Accurate Deep Syntactic Parsing of Graphs: The Case of French

Corentin Ribeyre[◦] **Eric Villemonte de la Clergerie**[⋆] **Djamé Seddah**[⋆ ⋄]
[◦]Université de Genève [⋆]Alpage, INRIA [⋄] Université Paris-Sorbonne
corentin.ribeyre@unige.ch eric.de_la_clergerie@inria.fr djame.seddah@paris-sorbonne.fr

Abstract

Parsing predicate-argument structures in a deep syntax framework requires graphs to be predicted. Argument structures represent a higher level of abstraction than the syntactic ones and are thus more difficult to predict even for highly accurate parsing models on surfacic syntax. In this paper we investigate deep syntax parsing, using a French data set (Ribeyre et al., 2014a). We demonstrate that the use of topologically different types of syntactic features, such as dependencies, tree fragments, spines or syntactic paths, brings a much needed context to the parser. Our higher-order parsing model, gaining thus up to 4 points, establishes the state of the art for parsing French deep syntactic structures.

Keywords: deep syntax, graph parsing, syntactic features

1. Introduction

The availability of many manually annotated syntactic corpora allows data-driven syntactic parsing to perform efficiently on English (routinely reaching 92% of labeled attachment accuracy) as well as on languages with a richer level of morphology (85% to 88%) (*e.g.* French, German, Arabic). However, despite this impressive level of performance, it has now become clear that such *surface* syntactic parses are often insufficient for semantically-oriented tasks such as question-answering systems (Berant et al., 2013). This is because many predicate-argument dependencies, such as those arising for instance in control-verb constructions, it-cleft constructions, participle clauses and so on, are lacking from the annotation schemes underlying most data set used by surface syntactic parsers.

Representing such constructs in a dependency scheme often leads to graph representations, which are a real challenge to predict, as shown, for example, by the performance obtained in Abstract Meaning Representation parsing (Banarescu et al., 2013; Flanigan et al., 2014; Artzi et al., 2015) or in graph-based semantic dependency parsing (Oepen et al., 2014). Even though some recent works have made good progresses in parsing full-fledged semantic structures (Beschke et al., 2014; Berant and Liang, 2014), they mostly focus on English.[1]

We propose to dig into that direction by relying on the recent release of a deep syntactic graphbank for French, the DEEPFTB, (Ribeyre et al., 2014a), whose deep syntactic graphs were automatically annotated on top of the dependency version of the French Treebank (Abeillé and Barrier, 2004; Candito et al., 2010). In order to draw meaningful comparisons with recent results for English, we use the same types of features as Ribeyre et al. (2015), which were extracted from the DM corpus (Oepen et al., 2014), an English predicate-argument structure corpus sharing similarities with the DEEPFTB.

Despite the differences of language and annotation scheme, we observe that the same combination of different topological syntactic information leads to the best models for both French and English.

[1]An exception is the work of Ballesteros et al. (2014) on deep syntacting parsing for Spanish, but their work is restricted to tree structure parsing.

2. Deep Syntactic French Annotation Scheme and Corpus

	DM CORPUS TRAIN	DM CORPUS DEV	DEEPFTB CORPUS TRAIN	DEEPFTB CORPUS DEV
# SENTENCES	32,389	1,614	14,759	1,235
# TOKENS	742,736	36,810	457,872	40,055
% VOID TOKENS	21.63	21.58	11.97	12.19
% VOID TOKENS (no punct.)	NA	NA	35.34	35.57
# PLANAR GRAPHS	18,855	972	8,292	664
# NON PLANAR	13,534	642	6,467	571
# DAGS	32,389	1,614	3,911	283
# EDGES	559,975	27,779	424,813	37,110
% CROSSING EDGES	4.24	4.05	3.70	3.87
LABEL SET	52	36	27	24

Table 1: DM and DEEPFTB Properties.

