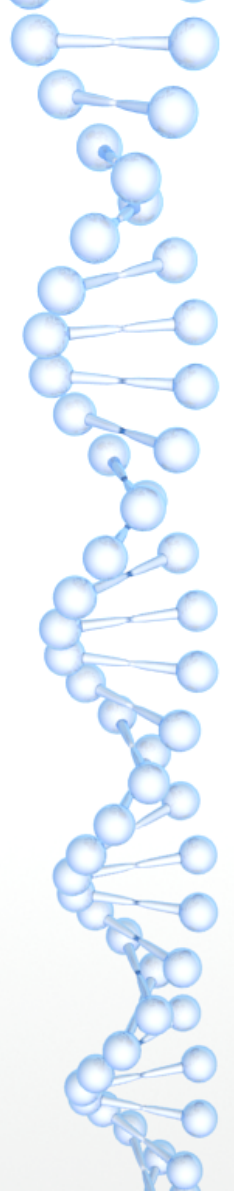
adfe

utadaliu

<x>

<y>

<z>



Ligatures

Varieties:
Solidal
variants



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

• còparaur

<s>
<t>

ut

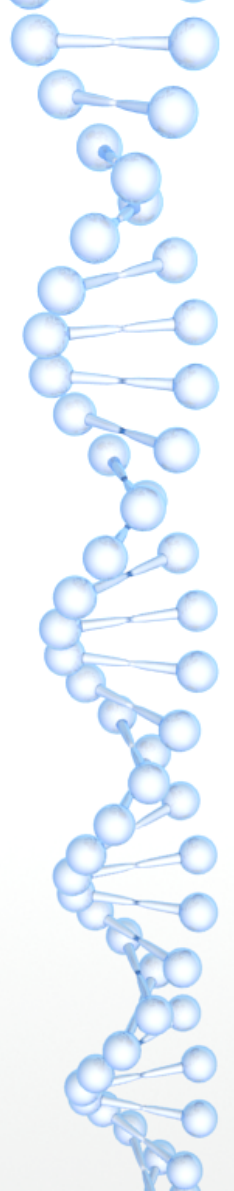
adfe

utadalium

<x>

<y>

<z>



Abbreviations

Abbreviations

-    comparatur ut adfe utadaliu

Abbreviations: one grapheme?

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



<O>

<ô>

<ō>

• cōparaEUR

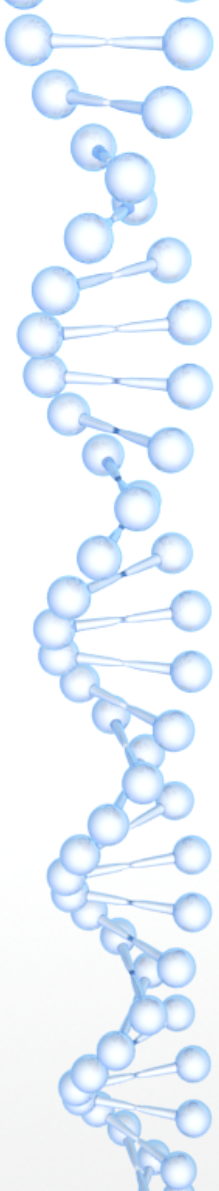
<p>

<q>

ut adfe utadanium



Abbreviations: one grapheme?



<O>

<ô>

<ō>

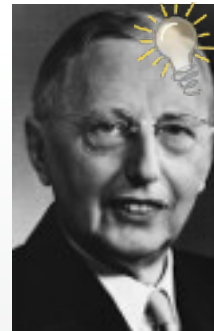
• cōparaεur

uł adfe

uładaliu

<p>

<q>



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)



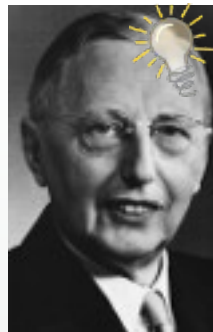
Abbreviations: function → ligature?

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

•  comparatur ut adfe utadadium

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 utadadium**

Function

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

Abbreviations: function → complementarity?

