Grew-match tutorial

Bruno Guillaume
LORIA / Inria Nancy Grand-Est

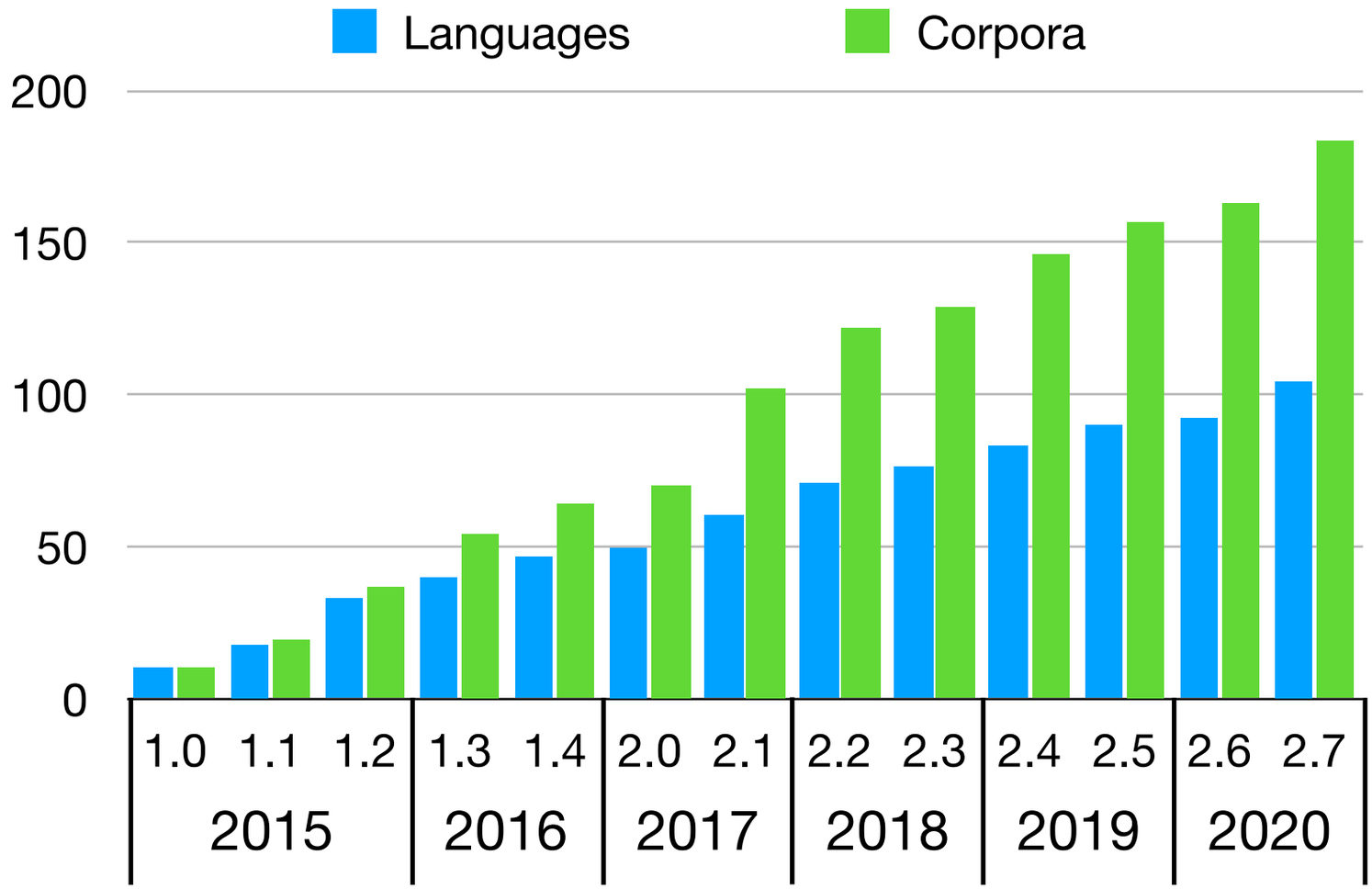
Sémagramme meeting
February 16, 2021

Universal Dependencies

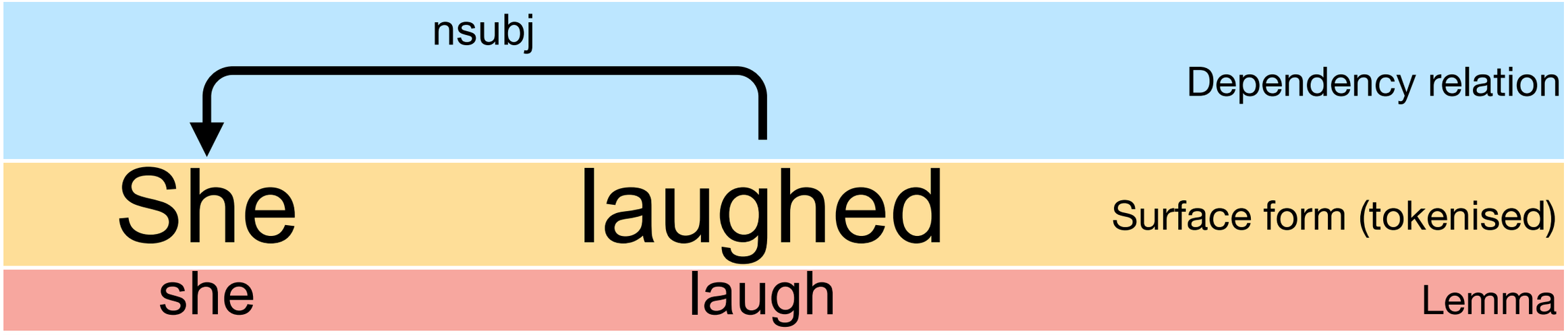
Universal Dependencies (UD):

- Collaborative project of “universal” dependency annotations
- Version 2.7: 104 languages, 183 corpora

http://universaldependencies.org



Universal Dependencies



PRON	VERB	POS
Case=Nom Gender=Fem Number=Sing Person=3 PronType=Prs	Mood=Ind Tense=Past VerbForm=Fin	Morphology