We exploit a French corpus annotated following the deep syntactic scheme presented in (Perrier et al., 2014) and already instantiated on the Deep Sequoia corpus (Candito et al., 2014). This scheme aims at abstracting away from some syntactic variations by expliciting which expressions fill the *canonical* subcategorization frames of verbs and adjectives. Canonical grammatical functions are roughly those that would be assigned to an argument, if its predicate were in the most unmarked construction. This results mainly in (i) normalized grammatical functions in the case of syntactic alternations (e.g. the subject of a passive verb is taken as the canonical object) (ii) added dependencies for the subject of non-finite verbs (in particular in control/raising constructions) (iii) added dependencies in the case of arguments shared by coordinated heads and (iv) inverted dependencies in the case of modifying verbs or adjectives (for instance, in (the French counterpart) of *'Children born after 2010 get free tickets'*, the participle *born* both modifies the noun *Children*, and has this noun as (deep) subject). Moreover, semantically empty functional words are marked as such, and "shunted off" (for instance in *'Anna a parlé à Paul' (Anna has talked to Paul)*, both the auxiliary and the preposition *à* are marked as empty, and *Paul* is directly linked to the verb).

As with the DM corpus (Oepen et al., 2014), DEEPFTB is comparable in the sense that the semantic arguments of verbs and adjectives are made explicit, but it leans a little less towards a semantic representation (hence the "deep syntactic" name). In particular it sticks to (canonical) syntactic labels `subj`, `obj`, ... instead of using numbered semantic labels `arg1`, `arg2`, ... Also, in the case of a predicate modifying one of its semantic argument (e.g. an attributive adjective), both the modifier dependency and the predicative dependency are kept in the deep graph: for instance for an attributive adjective like in *a perfect day*, *day* is taken as the subject of *perfect*, and *perfect* as a modifier of *day*. This choice was made in order to keep both the predicate-argument structures and the general information structure of a sentence. So for instance in Figure 1, while the copula is ignored in DM, it is kept in the French scheme.[2] It can be noted though that it causes a high number of cycles in the resulting graphs as seen in Table 1.

3. Graph Parsing

An increasing number of works have been proposed over the last few years to cope with graphs (Sagae and Tsujii, 2008; Flanigan et al., 2014; Martins and Almeida, 2014), whether acyclic or not. Because it has been shown to be one of the top performers on the SemEval 2014 shared task (Oepen et al., 2014), we reuse the higher-order arc-factored parser of (Martins and Almeida, 2014) (TSPARSER), which takes advantage of dual decomposition methods based on AD^3 (Martins et al., 2011). Another motivation for using this parser is that we want to assess whether using predicted syntactic features, which do provide additional context, is beneficial even when using a globally optimized parser.

4. Syntactic Features

Using syntactic features is widely known to help predicate-argument structure parsing by providing more context (Chen and Rambow, 2003; Moschitti et al., 2008). Following Farkas et al. (2011) and Ribeyre et al. (2015), we explore the impact of topologically different syntactic features extracted from the surfacic syntax and their respective combinations. Both our surfacic parsers use the French Treebank (Abeillé et al., 2003) in its (Seddah et al., 2013) instance, with predicted POS and morphological features. The constituency features come from the Berkeley Parser (Petrov et al., 2006) trained in a 10-fold jackkniffing setting. Respective parsers' performance scores are shown in Table 2.

	BKY	FRMG
Dev	80.19	83.41
Test	80.14	83.22

Table 2: Scores on surfacic syntax for the BKY (F_1) and FRMG (LAS) parsers.

TAG-based Dependencies (FRMG) As opposed to most previous works (Øvrelid et al., 2009), we use dependencies features extracted from a hand-written wide-coverage TAG-based[3] metagrammar (FRMG, (de La Clergerie, 2010)).[4]

Tree Fragments (BKY) These consist of fragments of syntactic constituency trees. They have been extracted using the same method as in (Carreras and Màrquez, 2005).

Spinal Elementary Trees (SPINES) They consist of the path of the maximal projection of a head in a constituency tree. They have been extracted using a *spine grammar* (Seddah, 2010) and the head percolation table of Dybro-Johansen (2004). The spines are assigned in a deterministic way.

Constituent Head Paths (PATHS) We use FRMG dependencies to extract the shortest path between a token and its lexical head and include the path length w (*i.e.* the number of traversed nodes) as a feature. The main idea is to use the phrase-based feature as well as the dependency feature to provide different kind of contexts and a generalisation over the functional label governing a token. These give a wider context to the parser whereas the spines are viewed as deterministic supertags bringing a vertical context. Table 2 presents an overview of the expected accuracy of our features via the scores of their respective source parsers. The feature set is shown in Figure 2.