•  comparatur ut adfe ut adalium

Function

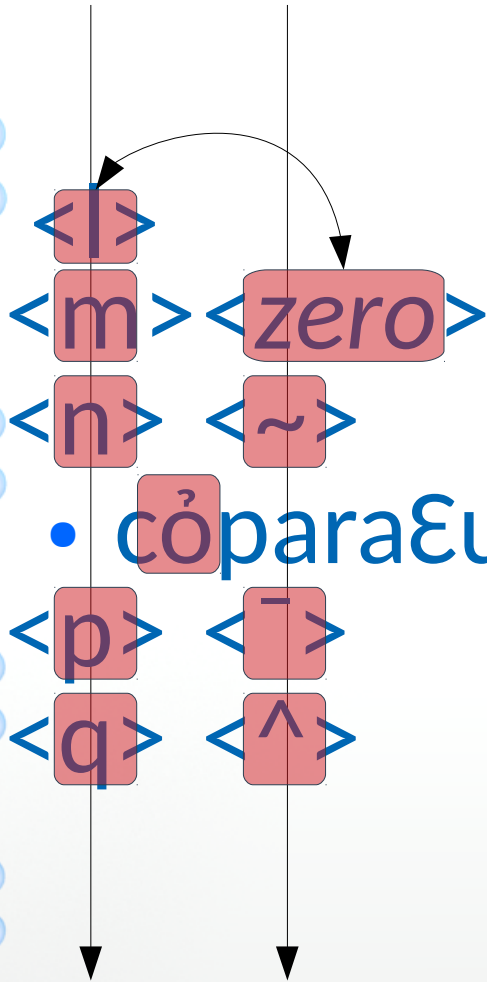
Complementarity
(interdependence in a system)

Example of
complementarity

[case + gender + number]

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

Abbreviations: function \rightarrow complementarity?



• $\langle \text{c} \rangle$ $\langle \text{o} \rangle$ $\langle \text{p} \rangle$ $\langle \text{a} \rangle$ $\langle \text{r} \rangle$ $\langle \text{a} \rangle$ $\langle \text{e} \rangle$ $\langle \text{u} \rangle$ $\langle \text{r} \rangle$

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

• c?ōparaEur uł adfe uładaliu

Abbreviations

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

• c?ōparaεur uł adfe uładaliu



Open issues

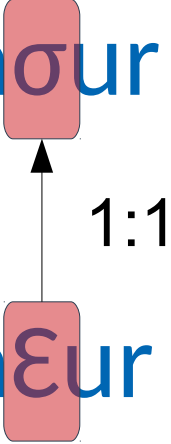


Issues: 1. Abbreviations → functions

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

- $\text{c}\acute{o}\text{p}\text{a}\text{r}\text{a}\sigma\text{u}\text{r}$
 - $\text{c}\acute{o}\text{p}\text{a}\text{r}\text{a}\epsilon\text{u}\text{r}$
- 

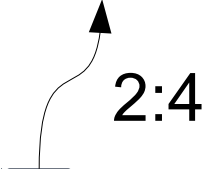


Issues: 2. Abbreviations → one:many

- **comp**paraσur
 - **c[?]o**paraεur
- 2:2 (2 ≠ 2)

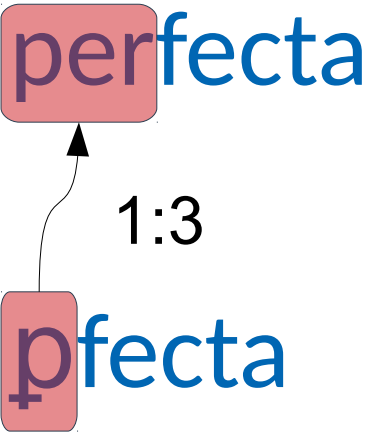


Issues: 2. Abbreviations → one:many

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



Issues: 2. Abbreviations → one:many

- **perfecta**
 - **p_♀fecta**
- 1:3
- 

Issues: 2. Abbreviations → one:many



- **perfecta**

Alphabemes
(alphabetical letters)

