



[ve]dph

Venice Centre for  
Digital and Public  
Humanities



UNIVERSITÀ  
DEGLI STUDI  
DI PALERMO

# *L'edizione di Orso Beneventano*

Dalla trascrizione grafematica ai LOD

## **Paolo Monella**

*ALIM and beyond*

Ca' Foscari, 28 gennaio 2020

**LiLa**

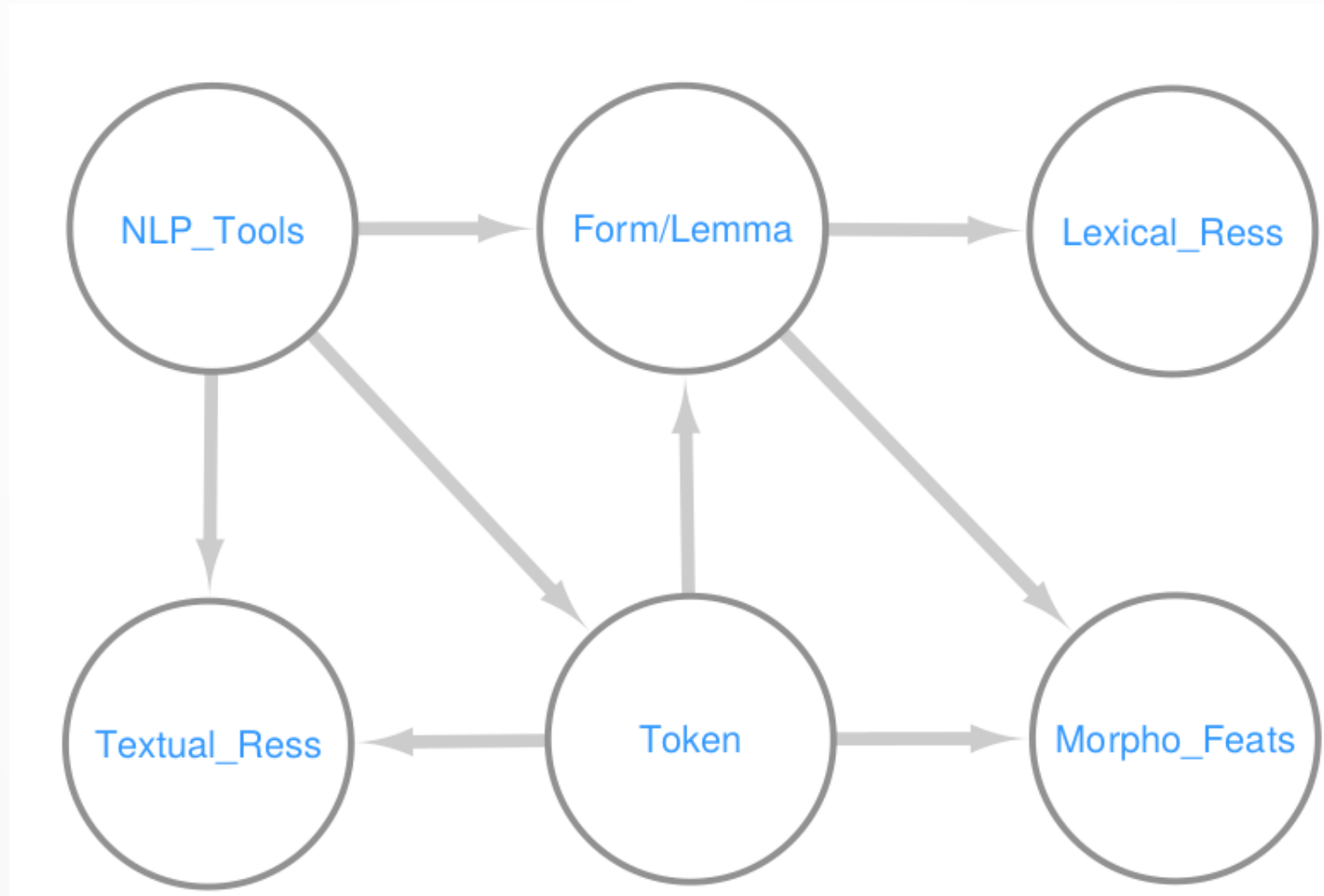
- Risorse testuali
- Risorse lessicali
- Strumenti NLP

- Risorse testuali → Orso Beneventano
- Risorse lessicali
  - Livello linguistico
  - Livello grafematico
- Strumenti NLP

