



1. About

Compare how two dependency parsers, one graph-based, the other transition-based, perform on Old French, facing some typical problems of medieval texts.

GPM: Graph-based *mate tools* parser (Bohnet, 2010) used with *Marmot* Tagger (Müller et al., 2013)

JTP: Joint transition-based parser Bohnet et al. (2013)

2. Objective

- ▶ Test both parsers on the *Syntactic Reference Corpus of Medieval French (SRCMF)* (Prévost & Stein, 2013).
- ▶ Questions: Which parser performs better, especially regarding free word order? What kind of effects do syntactic properties of the language have on both parsers?

3. Medieval texts: a challenge for NLP

Properties of Old French (OF):

- Variation:** The OF period stretches over more than 300 years. Considerable orthographical variation: type-token ratio 0.048 (= twice as high as in Modern French text).
- Syntax:** Relatively free word order in OF, similar to e.g. German or some other Germanic languages. "Null-subject": sentences need not have an overt subject (similar to e.g. Modern Italian or Spanish).
- Inflection:** OF is closer to e.g. Modern German than to Modern English. Verbs are marked for person, number, and tense/mood; nouns, pronouns, and adjectives are marked for number and two cases.

Previous results with the graph-based *mate tools* parser (complete pipeline including *mate tools* lemmatisation and tagging, Stein 2014): 82.62% labelled attachment score (LAS).

4. The training corpus

- ▶ 12 texts from the SRCMF, from 1000 to 1300, prose and verse
- ▶ Verified manual dependency and part of speech annotation
- ▶ Grammar model (see <http://srcmf.org> for details):
 - ▷ no coordination of main clauses
 - ▷ lexical heads dominate functional categories

5. Training of the parsers

- ▶ CoNLL export: 242 946 tokens (23 818 types), no punctuation
- ▶ Lemmatised with *TreeTagger* and OF parameters (unverified)
- ▶ Split training:evaluation 90:10, 10-fold cross-validated
 - ▷ GPM: improved pos/feature tagging using *Marmot*
 - ▷ JTP: joint model for tagging and parsing

6. Results per category and global scores

- ▶ For individual dependency relations (table not shown here):
 - ▷ Parsers diverge more often on the arguments of the verb (subject, direct/indirect object, etc.) than on other categories.
- ▶ Global scores:
 - ▷ slightly better LAS for JTP: +0.78
 - ▷ considerably better exact (per-sentence) match for JTP: +5.76

	GPM	JTP	diff.
part of speech (pos) accuracy	95.49%	95.78%	+0.29
feature accuracy	93.72%	93.21%	-0.51
unlabelled attachment score (UAS)	91.54%	91.75%	+0.21
labelled attachment score (LAS)	85.18%	85.96%	+0.78
label accuracy	88.51%	89.06%	+0.55
exact match/LAS	41.83%	47.59%	+5.76

Table: Global scores

7. Results per sentence

For whole sentences	GPM	JTP
① LA=true	965 (41.8%)	1098 (47.6%)
② LA=true for both parsers	835 (36.2%)	835 (36.2%)
③ LA=true for this parser	130 (5.6%)	263 (11.4%)
④ tokens per sent. if LA=true	6.6	6.9
⑤ pos=true	1611 (69.8%)	1646 (71.4%)
⑥ pos=false & LA=true	97 (4.2%)	116 (5.0%)

Table: Per-sentence results (LA=labelled attachment, pos=part of speech)

- ▶ Per-sentence scores
 - ▷ Line ①: LAS of exact sentence matches
 - ▷ Line ②: Most matches are predicted by both parsers.
 - ▷ Line ③: JTP has more exclusive exact matches.
 - ▷ Line ④: Length of JTP's exact matches is +0.3 tokens higher.
 - ▷ Line ⑤: Exact tagging: JTP outscores GPM by 1.5.
 - ▷ Line ⑥: JTP slightly better for sentences with tagging errors.

8. Example: Left dislocation

More deeply embedded in the SRCMF model (than e.g. in the Danish Treebank).

GPM:

JTP:

(gloss) of+the helmets bright the fire from+them shines

9. Example: overt subjects

False predictions, since subjects need not be overtly expressed in Old French.

GPM and JTP:

If something on these constructions climbed the door from up went down

GPM:

Three divisions has the emperor Charles

10. Conclusions

- ▶ Syntactic and morphological properties of the language matter for parser choice.
- ▶ JTP outperforms GPM in several respects.
- ▶ JTP predicts verb valency more reliably.
- ▶ Otherwise no consistent differences that have a purely linguistic explanation.

Resources (Old French lemmatiser, tagger, parser)

- ▶ <http://www.uni-stuttgart.de/lingrom/stein>
- ▶ <http://srcmf.org> (SRCMF corpus)
- ▶ <https://weblicht.sfs.uni-tuebingen.de>

[Download this poster here →](#)

References

Bohnet, Bernd. 2010. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 89–97. Beijing, China. <http://www.aclweb.org/anthology/C10-1011>.

Bohnet, Bernd, Joakim Nivre, Igor Boguslavsky, Richard Farkas, Filip Ginter & Jan Hajic. 2013. Joint Morphological and Syntactic Analysis for Richly Inflected Languages. In *TACL 1*, 415–428.

Müller, Thomas, Helmut Schmid & Hinrich Schütze. 2013. Efficient Higher-Order CRFs for Morphological Tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 322–332. Seattle, Washington, USA: ACL. <http://www.aclweb.org/anthology/D13-1032>.

Prévost, Sophie & Achim Stein. 2013. Syntactic Reference Corpus of Medieval French (SRCMF). <http://srcmf.org>.

Stein, Achim. 2014. Parsing Heterogeneous Corpora with a Rich Dependency Grammar. In Nicoletta Calzolari et al. (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 26–31.5.2014, Reykjavik, Iceland: ELRA. <http://www.lrec-conf.org/proceedings/lrec2014/pdf/239.Paper.pdf>.