5. Experiments & Discussion

Dev. set	LP	LR	LF	
BASELINE	83.04	78.80	80.86	
BKY	83.63	79.67	81.60	+0.74
SPINES	83.72	80.05	81.84	+0.98
PATHS	84.75	81.17	82.92	+2.06
FRMG	86.50	82.74	84.58	+3.72
FRMG+PATHS+BKY	86.11	83.68	84.88	+4.02
FRMG+PATHS+SPINES	86.15	83.71	84.91	+4.05

Table 3: Best results and gains (TSPARSER).

Experiments Table 3 displays the baseline scores, the scores for each type of features separately and our best models. As we can see, our baseline is weak especially in term of recall, leading us to believe that it is indeed difficult to recover the deep structure. Whereas the parser explores a large part of the search space, it seems to need more context to cope with complex linguistic structures. As expected, the use of each single feature increases the scores over the baseline, the improvement ranging from 0.74 using tree fragments (BKY) up to 3.72 points using the dependency features provided by the TAG parser. It is worth noticing that providing a wider context, namely using the PATHS features give an improvement that is closer to the best performing features (around 2 points), whereas the vertical context brought by the spines features does not give

[2]The authors wanted to differentiate graphs for *X is impossible* and *an impossible X*: if the copula is ignored, it only remains the same *impossible* $\longrightarrow$ *X* dependency.

[3]Tree Adjoining Grammars (Joshi and Schabes, 1997).

[4]This parser generates dependency trees after disambiguation and conversion from a shared derivation forest (Villemonte De La Clergerie, 2013b).

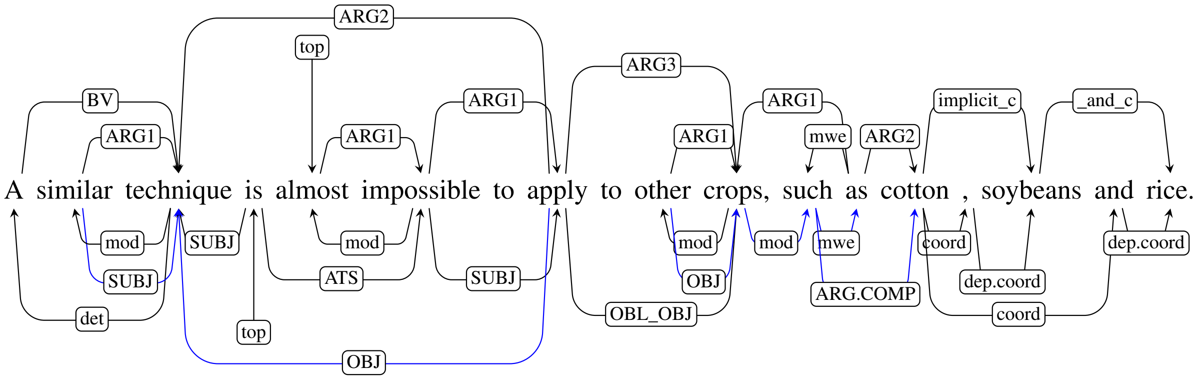


Figure 1: Comparison of DM and Deep Sequoia annotation schemes: Top: DM Annotation (Oepen et al., 2014) Bottom: Transposition to English of the Deep Sequoia annotation scheme.

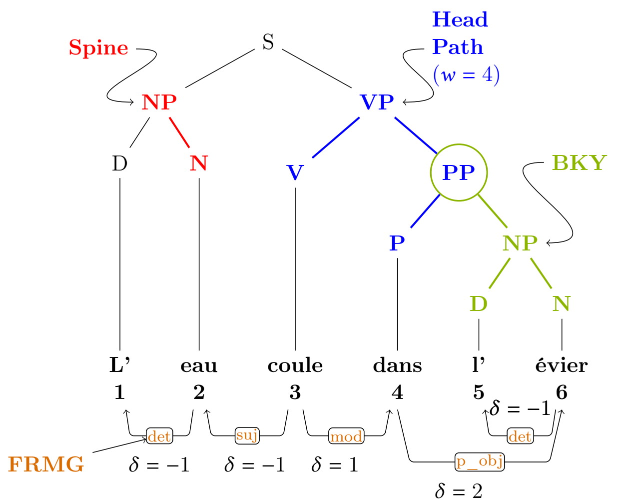
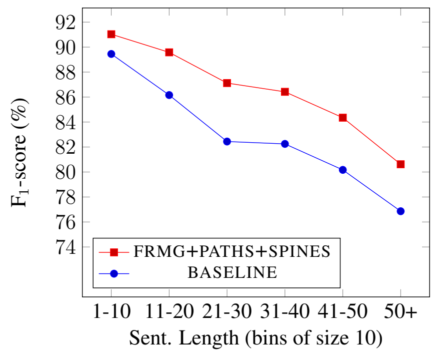