**Orso / LiLa**  
**Livello linguistico**

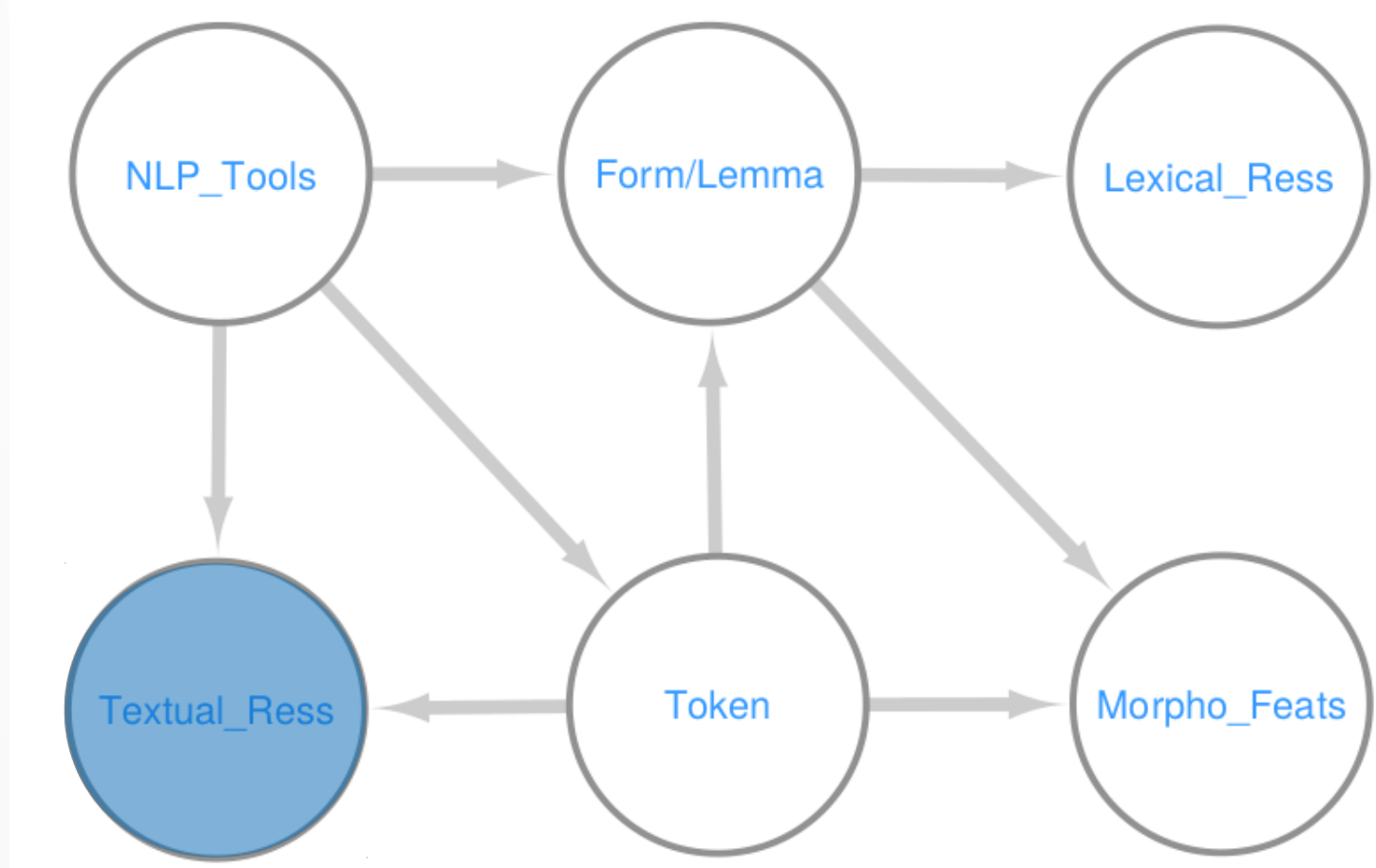
# Orso/LiLa

## Livello linguistico



From M. Passarotti, slideshow *LiLa: Linking Latin. A Knowledge Base of Linguistic Resources for Latin*, Conference *ALIM and beyond*, Venice 27-28 January 2020

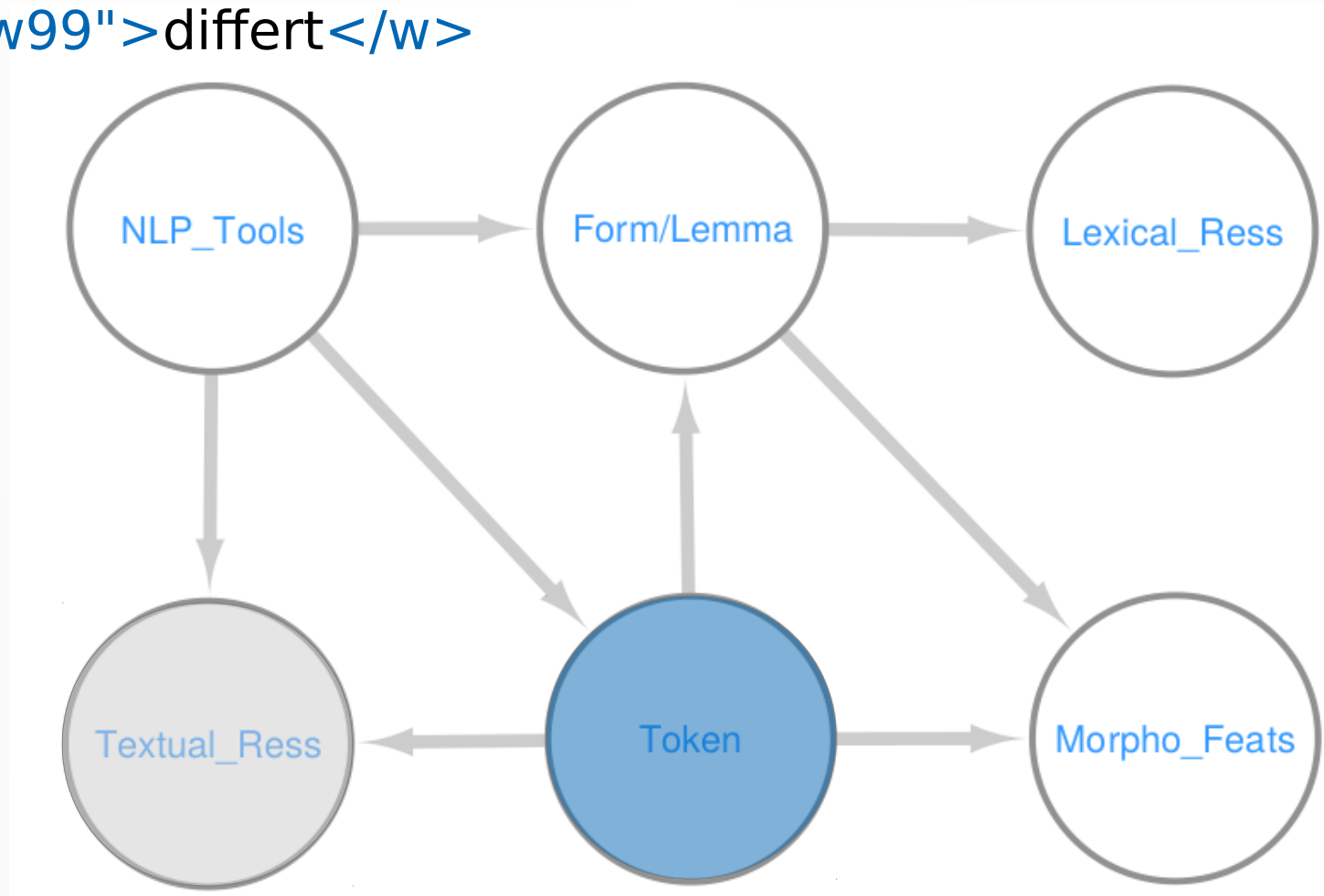
differt dictio a syllaba quoniam syllaba pars dictionis est