Grew & Grew-match

.grew https://grew.fr

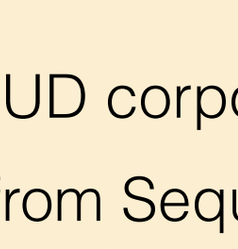
- Graph rewriting based software
- Graph rewriting adapted to NLP specificities
- Developed in the Sémagramme team (with G. Bonfante) in the last 10 years
- More on this in an upcoming tutorial…

.grew-match http://match.grew.fr

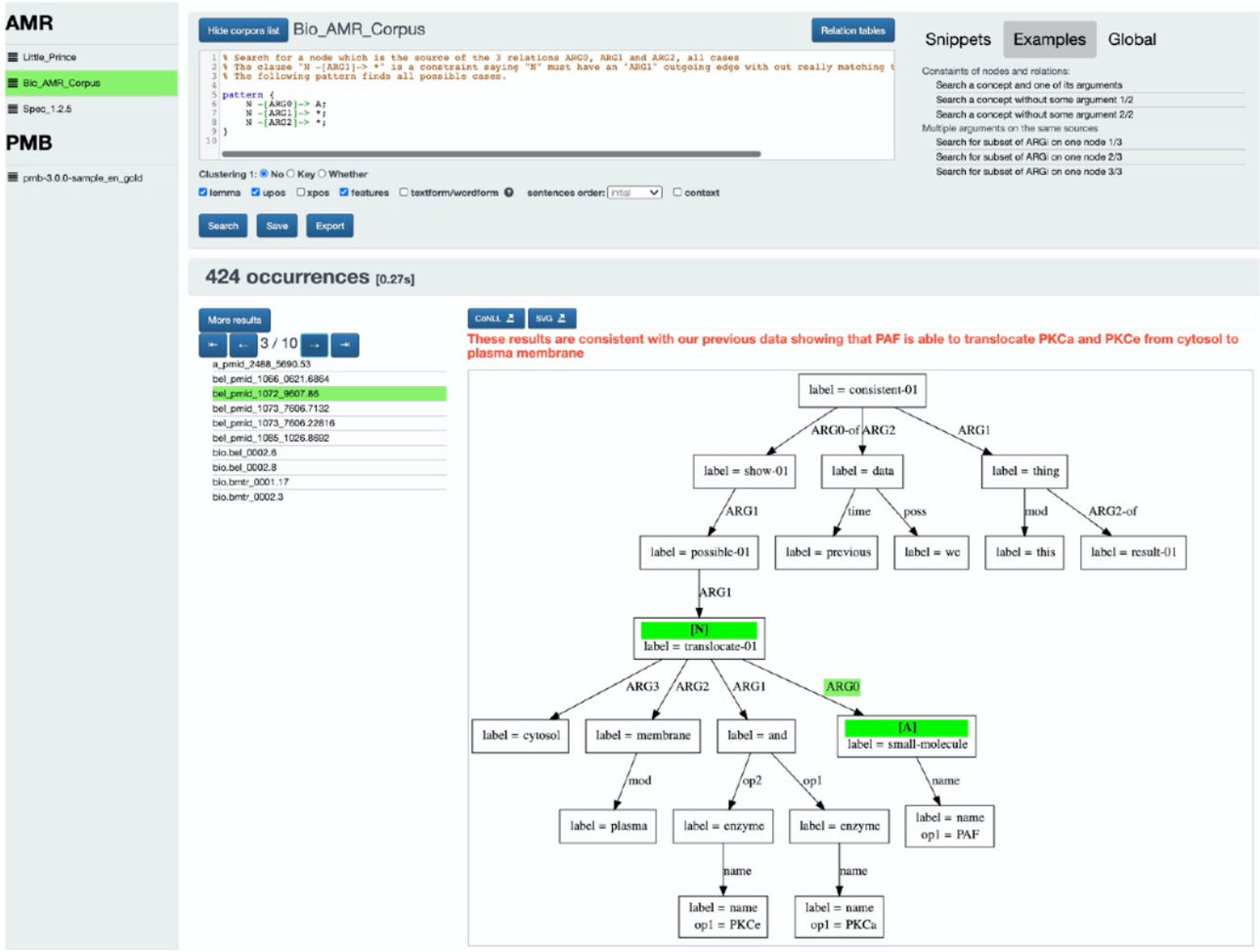
- In Grew: rule = (pattern, commands)
- In Grew-match, we use only the “pattern” part
- (graph, pattern) ⇨ list of occurrences
- (corpus, pattern) ⇨ list of occurrences

