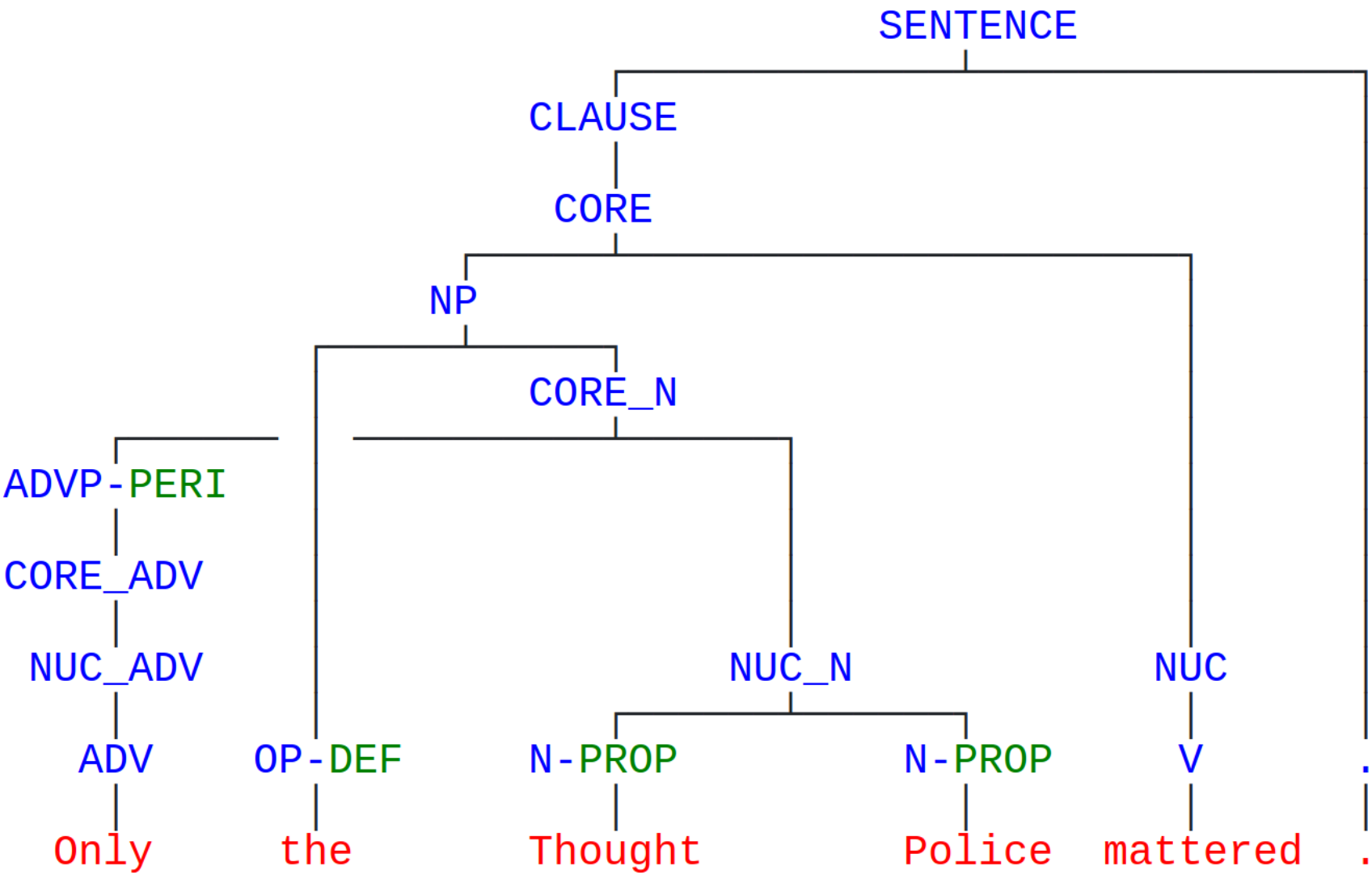
Cross-lingual RRG Parsing

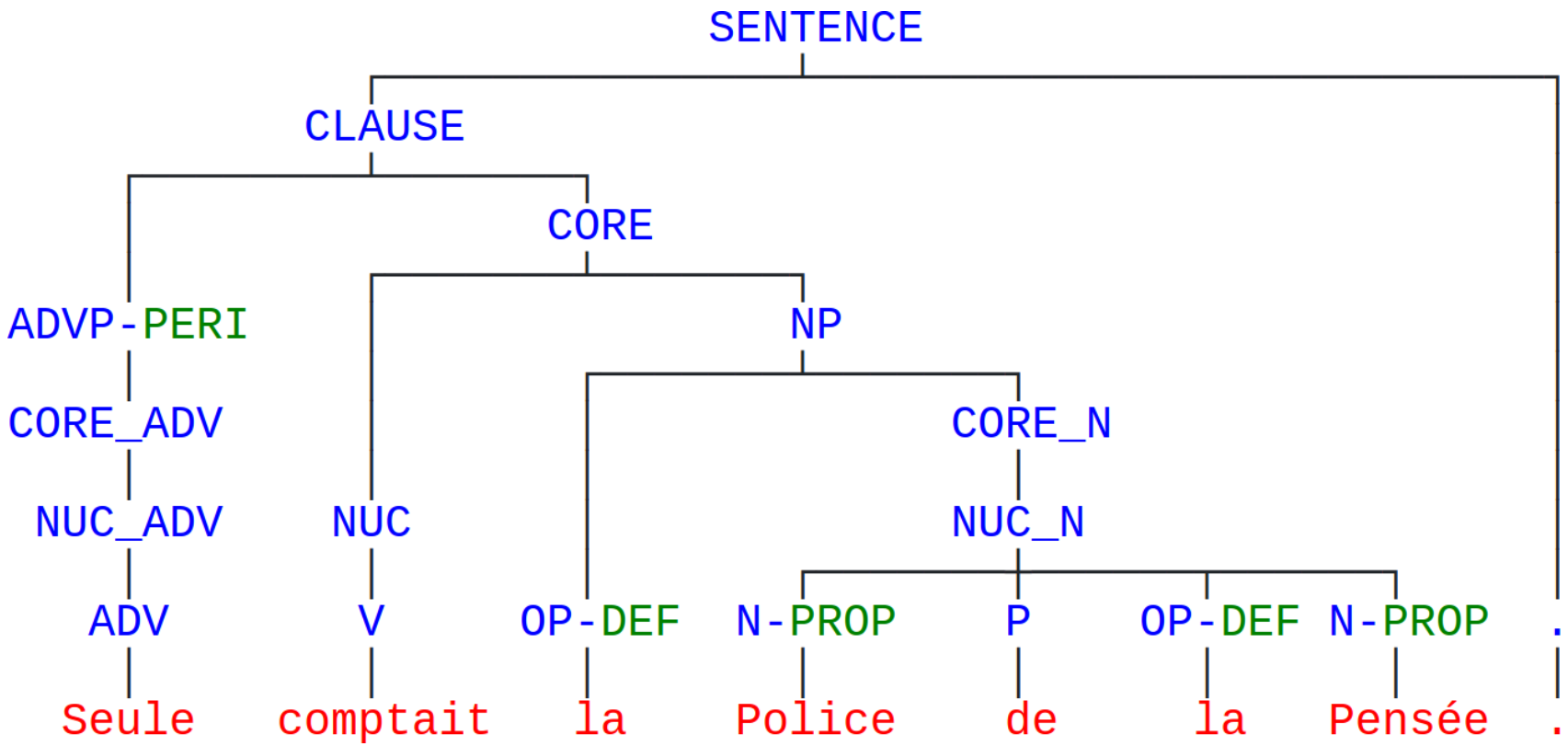
Jakub Waszczuk Kilian Evang

2020-11-11
TreeGraSP Meeting #5

RRGparbank: English example



RRGparbank: French example



Supertags

```
Only      (CORE_N* (ADVP-PERI (CORE_ADV (NUC_ADV (ADV <>)))))
the       (CORE_N* (OP-DEF <>))
Thought   (NUC_N* (N-PROP <>))
Police    (NP (CORE_N (NUC_N (N-PROP <>))))
mattered  (SENTENCE (CLAUSE (CORE (NP ) (NUC (V <>)))))
.         (SENTENCE* (. <>))

Seule     (CLAUSE* (ADVP-PERI (CORE_ADV (NUC_ADV (ADV <>)))))
comptait  (SENTENCE (CLAUSE (CORE (NUC (V <>)) (NP ))))
la        (NP* (OP-DEF <>))
Police    (NUC_N* (N-PROP <>))
de        (P <>)
la        (NUC_N* (OP-DEF <>))
Pensée    (NP (CORE_N (NUC_N (P ) (N-PROP <>))))
.         (SENTENCE* (. <>))
```