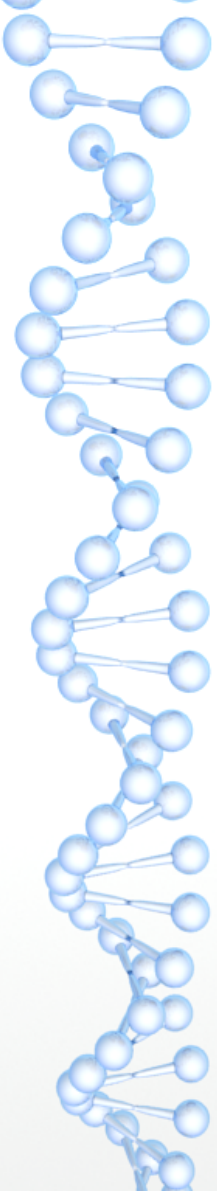
1:3

- **p_♀fecta**

Graphemes

Issues: 3. Ligatures

(syntagmatic)



Issues: 3. Ligatures



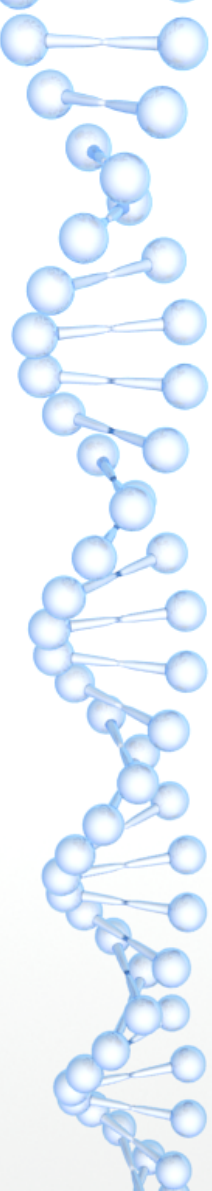
& (U+0026;
ASCII 38)

Issues: 3. Ligatures



& (U+0026;
ASCII 38)

Historical/“etymological”
considerations



Issues: 3. Ligatures



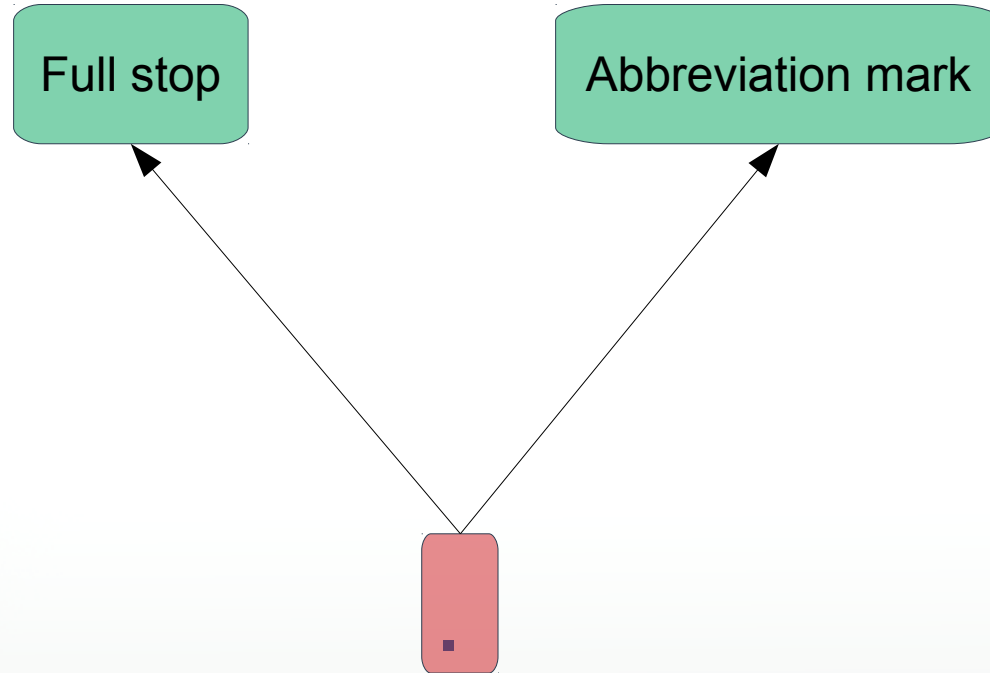
& (U+0026;
ASCII 38)

