# Parse Correction with Specialized Models for Difficult Attachment Types


Enrique Henestroza Anguiano & Marie Candito

Alpage (Université Paris Diderot / INRIA)



## Summary

• Dependencies in an input parse tree are revised by selecting, for a given dependent, the best governor from a small set of candidates. (Based on work of [3])

• A discriminative linear ranking model selects the best governor using a rich feature set that encodes syntactic structure.

• Both a generic model and specialized models tailored to coordination and pp-attachment are tested. Combining a generic model with specialized models successfully improves parse quality for French.


## Corrective modeling

• Scoring function $S(c, d, T)$ is trained using an online passive-aggressive algorithm, with a ranking approach that maximizes the margin between the gold governor and the highest-scoring incorrect candidate.

• The *generic* model corrects any dependent $d$, and uses combinations of POS, lemma, distance, tree path, and other features over $c$ and $d$ in tree $T$.

• The two *specialized* models are focused on correcting coordinating conjunction and preposition dependents (using extra phenomenon-specific features).

## Experiments and results

• Each parser was used to auto-parse (jackknife) the French Dependency Treebank [2] training set. The auto-parse and gold parse of each sentence were paired to create training examples for parse correction.

• Oracle results (on dev set) show that parse correction has the potential to greatly improve UAS scores, both overall and for difficult attachment types.

• Final results (on test set) show that parse correction significantly improves coordination UAS across all parsers, while modestly improving pp-attachment and overall UAS for some parsers.