Current Supertags in RRGparbank

Language	Sentences
English	5 441
French	612
German	630
Russian	2 117

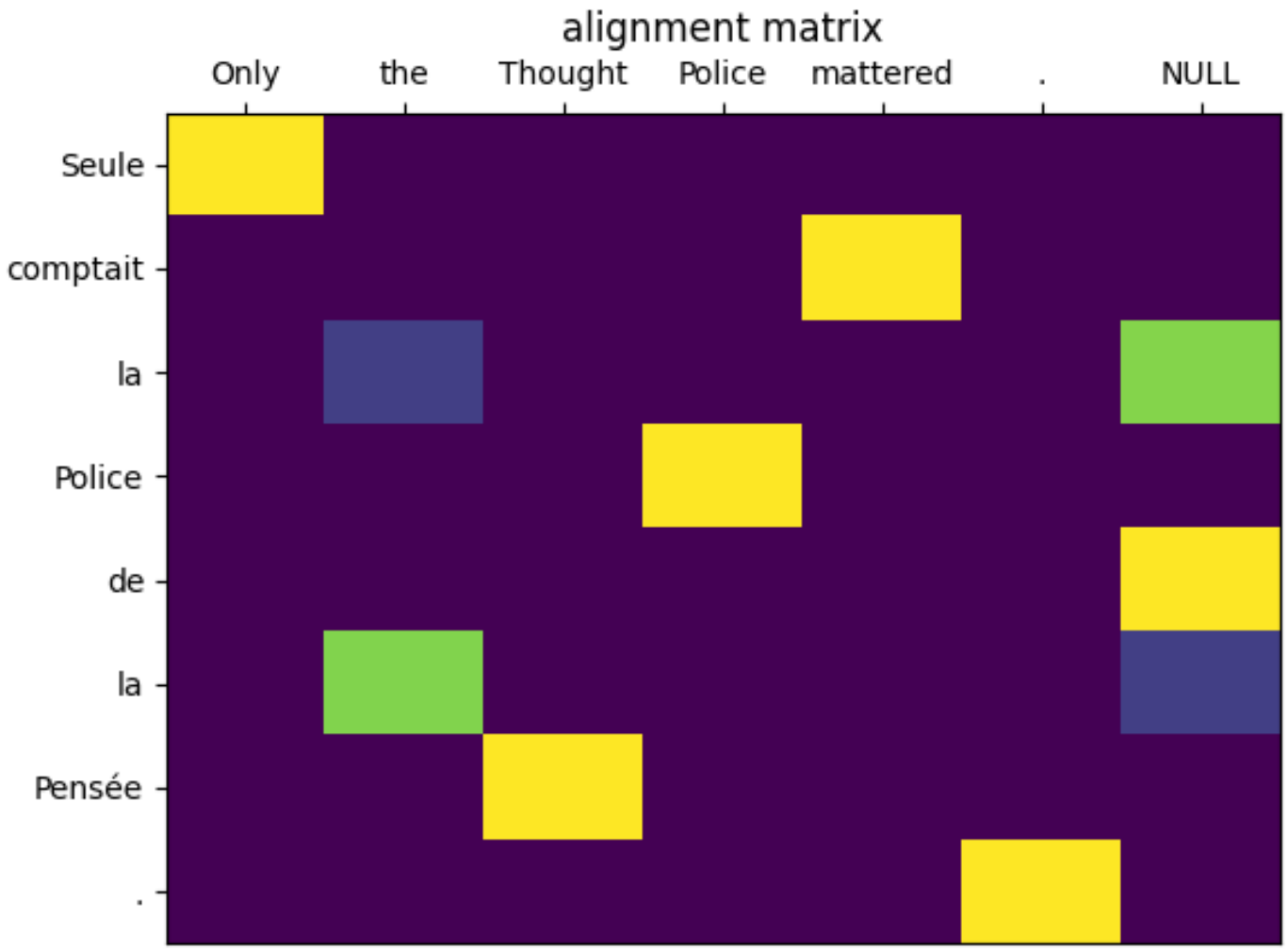
Table: Number of gold trees (automatically decomposed into supertags) in RRGparbank

A case for cross-lingual learning?

Cross-lingual Learning for French RRG Parsing: Initial Idea

- monolingual parser for French: try to predict supertags, dependencies from French word embeddings only
- cross-lingual parser for French: also look at (embeddings of) (predicted) English supertags for aligned words

Word Alignments (IBM Model 1)



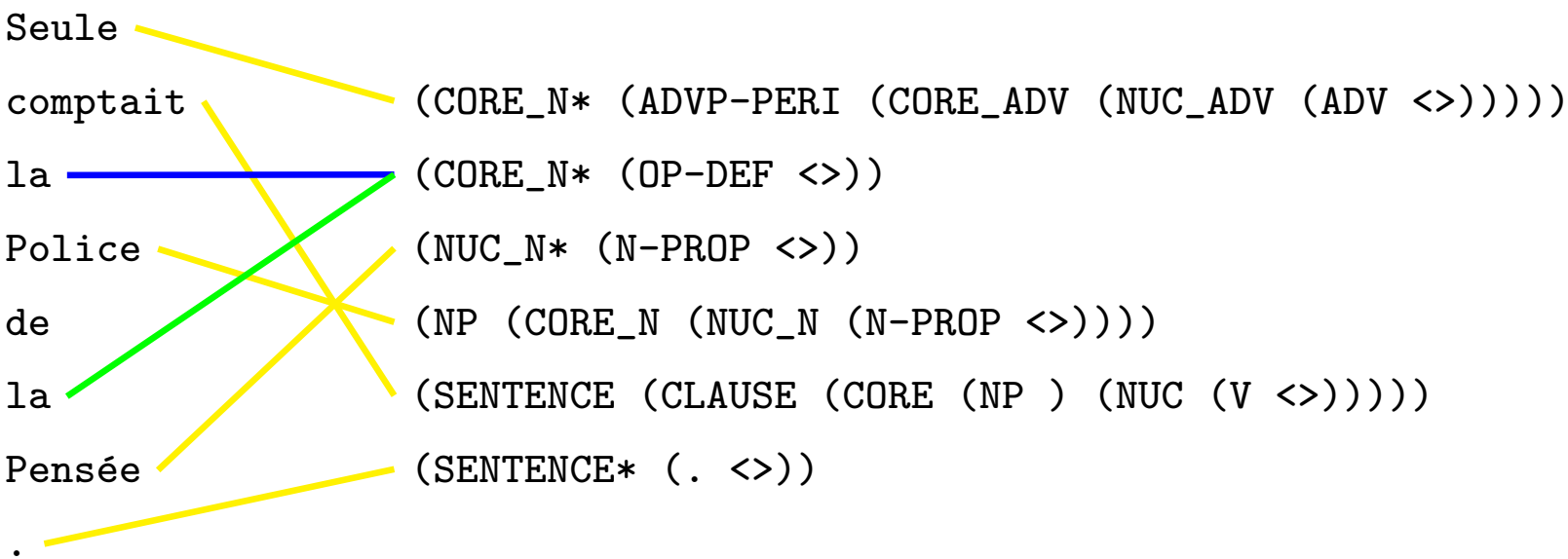
Word Alignments (Formally)

A word alignment function $a(j, i)$ which provides the probability of aligning the j-th target word with the i-th source word. It should satisfy the following equation:

$$\forall_{j \in 1..M} \left(\sum\nolimits_{i \in 1..N} a(j, i) \right) = 1, \tag{1}$$

where N and M is the size of the source and target sentence, respectively.