# Orso/LiLa

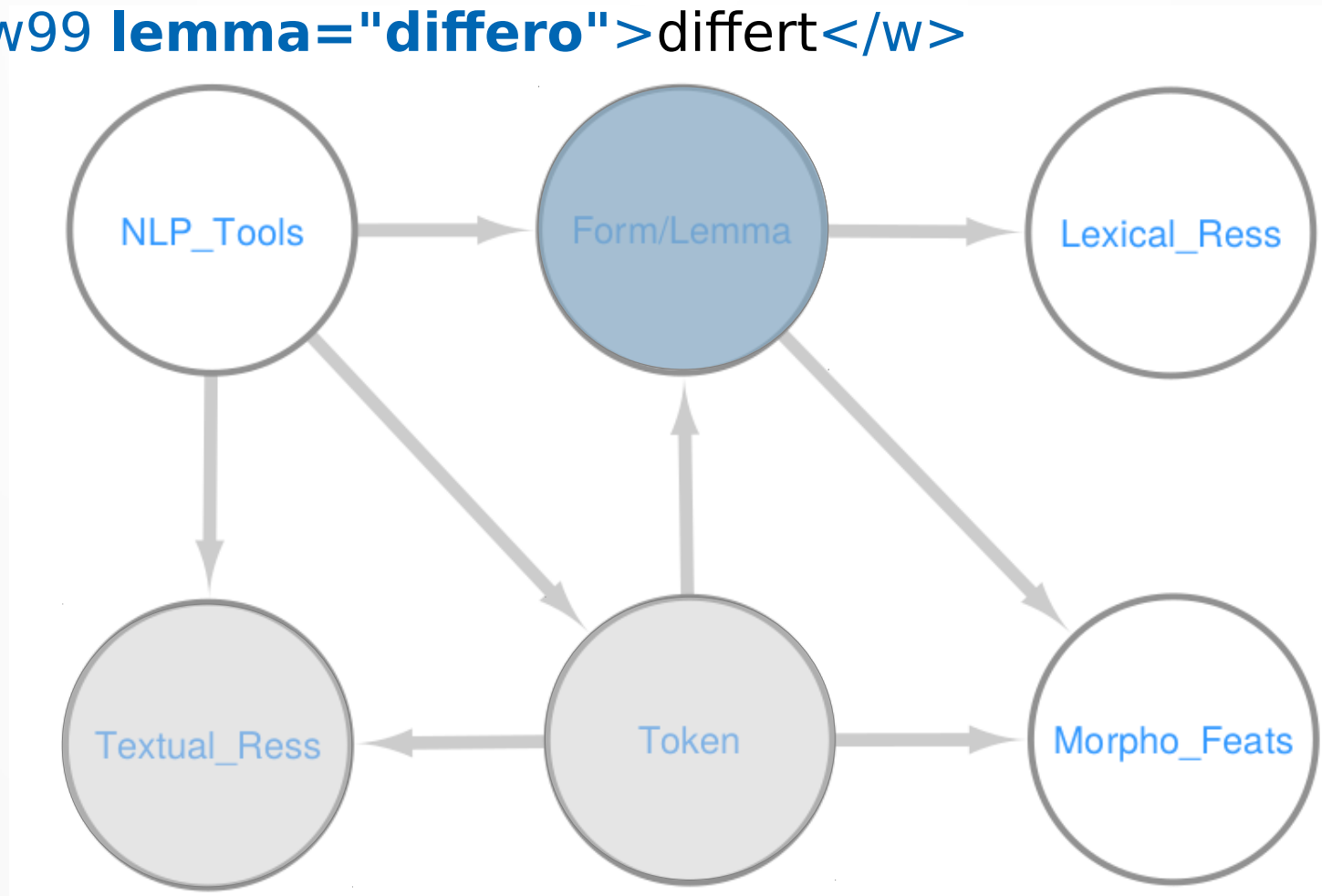
## Livello linguistico

`<w xml:id="w99">differt</w>`

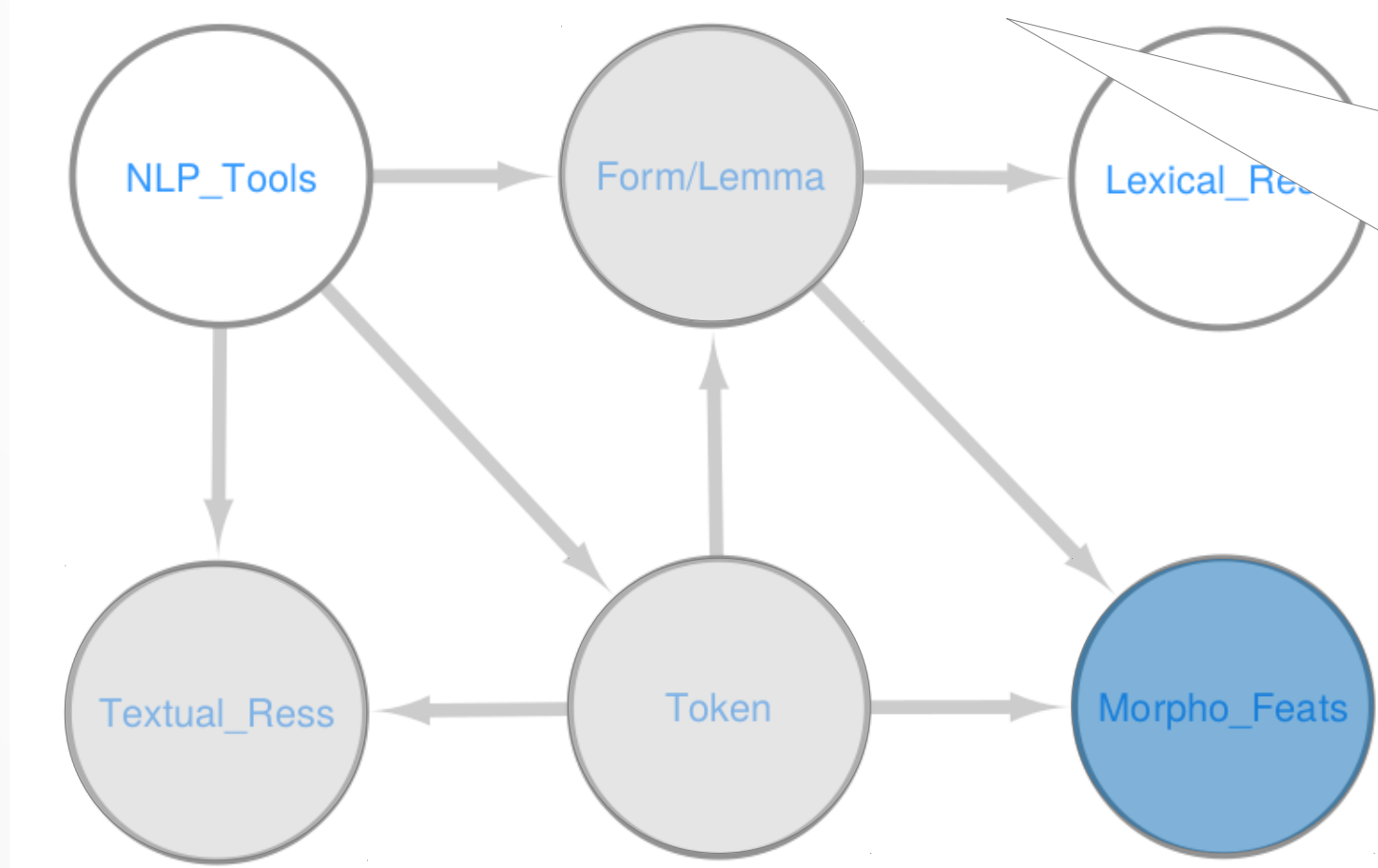




<w xml:id="w99 lemma="differo">differt</w>



