

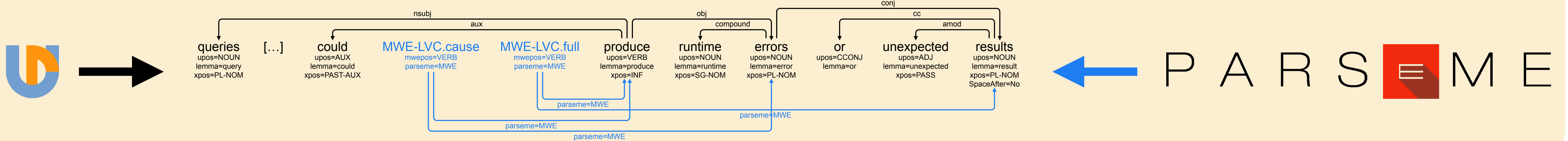
# Multi-layer querying in Corpora: Example of Parseme and UD

Bruno Guillaume — LORIA / Inria Nancy Grand-Est

Most of images and pieces of code are clickable on the electronic version!

## Universal dependencies and Parseme annotations in the same structure

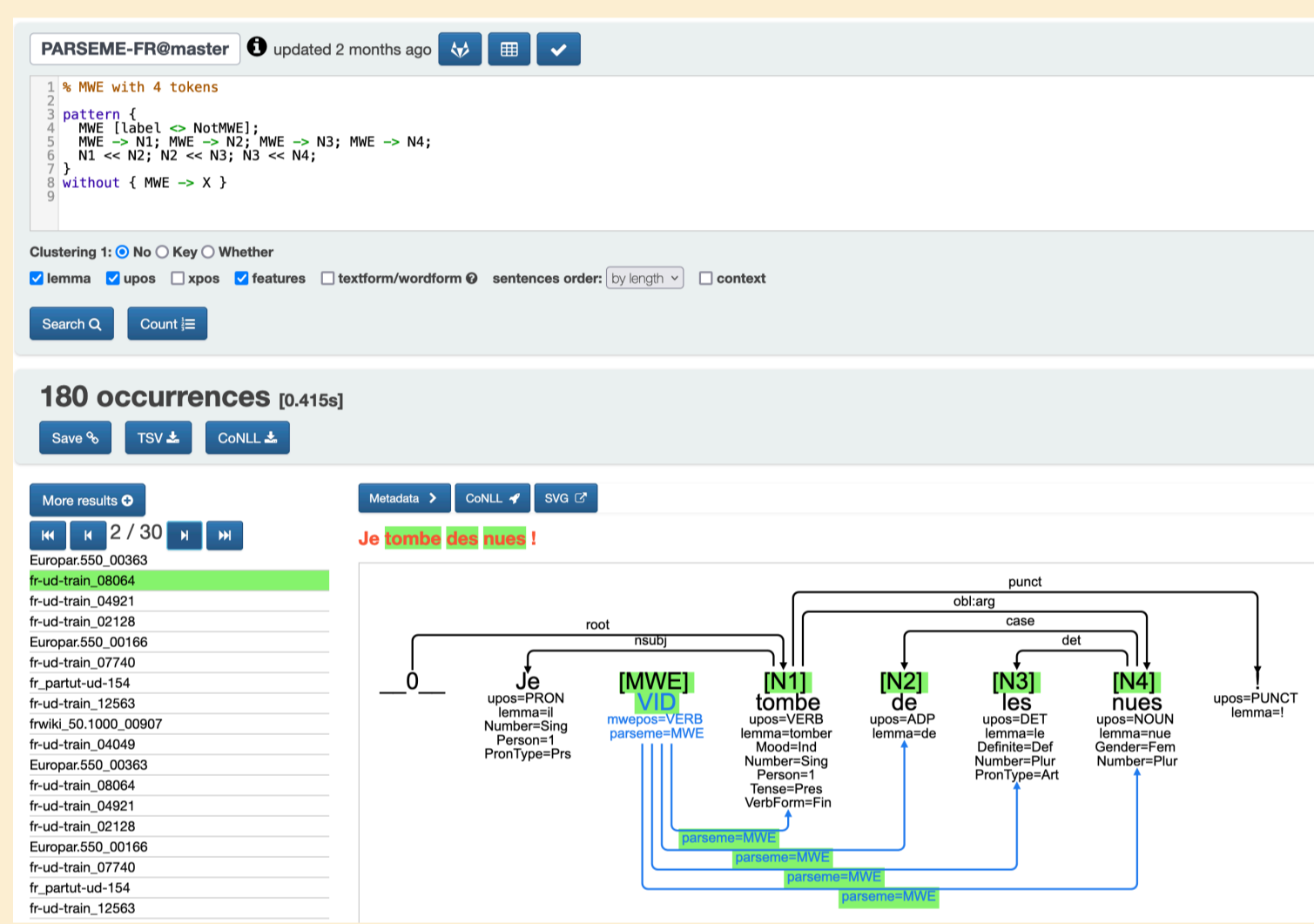
- Parseme **MWEs** can be **discontinuous**
- Several Parseme **MWEs** can **overlap**
- The two **layers** can be encoded in a **single graph structure**
- Each **MWE** is a new **node**
- Edges** map each MWE node to all its **tokens**



## Parseme: online queries on 26 languages



<http://parseme.grew.fr>



Je tombe des nues !  
I fall off the clouds (old form) !  
'I can't believe it!'