Grew-match

- Available online on a large set of corpora
  - All (S)UD corpora
  - Data from Sequoia, Parseme and Orfeo projects
  - Some semantics graphs
    - AMR (The Little Prince, Bio_AMR_Corpus)
    - 10 sentences from the PMB (ongoing project…)

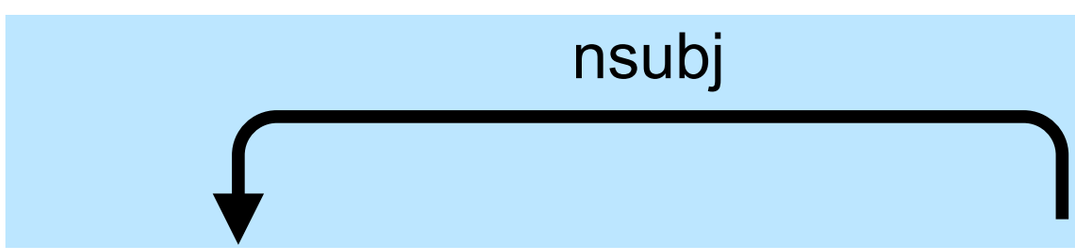


http://match.grew.fr



UD & Grew-match

UD annotation



She
she
PRON
Case=Nom
Gender=Fem
Number=Sing
Person=3
PronType=Prs

laughed
laugh
VERB
Mood=Ind
Tense=Past
VerbForm=Fin

Grew-match syntax

```
pattern { M -[nsubj]-> N }
```

```
pattern { N [form="laughed"] }
```

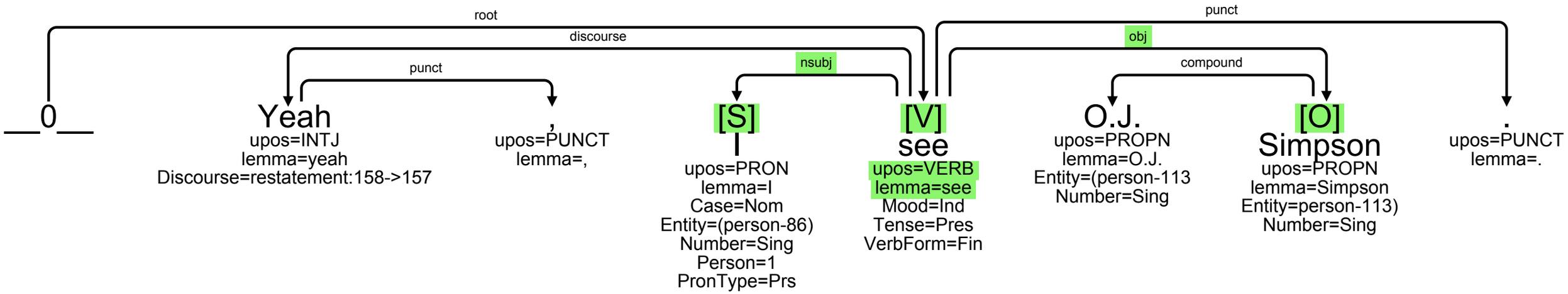
```
pattern { N [lemma="laugh"] }
```

```
pattern { N [upos=VERB] }
```

```
pattern { N [Tense=Past] }
```