Cross-lingual Supervision with Supertags



Parsing architecture

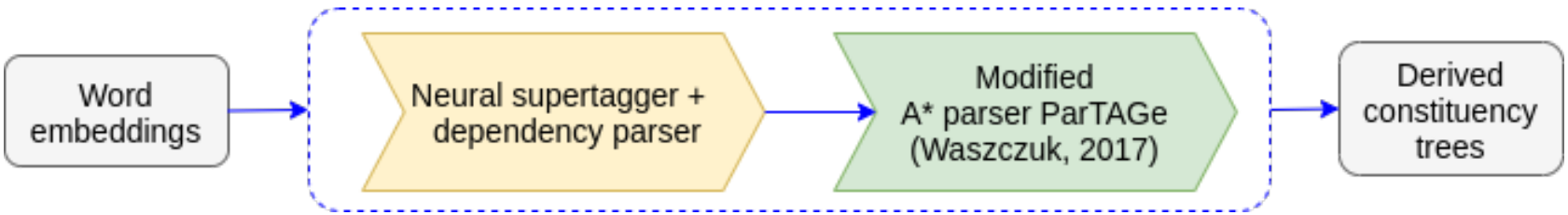


Figure: TWG parsing architecture

Mono-lingual Supertagging (Formally)

Input

Contextualized word embeddings $(h_i)_{i=1}^{n}$

Task-specific adaptation

$$(h_1^{(sup)}, \ldots, h_n^{(sup)}) = \texttt{BiLSTM}_s(h_1, \ldots, h_n) \tag{2}$$

$$(h_1^{(dep)}, \ldots, h_n^{(dep)}) = \texttt{BiLSTM}_d(h_1, \ldots, h_n) \tag{3}$$

Scoring (supertags and POS tags)

$$\Pr(sup(i)) = \texttt{softmax}(\texttt{Linear}_s(h_i^{(sup)})) \tag{4}$$

$$\Pr(pos(i)) = \texttt{softmax}(\texttt{Linear}_p(h_i^{(sup)})) \tag{5}$$

Cross-lingual Supervision with Supertags (Formally)

Projection

$$t_i = \sum_j a(j,i) t_j^{(src)} \tag{6}$$

where $t_i^{(sup)}$ is the source supertag embedding take from the last BiLSTM layer before scoring of the source (EN) supertagger.[1]

Adaptation

$$(t'_1, \ldots, t'_M) = \texttt{BiLSTM}(t_1, \ldots, t_M) \tag{7}$$

Scoring

$$\Pr_{prj}(sup(i)) = \texttt{softmax}(\texttt{Linear}_s(t'_i)) \tag{8}$$

[1] Dependency-related embeddings are projected in a similar way.

Initial Results

	POS	UAS	STag
Baseline EN $\rightarrow$ FR	70.50	63.16	53.96
Cross-lingual EN $\rightarrow$ FR	77.49	70.18	59.84
Monolingual FR	86.47	80.88	68.30
Cross-lingual w/o ParTAGe	77.55	65.17	61.52
Cross-lingual w/o Adaptation	70.39	58.89	56.68
Monolingual w/o ParTAGe	88.03	76.06	67.28