## Error Mining: UD/Parseme consistency

a MWE with only one token?

```
1 pattern {
2   MWE [label <> NotMWE];
3   MWE -> N;
4 }
5 without { MWE -> X }
```

Language	Occ
Hungarian	5745
Chinese	4416
Swedish	1616
German	1268
Hebrew	42

a MWE without any verb?

```
1 pattern {
2   MWE [label <> NotMWE]
3 }
4 without {
5   MWE -> V;
6   V [upos=VERB | AUX]
7 }
```

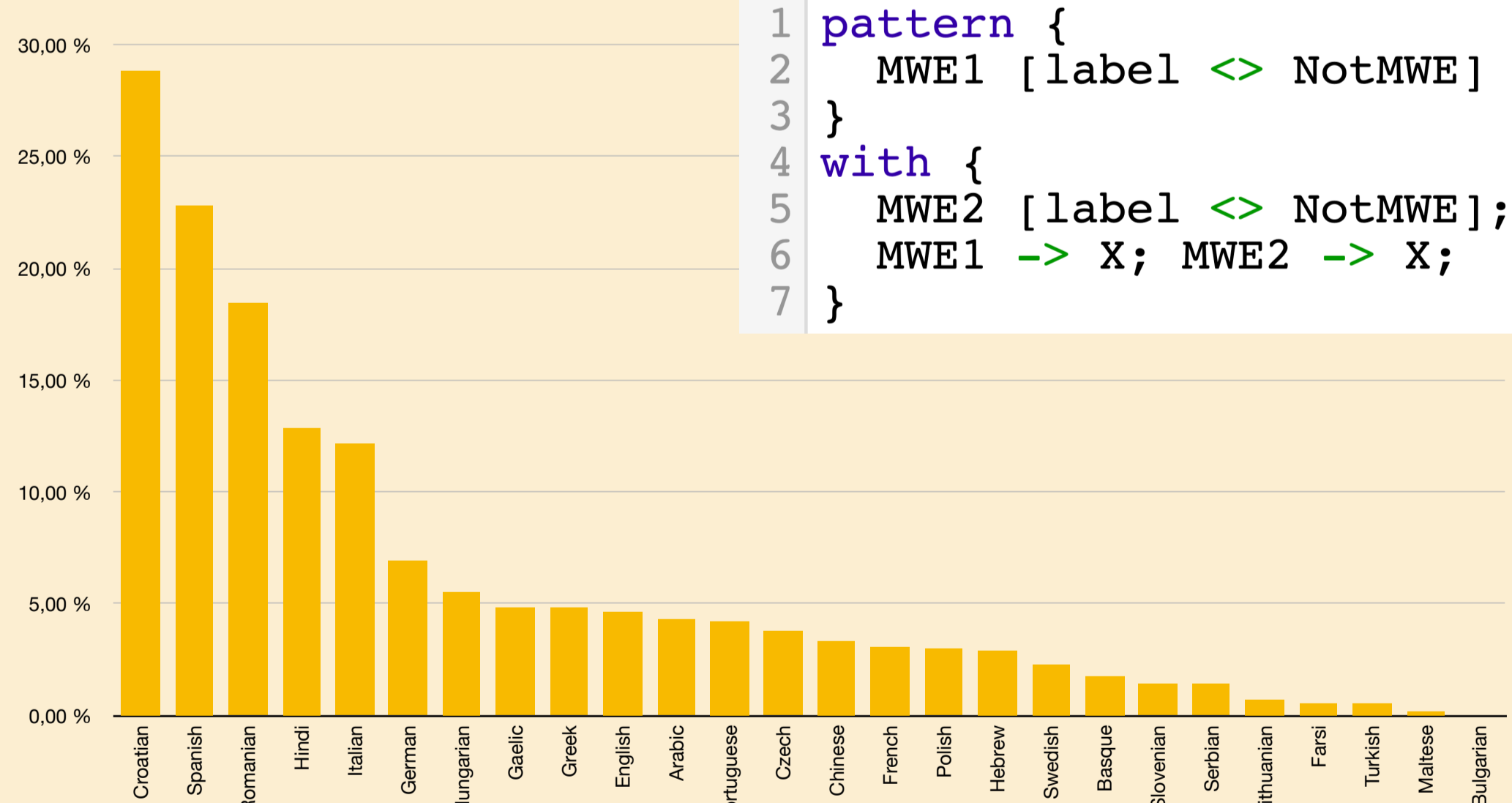
Language	Occ
Hungarian	5901
Slovenian	3378
Arabic	1291
Polish	870
Czech	790

an IRV without a reflexive pronoun?

```
1 pattern {
2   MWE [label = "IRV"];
3 }
4 without {
5   MWE -> P;
6   P [upos=PRON, Reflex=Yes]
7 }
```

Language	Occ
Slovenian	1631
Italian	1144
Portuguese	1021
Swedish	237
Romanian	206

## ratio of overlapping relations



```
1 pattern {
2   MWE1 [label <> NotMWE]
3 }
4 with {
5   MWE2 [label <> NotMWE];
6   MWE1 -> X; MWE2 -> X;
7 }
```

## Lemmas used in MVC annotations

English

```
1 pattern {
2   MWE [label="MVC"];
3   MWE -> N1; MWE -> N2; N1 << N2
4 }
```

N1.Lemma	N2.Lemma	45	know	rid	examine	go
let		45				1
get			4			
cross						1

French

N1.Lemma	N2.Lemma	7	parler	remarquer	savoir	tomber	valoir	passer
faire		12	4	4		3		1
entendre		7						
laisser					3			

## Conclusion

**Graphs** can be used as a efficient way of connecting different **annotation layers**

**Grew** implements graph-based structures for NLP:

- Pattern Graph matching** in **Grew-match** (linguistic observations and error mining)
- Graph Rewriting** in **Grew** (conversion, consistent updates)
- Grew Command Line Interface
- Python Library**
- Grew-web** (online rewriting, for testing and debugging)



<https://grew.fr>