<w xml:id="w99 lemma="differo" **msd="3-NA1--6---**" >differt</w>



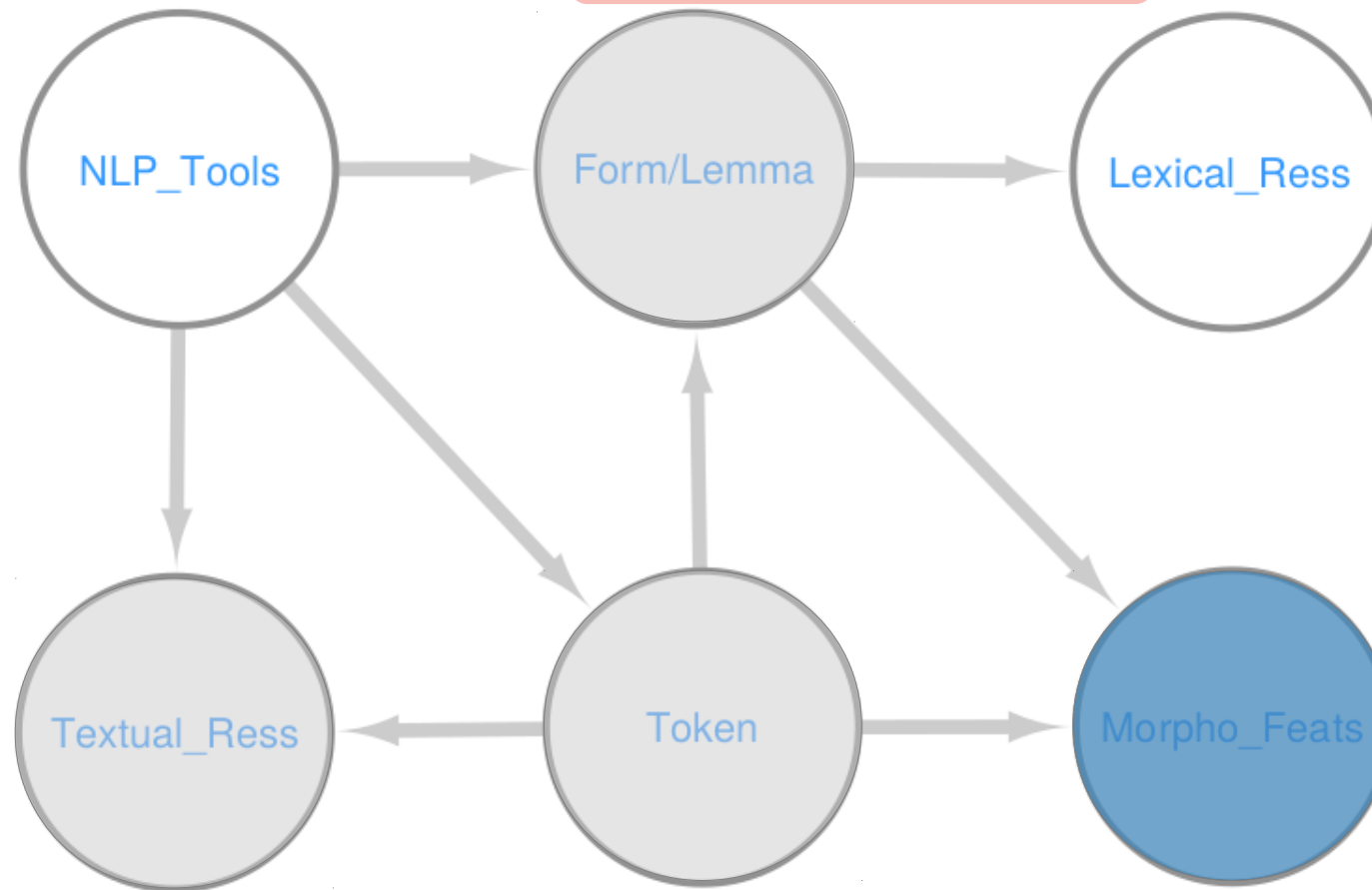
Verbal,  
Regularly  
irregular conjug,  
Active indicative,  
Present,  
III singular

(Software:  
TreeTagger;  
IT-TB da UD)