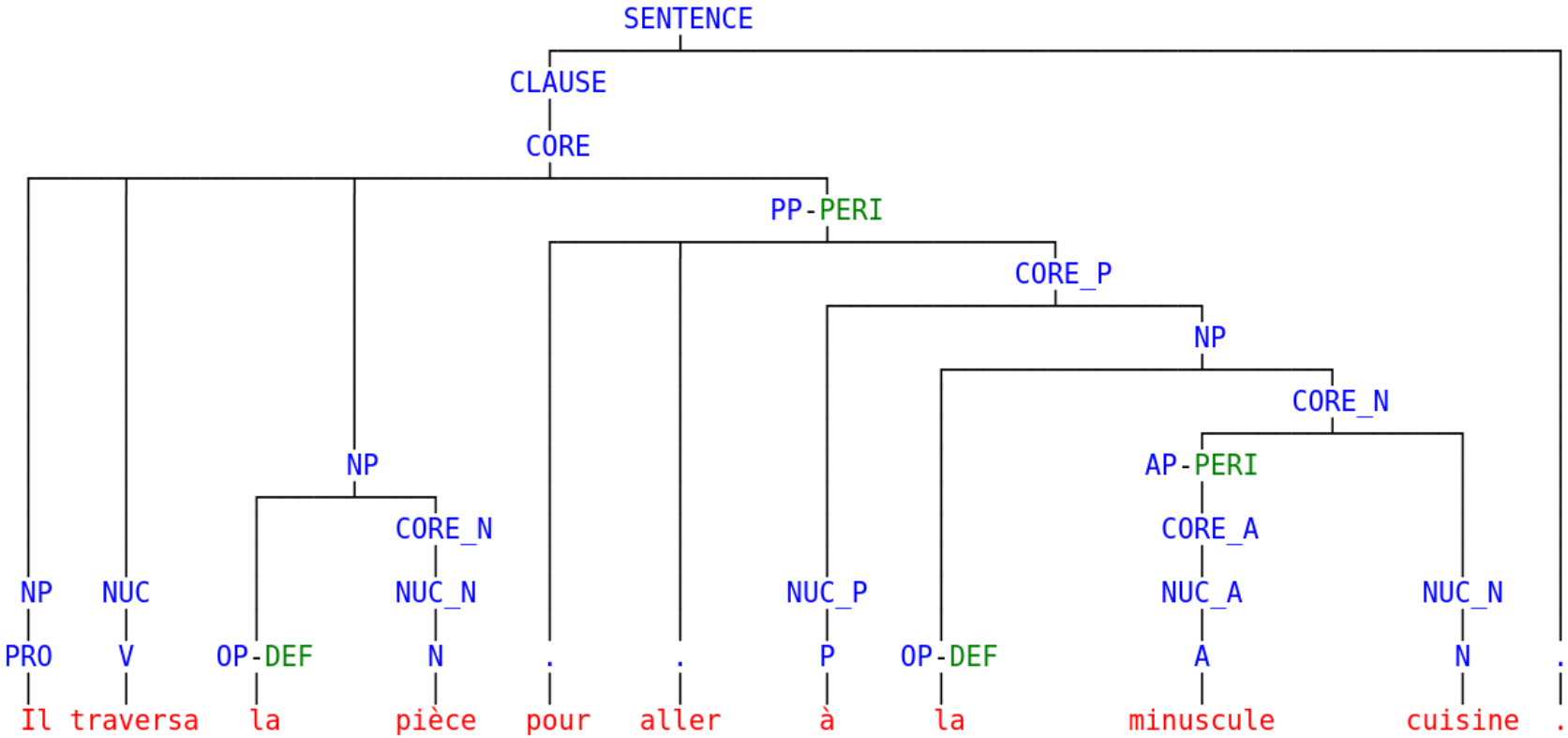
Table: Neural parsing results (POS tagging accuracy, unlabeled attachment score, and supertagging accuracy)

Error Analysis

Freq.	Gold/Pred.
4	(NP (CORE_N (NUC_N (N <>)) (PP)))
	(NP (CORE_N (NUC_N (N <>))))
3	(NP (CORE_N (NUC_N (N <>))))
	(NP (CORE_N (NUC_N (N-PROP <>))))
3	(PP (CORE_P (NUC_P (P <>)) (NP)))
	(CORE_N* (PP-PERI (CORE_P (NUC_P (P <>)) (NP))))
3	(CORE_N* (AP-PERI (CORE_A (NUC_A (V-PART <>)))))
	(CORE_N* (AP-PERI (CORE_A (NUC_A (A <>)))))
3	(NP (PRO-CLT <>))
	(PRO-CLT <>)
3	(PP (P-DEF <>) (NP))
	(NP* (OP-DEF <>))
3	(NP (PRO-CLT <>))
	(CORE* (OP-NEG <>))