## Statistical dependency parsing

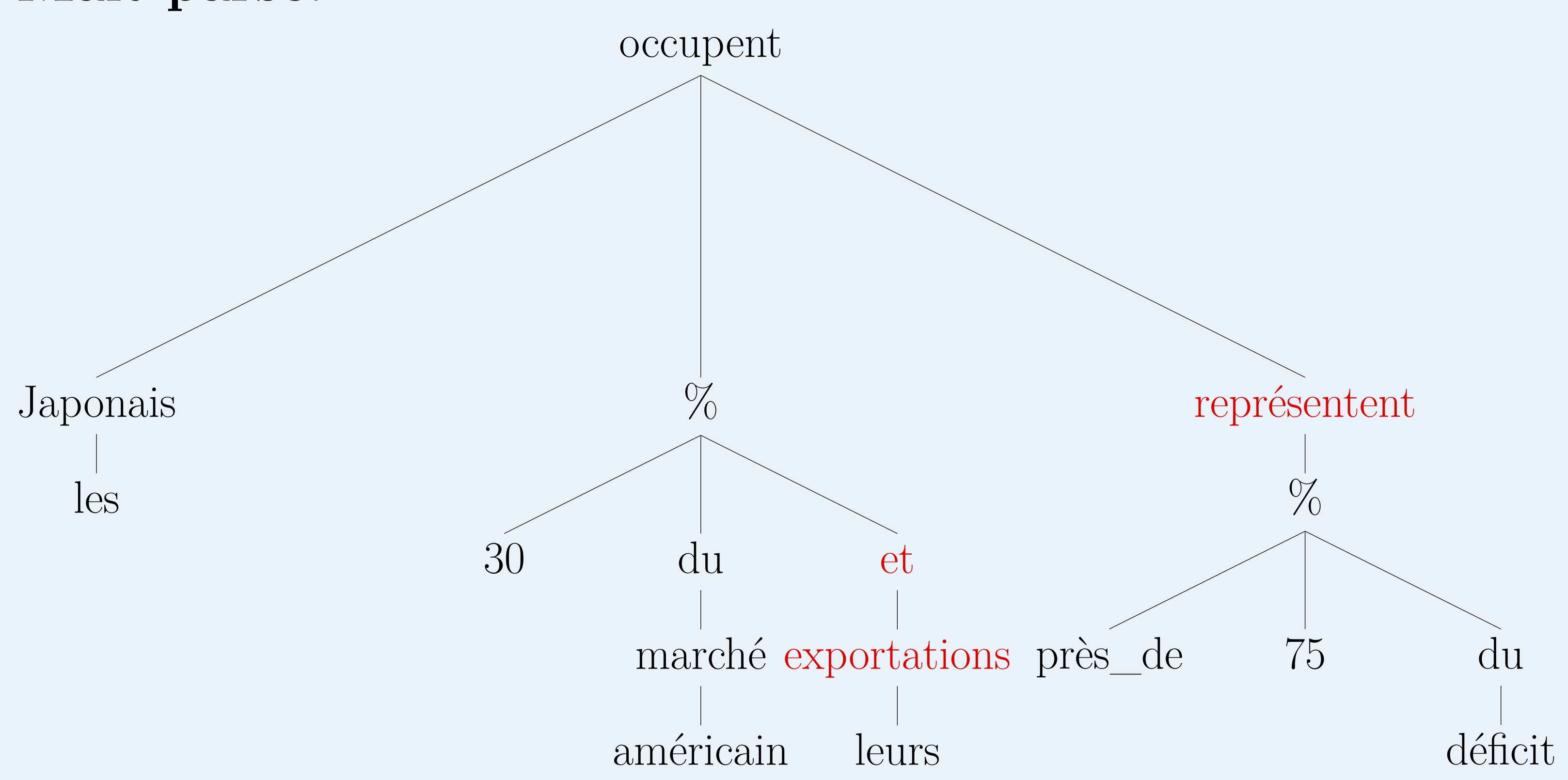
**Parser baselines (UAS %):**

| | Coords | Preps | Overall |
|---|---|---|---|
| Berkeley | 68.3 | 83.8 | 90.73 |
| Bohnet | 70.5 | 86.1 | 91.78 |
| Malt | 59.8 | 83.2 | 89.78 |
| MST | 60.5 | 85.9 | 91.04 |

**Example sentence:**

*... les Japonais occupent 30% du marché américain et leurs exportations représentent près de 75% du déficit ...*
*(... the Japanese occupy 30% of the American market and their exports represent close to 75% of the deficit ...)*

**Malt parse:**




## Parse correction

**Correction algorithm:**

INPUT: Predicted parse tree $T$

LOOP: For each chosen dependent $d \in D$

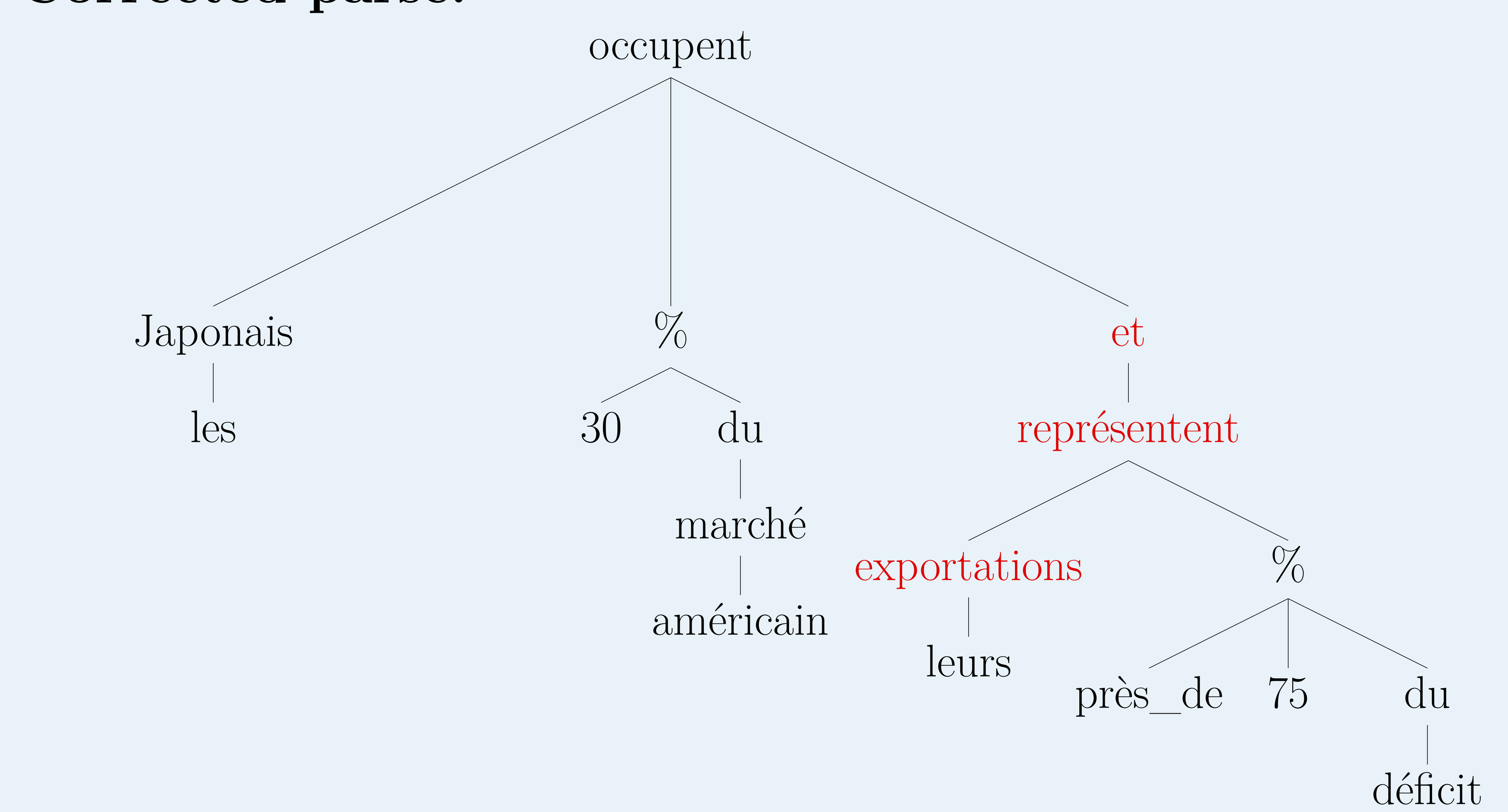
- Identify candidates $C_d$ from $T$
- Predict $\hat{c} = \underset{c \in C_d}{\text{argmax}}\ S(c, d, T)$
- Update $T\{gov(d) \leftarrow \hat{c}\}$