# Orso/LiLa

## Livello linguistico

<w xml:id="w99 lemma="differo" **msd="3-NA1--6----**">differt</w>

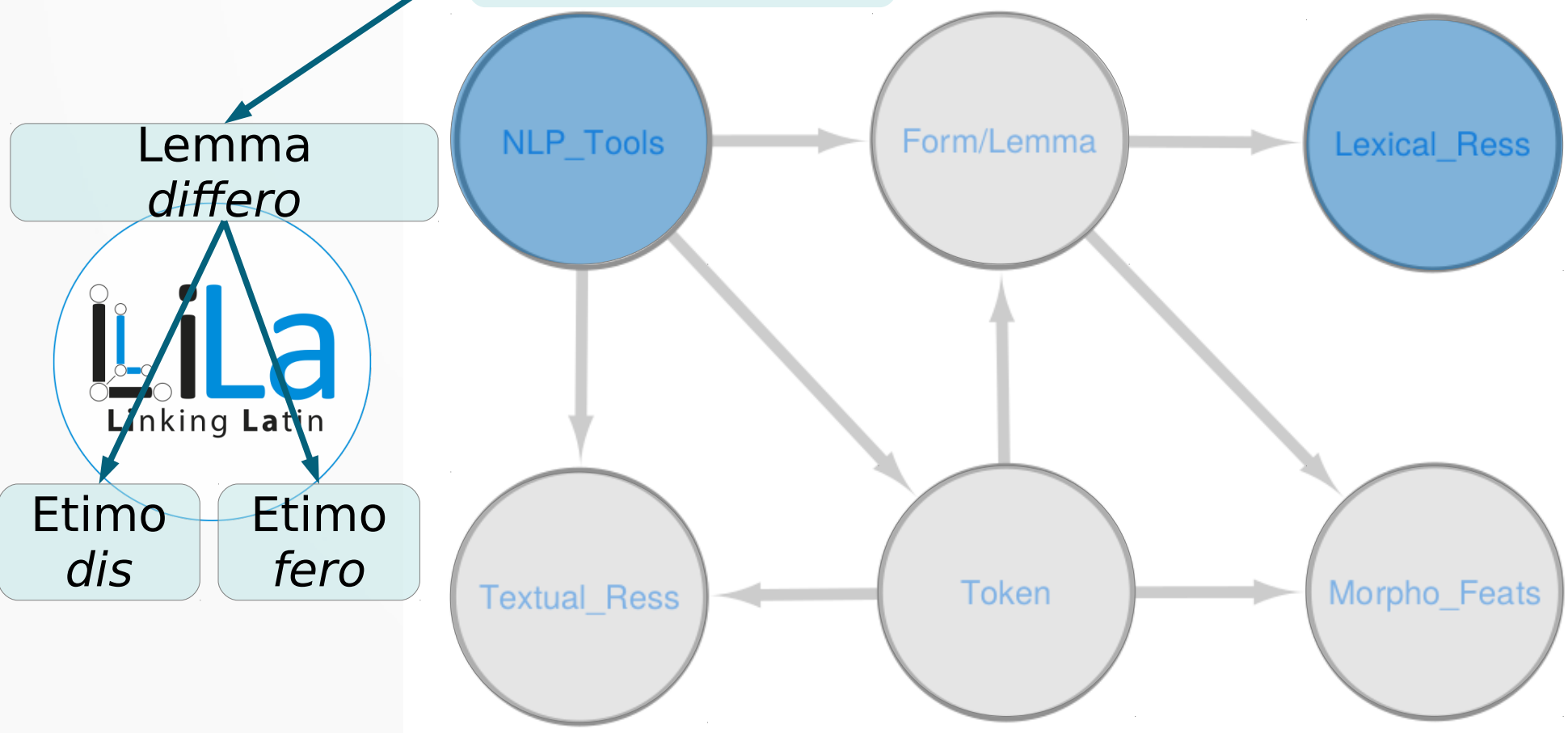


Aggiungere  
parte del  
discorso  
(tagset  
IT-TB →)

# Orso/LiLa

## Livello linguistico

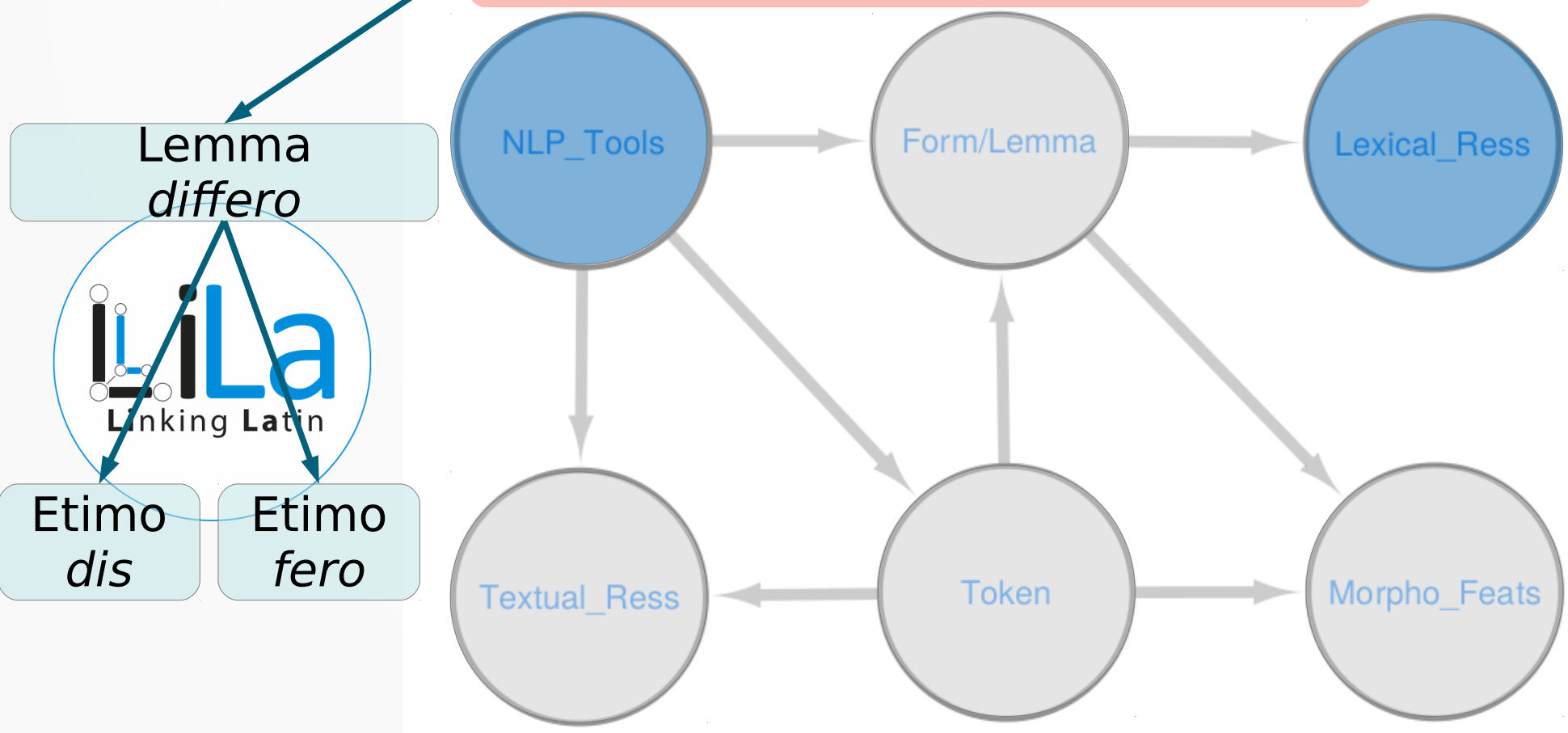
<w xml:id="w99 lemma="differo" msd="3-NA1--6---" >differt</w>