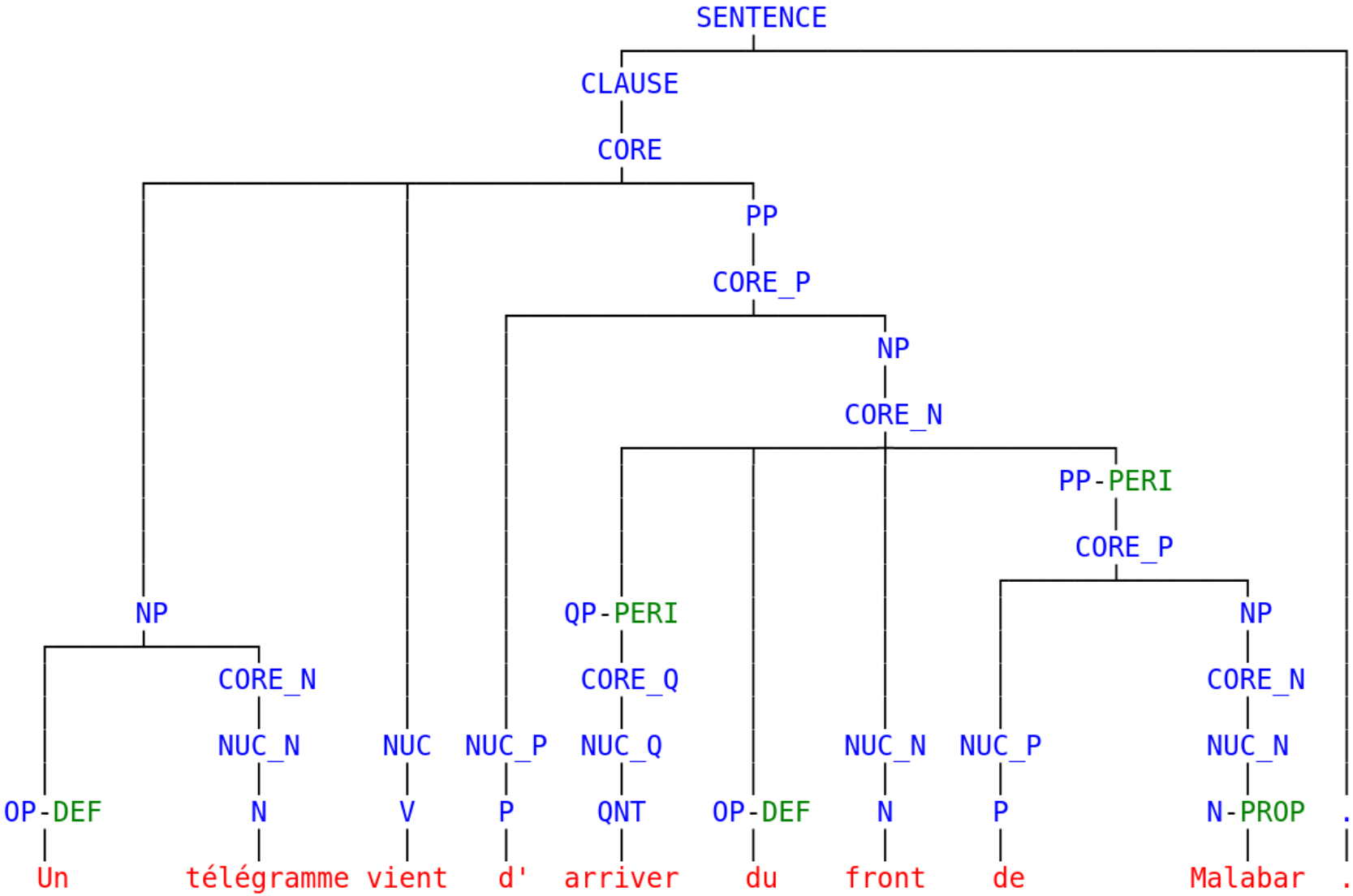
Table: Most frequently confused supertags (fr, dev)