Grew-match syntax

```
pattern { V [upos=VERB, lemma="see"]; V -[nsubj]-> S; V -[obj]-> O }
```



Your turn · Step 1

```
pattern { M -[nsubj]-> N }
pattern { N [form="laughed"] }
pattern { N [lemma="laugh"] }
```

```
pattern { N [upos=VERB] }
pattern { N [Tense=Past] }
```

```
pattern { V [upos=VERB, lemma="see"]; V -[nsubj]-> S; V -[obj]-> O }
```

```
pattern { N [lemma <> "see"] }
```

The lemma is different from “see”

```
pattern { N [upos=VERB | AUX] }
```

The POS is either VERB or AUX

1. In the corpus UD_English-GUM (selected by default), what are the words used with the POS “PART”?
2. In the corpus UD_French-GSD (left pane), what are the possible lemmas for POS “AUX”?
3. Chose the corpus you want, observe if it is possible to have two subjects on the same verb, two objects on the same verb.

Add `% your_name` as first line in your requests!

Grew-match · more syntax

`N1 < N2` The node N1 is immediately before N2

```
pattern {
 N1 [upos=DET]; N2 [upos=NOUN];
 N1 < N2 }
```

`N1 << N2` The node N1 is before N2

```
% left-headed nsubj
pattern { G -[nsubj]-> D; G << D }
```

`N1.f = N2.f`

`N1.f <> N2.f`

(In)equality of features

```
pattern {
 N1 [upos=DET]; N2 [upos=NOUN];
 N1 < N2;
 N1.Gender <> N2.Gender }
```

`without` Filter out some occurrences

```
pattern { V [upos=VERB] }
without { V -[nsubj]-> S }
```

```
pattern { N1 [upos=DET]; N2 [upos=NOUN]; N1 < N2 }
without { N2 -[det]-> N1}
```

Your turn · Step 2

```
pattern { V [upos=VERB, lemma="see"]; V -[nsubj]-> S; V -[obj]-> O }
```

```
pattern { N [lemma<>"see"] }
```

```
pattern { N [upos=VERB | AUX] }
```

```
pattern { N1 [upos=DET]; N2 [upos=NOUN]; N1 < N2 }
```

```
pattern { G -[nsubj]-> D; G << D }
```

```
pattern { N1 [upos=DET]; N2 [upos=NOUN]; N1 < N2; N1.Gender <> N2.Gender }
```

```
pattern { V [upos=VERB] } without { V -[nsubj]-> S }
```

```
pattern { N1[upos=DET]; N2[upos=NOUN]; N1 < N2 } without { N2 -[det]-> N1}
```

1. How the trigram “in order to” is annotated in English corpora?
   ➡ use “n-grams” from the snippets (on the right of the textarea)
2. Explore Verb/Subject number agreement in French
   ➡ use several successive `without` clauses

Grew-match • Other stuff

- **Relation tables**: See in one table what are the POS used for a given relation

See *aux*, *amod*

- **Clustering**: put occurrences of a request in clusters, given a clustering key

French AUX lemmas

Relations between two verbs

- **Whether**: split occurrences given by the pattern in two clusters:
  - the Yes cluster where the additional contraints hold
  - the No clusters where the additional contraints do not hold (without)

How much is the *amod* left-headed in French?

Does NOUN alway have a determiners?

How to make stats running several requests on a set of 183 corpus?

- use Grew command line interface
- dedicated web service: to be checked…

Grew-match • What's next?

- More corpora:
  - Focus on semantically annotated data (PMB…)
  - Difficulty of building readable graphical display
- Other kinds of graph used in NLP
  - Lexical database: experiment in RLF
- Multi-corpora request in the web interface
- Double clustering