# Orso/LiLa

## Livello linguistico

<w xml:id="w99 lemma="differo" msd="3-NA1--6---" >differt</w>



**Revisione**

Test su  
397 parole:  
@lemma 82%  
@msd 79%

Lunghezza  
Testo: 14.354  
parole

**Orso / LiLa**  
**Livello grafematico**



# Orso/LiLa

## Livello grafematico

`<w>infiniti</w>`

`<pc n=",">•</pc>`

`<w>mutata</w>`

`<pc n="quote">•</pc>`

`<w>tum</w>`

`<pc n="quote">•</pc>`

Punteggiatura come  
'markup grafico'

# Orso/LiLa

## Livello grafematico

`<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>`

Punteggiatura come  
'markup grafico'

Spazi come  
'markup grafico'



# Orso/LiLa

## Livello grafematico

`<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>`

Distinzione tra  
'parola grafica' e  
'parola linguistica'

# Orso/LiLa

## Livello grafematico

`<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>`

Distinzione tra

'parola grafica' e

'parola linguistica'

# Orso/LiLa

## Livello grafematico

`<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>`

Distinzione tra

‘parola grafica’ e

‘parola linguistica’

# Orso/LiLa

## Livello grafematico

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

**ordin**

```
<choice>
```

```
<abbr type="superscription">
```

```
  e<am>2</am>
```

```
</abbr>
```

```
<expan>em</expan>
```

```
</choice>
```

```
</w>
```

Abbreviazioni

# Orso/LiLa

## Livello grafematico

**ordin**

ordin<sup>2</sup>

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

**quo**

quoq;

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

**bus**

qbus

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

**cor**

cor

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

**bus**

Tipi di abbreviazione



# Orso/LiLa

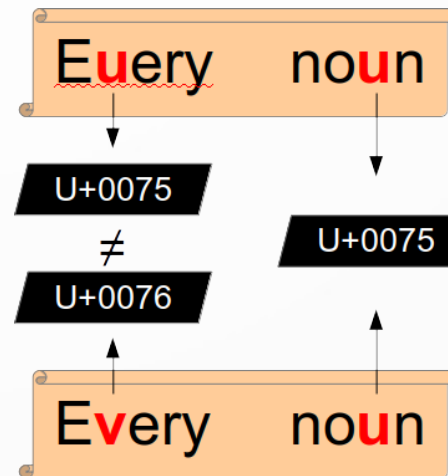
## Livello grafematico

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 . , :

# Orso/LiLa

## Livello grafematico

<u>Grapheme</u>	<u>Alphabeme(s)</u>	<u>Grapheme visualization</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>e</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>

Tabella dei segni

# Orso/LiLa

## Livello grafematico

<u>Grapheme</u>	<u>Alphabeme(s)</u>	<u>Grapheme visualization</u>	<u>Type</u>
k	k	k	<u>Alphabetic</u>
l	l	l	<u>Alphabetic</u>
m	m	m	<u>Alphabetic</u>
n	n	n	<u>Alphabetic</u>
o	o	o	<u>Alphabetic</u>
p	p	p	<u>Alphabetic</u>
q	q	q	<u>Alphabetic</u>
r	r	r	<u>Alphabetic</u>
s	s	s	<u>Alphabetic</u>
t	t	t	<u>Alphabetic</u>

Tabella dei segni



# Orso/LiLa

## Livello grafematico

<u>Grapheme</u>	<u>Alphabeme(s)</u>	<u>Grapheme visualization</u>	<u>Type</u>
u	u	u	<u>Alphabetic</u>
x	x	x	<u>Alphabetic</u>
y	y	y	<u>Alphabetic</u>
z	z	z	<u>Alphabetic</u>
			Space
.		.	<u>Punctuation</u>
.		.	<u>Punctuation</u>
		,	<u>Punctuation</u>
~		~	<u>Punctuation</u>
:		:	<u>Punctuation</u>

Tabella dei segni

# Orso/LiLa

## Livello grafematico

<u>Grapheme</u>	<u>Alphabeme(s)</u>	<u>Grapheme visualization</u>	<u>Type</u>
‡		;	<u>Punctuation</u>
¨		¨	<u>Punctuation</u>
¡		¡	<u>Punctuation</u>
8		ˆ	<u>Abbreviation mark</u>
2		˘	<u>Abbreviation mark</u>
7		˙	<u>Abbreviation mark</u>