Error Analysis: Example 1

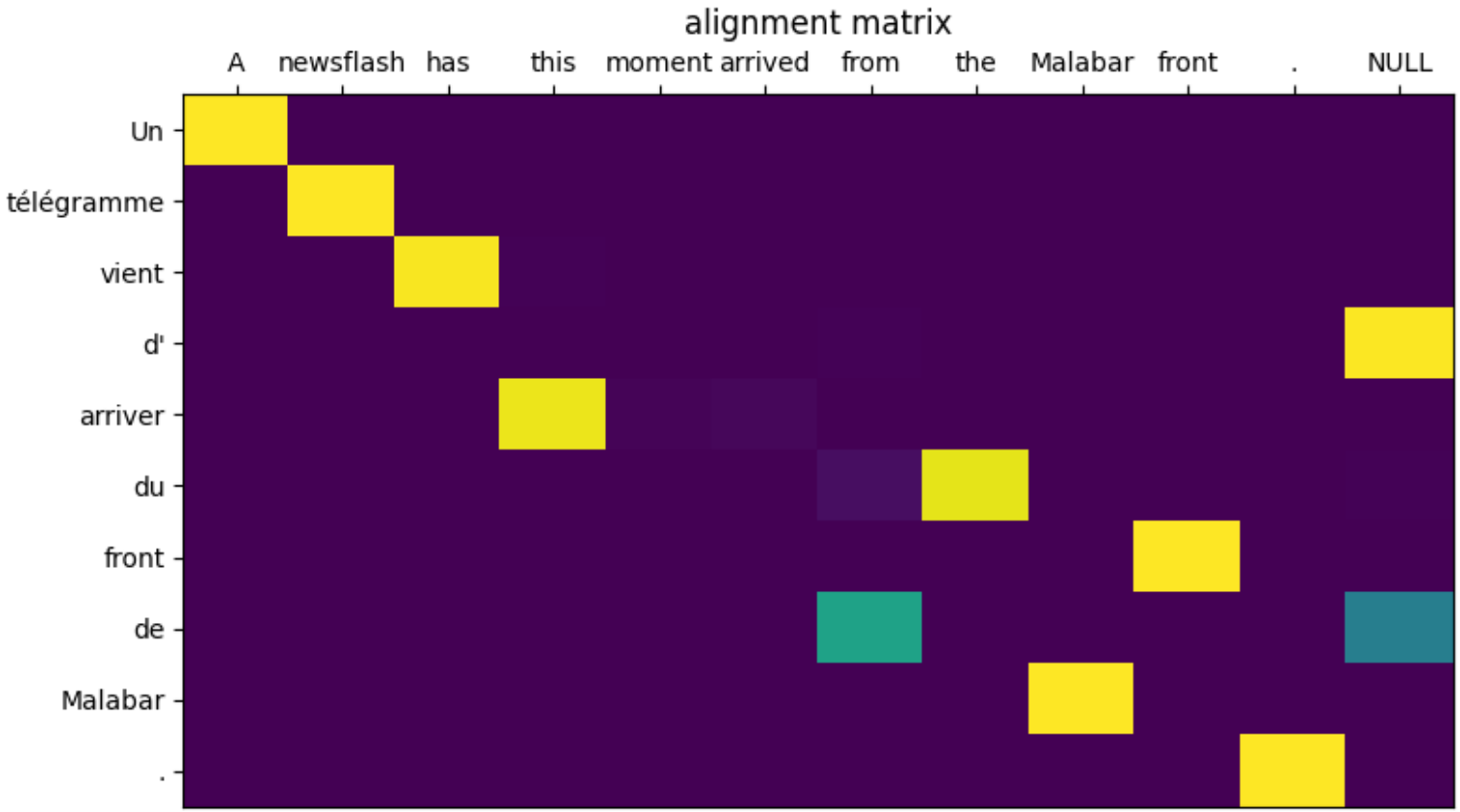


Error Analysis: Example 1

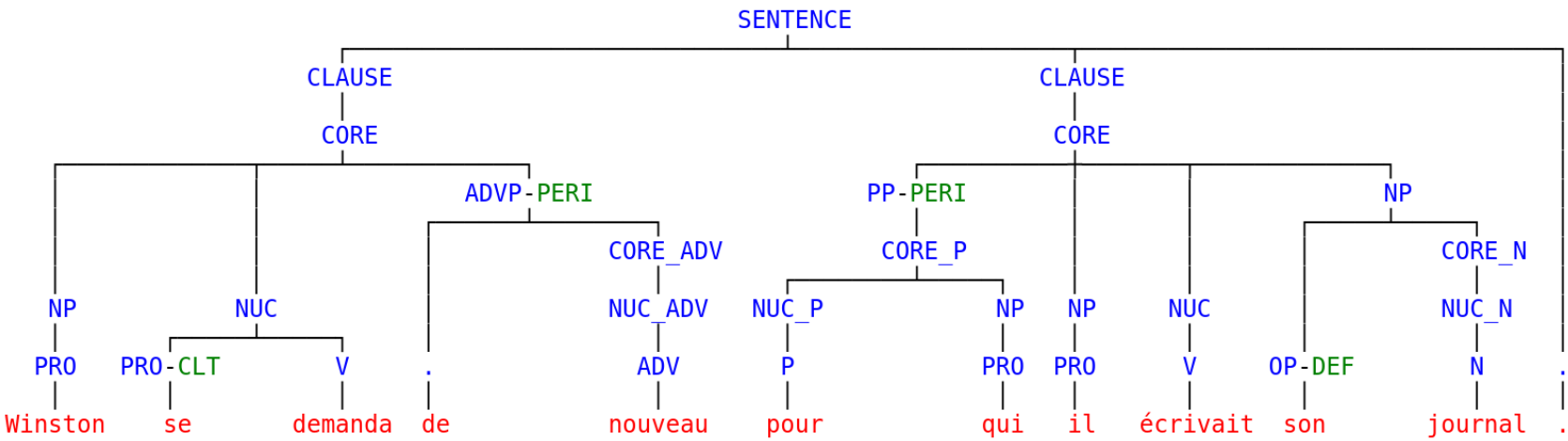