Historical/“etymological”
considerations



Issues: 4. Grapheme definition

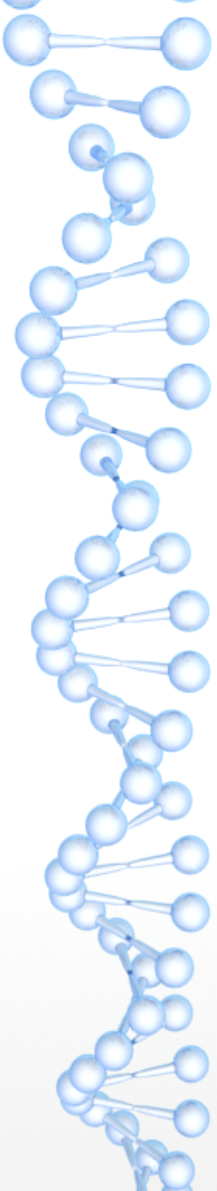
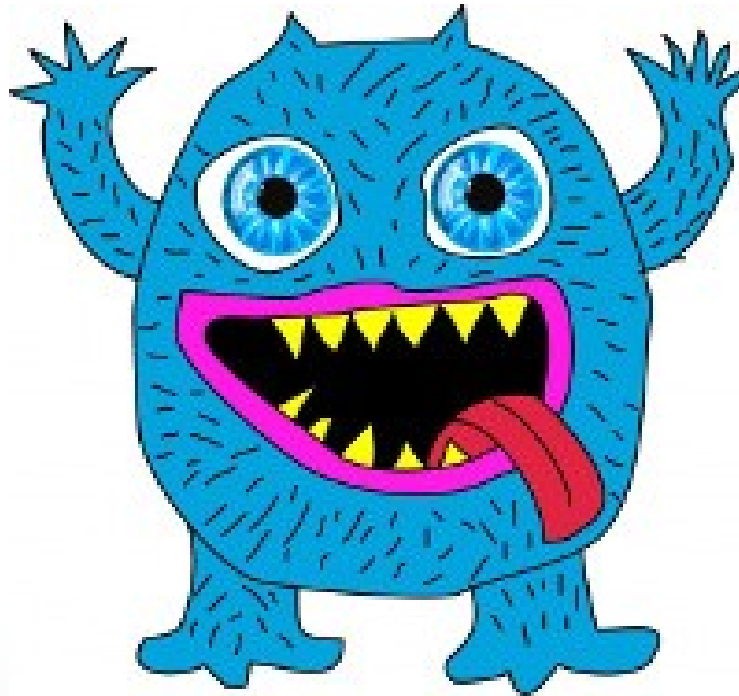
(paradigmatic)





Issues: 5. Allograph definition (metasemiology)

Issues: 5. Allograph definition (metasemiology)



Issues: 5. Allograph definition (metasemiology)

Allographs
(variants)

« σ »

• còparaEur

uł adfe

uładaliu

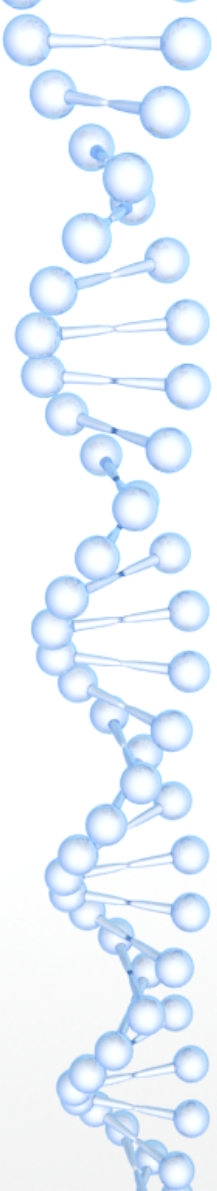
« $\sqrt{\quad}$ »

Issues: 5. Allograph definition (metasemiology)

« σ »
« σ »
« σ »
« σ »
« σ »
« σ »

- 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 a nutshell



In a nutshell

- The interoperability issue
- Interoperability through modelling
- Open issues