Tabella dei segni

# Orso/LiLa

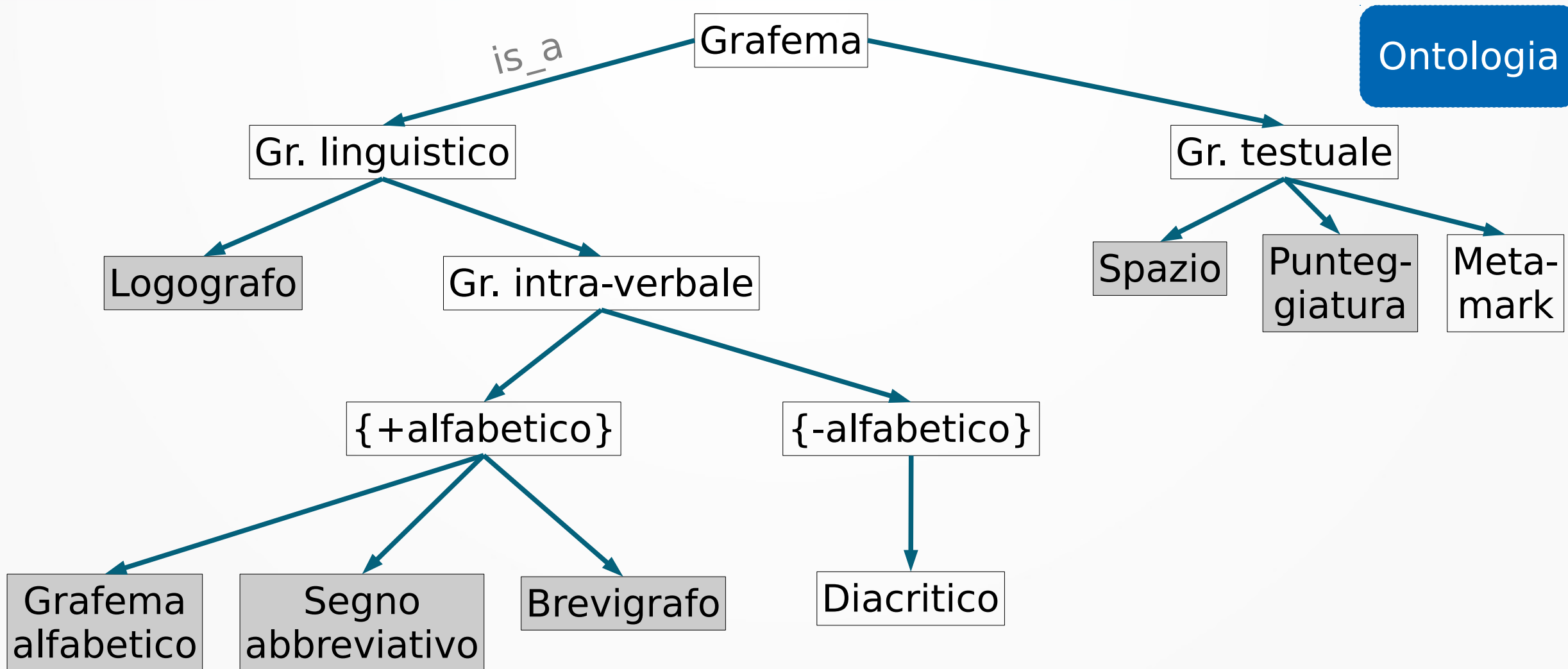
## Livello grafematico

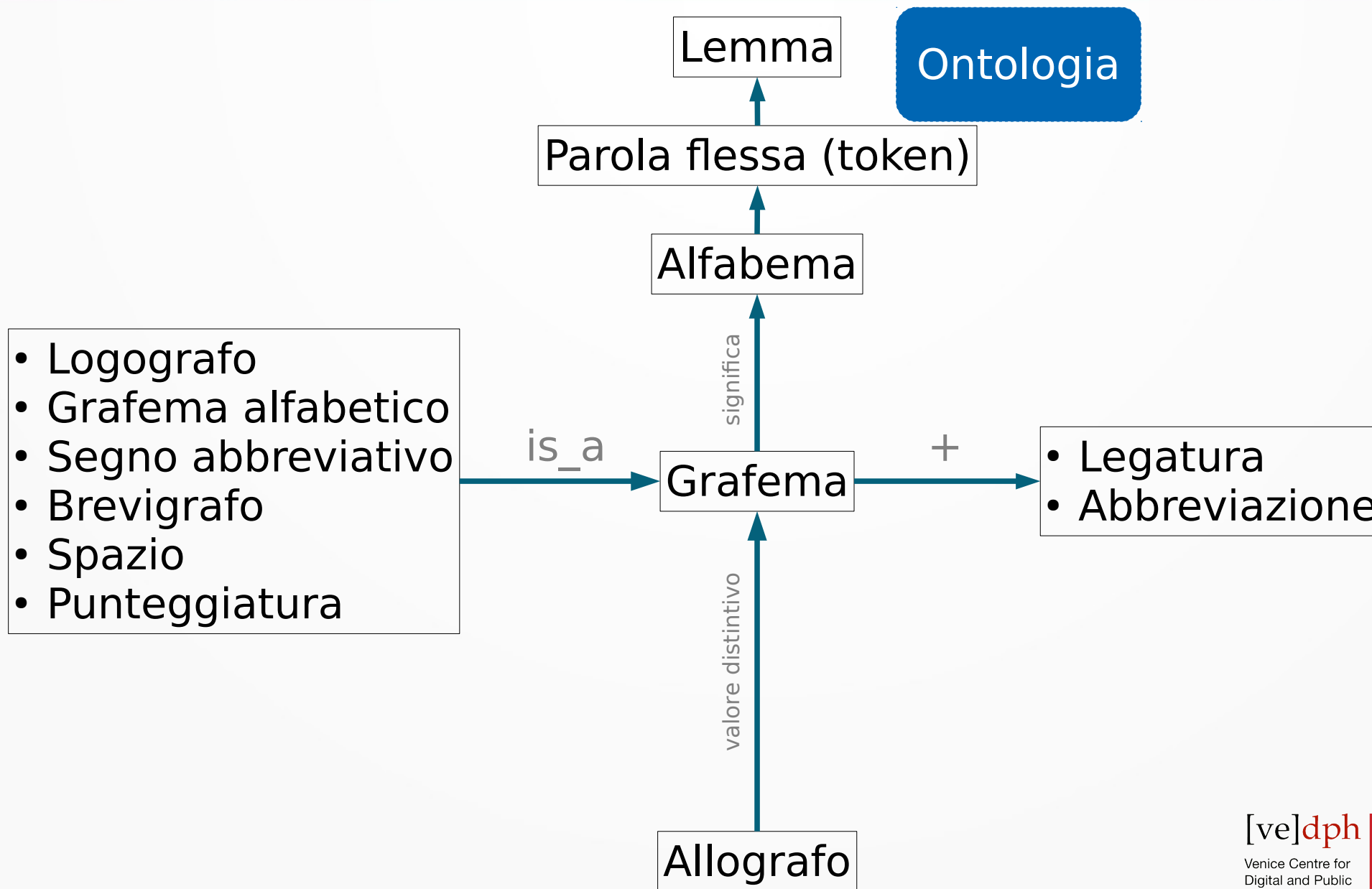
<u>Grapheme</u>	<u>Alphabeme(s)</u>	<u>Grapheme visualization</u>	<u>Type</u>
—		◌̄	<u>Abbreviation mark</u>
;		;	<u>Abbreviation mark</u>
ȳ	u	◌̄	<u>Abbreviation mark</u>
<u>p</u>	per	<u>p</u>	<u>Brevigraph</u>
<u>ŋ</u>	pro	<u>p</u>	<u>Brevigraph</u>
<u>ç</u>	qui	<u>q</u>	<u>Brevigraph</u>
÷	est	≈	<u>Logograph</u>