Error Analysis: Example 2



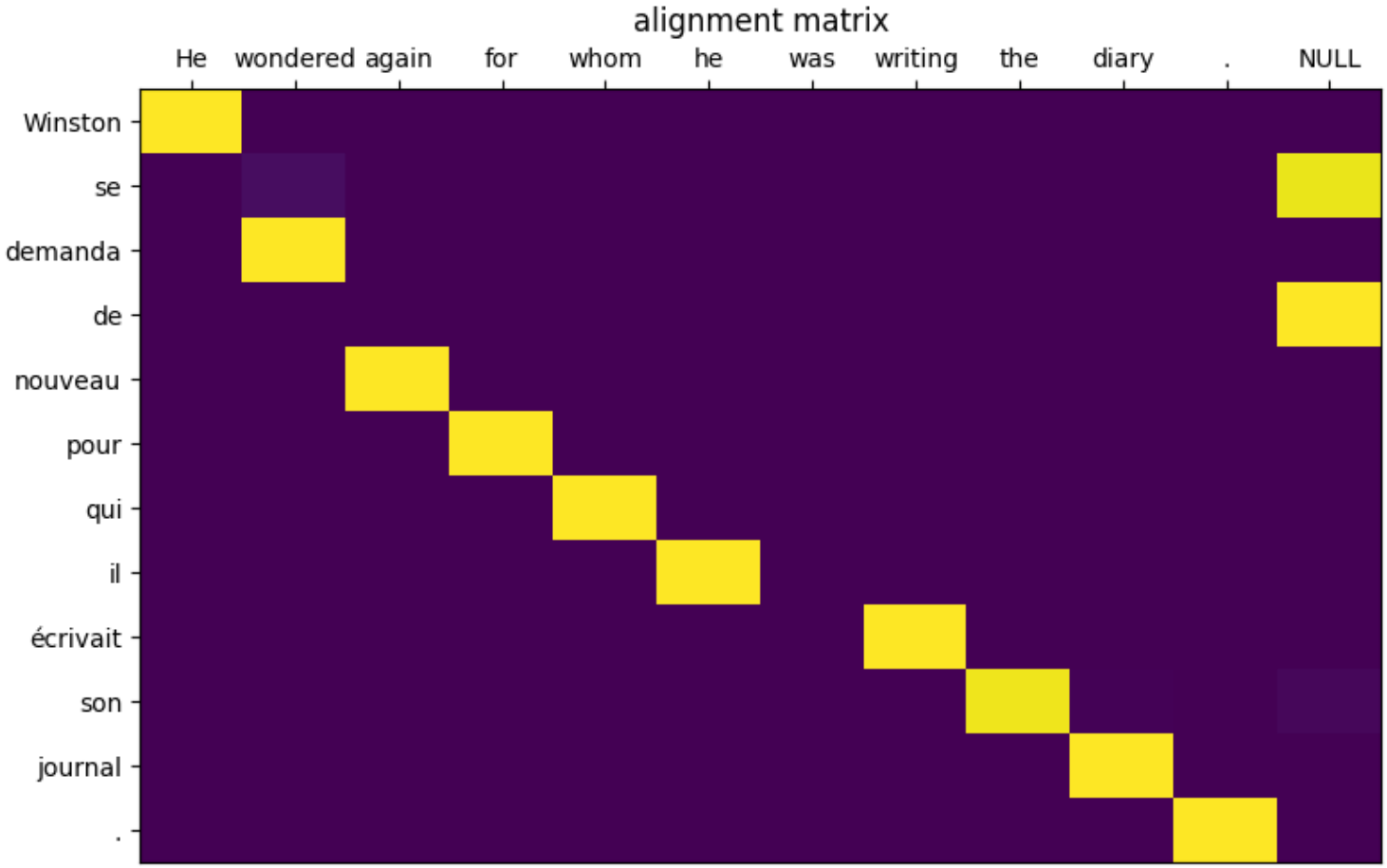
Error Analysis: Example 2 (Alignment)