OUTPUT: Corrected version of parse tree $T$

**Sequential corrections:**

1. $gov$(et) ← occupent
2. $gov$(exportations) ← représentent
3. $gov$(représentent) ← et

**Corrected parse:**




## Oracle and final evaluations

**Oracle evaluation (UAS %):**

| | | Neighborhood Size ($m$) | | | |
|---|---|---|---|---|---|
| | | Base | 2 | 3 | 4 |
| Berkeley | Coords | 67.2 | 76.5 | 82.8 | 84.8 |
| | Preps | 82.9 | 88.5 | 92.2 | 93.2 |
| | Overall | 90.1 | 94.0 | 96.0 | 96.5 |
| Bohnet | Coords | 70.1 | 80.6 | 85.6 | 87.7 |
| | Preps | 85.4 | 89.4 | 93.4 | 94.5 |
| | Overall | 91.2 | 94.4 | 96.1 | 96.6 |
| Malt | Coords | 60.9 | 72.2 | 78.2 | 80.5 |
| | Preps | 82.6 | 88.1 | 92.6 | 93.7 |
| | Overall | 89.3 | 93.2 | 95.1 | 95.8 |
| MST | Coords | 63.6 | 73.7 | 80.7 | 84.4 |
| | Preps | 84.7 | 89.4 | 93.4 | 94.4 |
| | Overall | 90.2 | 93.7 | 95.6 | 96.2 |

**Final evaluation (UAS %):**

| | Corrective Configuration | UAS (%) | | |
|---|---|---|---|---|
| | | Coords | Preps | Overall |
| Berkeley | Baseline | 68.3 | 83.8 | 90.73 |
| | Generic | 69.4 | 84.9* | 91.13* |
| | Specialized | 71.5* | 85.1* | 91.23* |
| Bohnet | Baseline | 70.5 | 86.1 | 91.78 |
| | Generic | 71.2 | 86.4 | 91.88 |
| | Specialized | 72.7* | 86.2 | 91.88 |
| Malt | Baseline | 59.8 | 83.2 | 89.78 |
| | Generic | 63.2* | 84.5* | 90.39* |
| | Specialized | 64.0* | 85.0* | 90.47* |
| MST | Baseline | 60.5 | 85.9 | 91.04 |
| | Generic | 64.2* | 86.2 | 91.25* |
| | Specialized | 68.0* | 86.2 | 91.36* |

## Baseline parsers

• Four parsers tested: one constituent-based parser (Berkeley [6]) with conversion to dependency structure, two global optimisation parsers (Bohnet [1] and MST [4]), and one transition-based parser (Malt [5]).

• Attachments are generally accurate, with around 90% unlabeled attachment score overall.

• Errors are more common on coordination (60-70% UAS) and pp-attachment (83-86% UAS). At left, a coordination error: (% → et → exportations) instead of (occupent → et → représentent).

## Running times

• French Dependency Treebank test set containing 1,235 sentences, in decreasing speed:

- Malt: 45 seconds, $O(n)$
- Parse Correction (rough): 200 seconds, $O(n)$
- Berkeley: 600 seconds, $O(n^3)$
- Bohnet: 800 seconds, $O(n^3)$
- MST: 1,000 seconds, $O(n^3)$

• **An attractive system: Malt+Correction?** Correction improves Malt most, and system is $O(n)$.

## Selected references


[1] B. Bohnet. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of COLING '10*.

[2] M. Candito, B. Crabbé, and P. Denis. 2010. Statistical french dependency parsing: Treebank conversion and first results. In *Proceedings of LREC '10*.

[3] K. Hall and V. Novák. 2005. Corrective modeling for non-projective dependency parsing. In *Proceedings of IWPT '05*.

[4] R. McDonald, K. Crammer, and F. Pereira. 2005. Online large-margin training of dependency parsers. In *Proceedings of ACL '05*.

[5] J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*.

[6] S. Petrov, L. Barrett, R. Thibaux, and D. Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of COLING-ACL '06*.