Tabella dei segni

# Orso/LiLa

## Livello grafemico







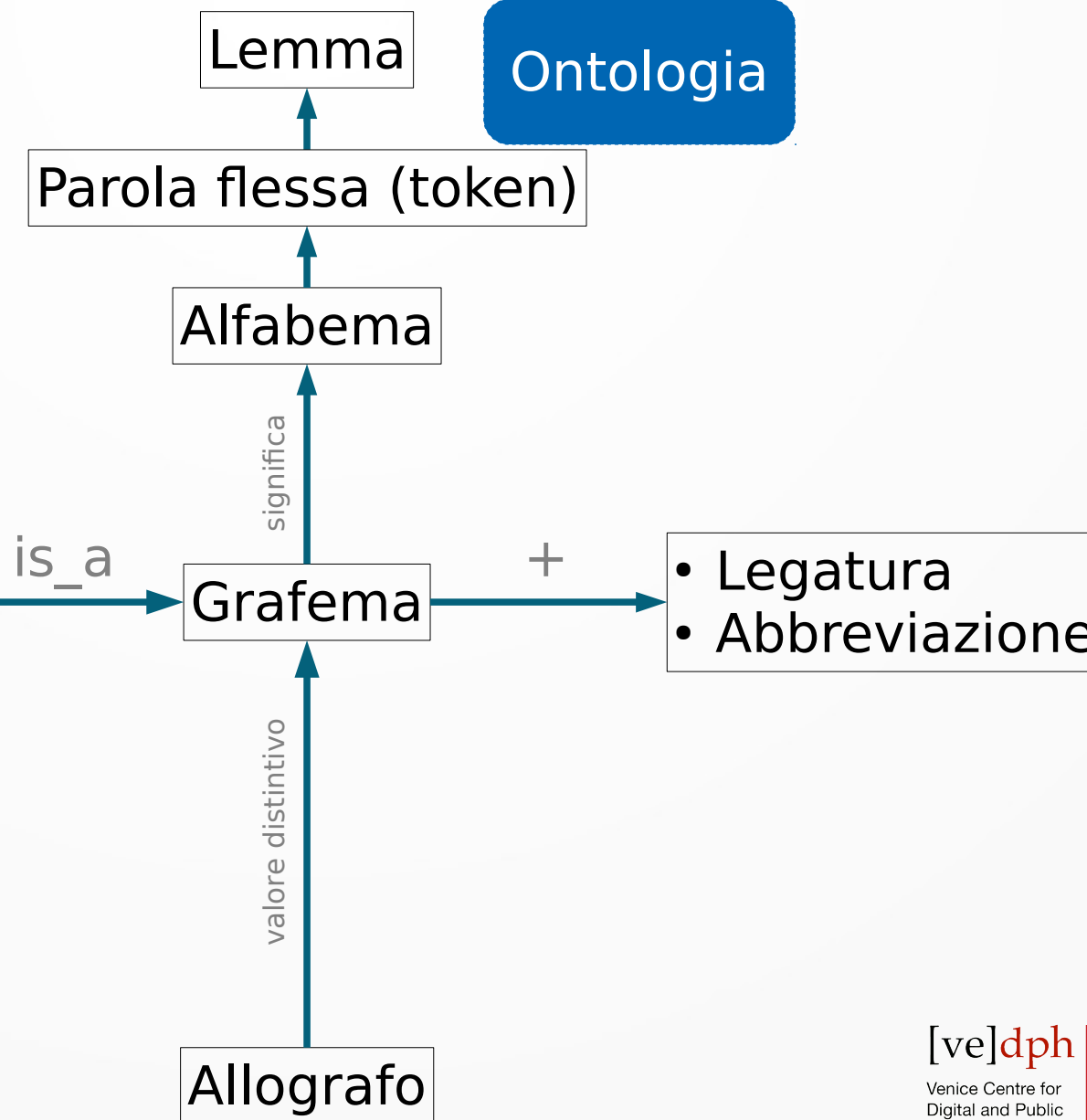
# Orso/LiLa

## Livello grafemico

Orso (risorsa testuale)

```
<pc n=",">•</pc>
```

- Logografo
- Gr. alfabetico
- Segno abbreviativo
- Brevigrafo
- Spazio
- Punteggiatura



# Orso/LiLa

## Livello grafemico

Orso (risorsa testuale)

```
<pc n=",">•</pc>
```

- Logografo
- Gr. alfabetico
- Segno abbreviativo
- Brevigrafo
- Spazio
- Punteggiatura



Lemma

Ontologia

Parola flessa (token)

Alfabema

significa

is\_a

Grafema

+

- Legatura
- Abbreviazione

valore distintivo

Allografo

# Conclusioni



# Conclusioni

- Risorse testuali
- Risorse lessicali
- Strumenti NLP



# Conclusioni

- Risorse testuali → Orso Beneventano
- Risorse lessicali
- Strumenti NLP



# Conclusioni

- Risorse testuali → Orso Beneventano
- Risorse lessicali – Livello linguistico
- Strumenti NLP



– Livello grafematico

# Conclusioni

- Risorse testuali → Orso Beneventano
- Risorse lessicali
  - Livello linguistico
    - Revisione lemmatizzazione e PoS tagging Orso
  - Livello grafematico
- Strumenti NLP



# Conclusioni

- Risorse testuali → Orso Beneventano
  - Risorse lessicali
  - Strumenti NLP
- Livello linguistico
    - Revisione lemmatizzazione e PoS tagging Orso
    - Connessione Orso/ LiLa
  - Livello grafematico



# Conclusioni

- Risorse testuali → Orso Beneventano
  - Risorse lessicali
  - Strumenti NLP
- Livello linguistico
    - Revisione lemmatizzazione e PoS tagging Orso
    - Connessione Orso/ LiLa
  - Livello grafematico
    - Elaborazione ontologia (VeDPH)





# Conclusioni

• Risorse testuali → Orso Beneventano

• Risorse lessicali

• Strumenti NLP

– Livello linguistico

- Revisione lemmatizzazione e PoS tagging Orso

- Connessione Orso/ LiLa

– Livello grafematico

- Elaborazione ontologia (VeDPH)

- Implementazione ontologia in LiLa



# Conclusioni

• Risorse testuali → Orso Beneventano

• Risorse lessicali

• Strumenti NLP



– Livello linguistico

- Revisione lemmatizzazione e PoS tagging Orso

- Connessione Orso/ LiLa

– Livello grafematico

- Elaborazione ontologia (VeDPH)

- Implementazione ontologia in LiLa

- Connessione Orso / LiLa