Error Analysis: Example 3



Error Analysis: Example 3 (Alignment)



Error Analysis: Example 3 (Gold Supertags)

```
1       He       2:1.0   (NP (PRO <>)):1.0
2       wondered         0:1.0   (SENTENCE (CLAUSE (CORE (NP ) (NUC (V <>))) (CLAUSE ))):1.0
3       again    2:1.0   (CORE* (ADVP-PERI (CORE_ADV (NUC_ADV (ADV <>))))):1.0
4       for      8:1.0   (PP-WH (CORE_P (NUC_P (P <>)) (NP-WH ))):1.0
5       whom     4:1.0   (NP-WH (PRO-WH <>)):1.0
6       he       8:1.0   (NP (PRO <>)):1.0
7       was      8:1.0   (CORE* (OP-TNS <>)):1.0
8       writing  2:1.0   (CLAUSE (PrCS (PP-WH )) (CORE (NP ) (NUC (V <>)) (NP ))):1.0
9       the      10:1.0  (NP* (OP-DEF <>)):1.0
10      diary    8:1.0   (NP (CORE_N (NUC_N (N <>)))):1.0
11      .        2:1.0   (SENTENCE* (. <>)):1.0

1       Winston  3:1.0   (NP (PRO <>)):1.0
2       se       3:1.0   (PRO <>):1.0
3       demanda  0:1.0   (SENTENCE (CLAUSE (CORE (NP ) (NUC (PRO ) (V <>))) (CLAUSE ))):1.0
4       de       5:1.0   (P <>):1.0
5       nouveau  3:1.0   (CORE* (ADVP-PERI (CORE_ADV (NUC_ADV (P ) (A <>))))):1.0
6       pour     9:1.0   (PP-WH (CORE_P (NUC_P (P <>)) (NP-WH ))):1.0
7       qui      6:1.0   (NP-WH (PRO-WH <>)):1.0
8       il       9:1.0   (NP (PRO <>)):1.0
9       écrivait         3:1.0   (CLAUSE (PrCS (PP-WH )) (CORE (NP ) (NUC (V <>)) (NP ))):1.0
10      son      11:1.0  (NP* (OP-DEF <>)):1.0
11      journal  9:1.0   (NP (CORE_N (NUC_N (N <>)))):1.0
12      .        3:1.0   (SENTENCE* (. <>)):1.0
```

Error Analysis: Example 3 (Predicted Supertags)

```
1       Winston 3:1.0   (NP (PRO <>)):1.0
2       se      3:1.0   (PRO-CLT <>):1.0
3       demanda 0:1.0   (SENTENCE (CLAUSE (CORE (NP) (NUC (PRO-CLT) (V <>))))):1.0
4       de      5:1.0   (ADVP-PERI* (. <>)):1.0
5       nouveau 3:1.0   (CORE* (ADVP-PERI (CORE_ADV (NUC_ADV (ADV <>))))):1.0
6       pour    9:1.0   (CORE* (PP-PERI (CORE_P (NUC_P (P <>)) (NP)))):1.0
7       qui     6:1.0   (NP (PRO <>)):1.0
8       il      9:1.0   (NP (PRO <>)):1.0
9       écrivait        3:1.0   (SENTENCE* (CLAUSE (CORE (NP) (NUC (V <>)) (NP)))):1.0
10      son     11:1.0  (NP* (OP-DEF <>)):1.0
11      journal 9:1.0   (NP (CORE_N (NUC_N (N <>)))):1.0
12      .       3:1.0   (SENTENCE* (. <>)):1.0
```