Figure 2: Syntactic Features Overview *(The water flows down the drain)*

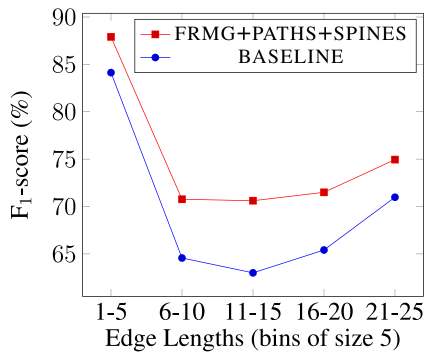
much more than the constituency fragments. We also observe that improvements for BKY and SPINES features are almost the same. This tendency gets stronger when combining the features with PATHS and FRMG, where the difference between the best models is of 0.03 points.

As regards FRMG features, because of the extended domain of locality of its elementary unit (tree-based), attachment decisions are taken with a more global view than classical transition-based parsers. In facts, these decisions ought to be more accurate in the case of complex linguistic phenomena such as coordinations, *etc.* This was suggested by the state-of-the-art results on French using a transition-based parser and such TAG-based features (Villemonte de la Clergerie, 2014). As a matter of fact, the parser is able to cope, for example, with a few cases of elliptic coordinations and so we expect that the resulting surface trees would provide more accurate guiding information for building a deep representation.

Expected results are observed using syntactic features that improve over a baseline as it was already demonstrated for DM (Ribeyre et al., 2015). However, it is important to understand what is indeed improved with those features. Figure 3 gives a detailed analysis when increasing the length of the sentence and the length of edges. We observe that the increase is two times higher with longest dependencies than with short dependencies (Fig. 3(b)). This is expected when considering our low recall: when we include wider contexts into the parsing model, we enable it to recover longest dependencies that are common in complex constructions such as elliptic coordinations. This is corroborated by the increase in performances with respect to the sentence length. For short sentences (between 1 and 10 words), the improvements is small (around 1.5 points), whereas it increased by a factor of 4 (around 6 points) for longer sentences.



(a) Sentence Lengths.



(b) Long-distance Dep.

Figure 3: Detailed Analysis TSPARSER (dev. set)

Discussion To assess the validity of our approach, we used a beam-based transition parser with early aggressive updates (DYALOG-SR, (Villemonte De La Clergerie, 2013a)), capable of handling general graphs through an extended set of transitions[5] described in (Ribeyre et al., 2014b).

Test set	DEEPFTB	DM
BEST (TSPARSER)	85.18	89.70
BEST (DYALOG-SR)	82.92	85.66
BASELINE (TSPARSER)	80.79	88.08
BASELINE (DYALOG-SR)	75.42	83.91
TSPARSER (SURF.)+RULES	80.45	-

Table 4: Comparison of baselines and best LF results for DEEPFTB and DM. (DEEPFTB's best: FRMG+PATHS+SPINES & DM's best: BN+SPINES+PATHS).

Table 4 reports a comparison between the best results for DEEPFTB and DM and the baseline for both parsers. The last row includes results from the TSPARSER, trained on the FTB surface dependencies,[6] whose outputs were fed to a tree-to-graph rewriting system (Ribeyre et al., 2012), following Ribeyre (2016).[7] This setup provides slightly inferior performance than the TSPARSER baseline parser and is vastly over-performed by our best setup (by almost 5pt). Despite validating our approach, this leads us to wonder if a graph-to-graph rewriting system could not be developed to push the envelope even further. This is left for future work. Interestingly enough, except for the fact that our feature set generalizes well with another parser, we see that the best model for both corpora are of the same kind: mixing dependencies information with spinal trees and head paths. Even tough these corpora differ in terms of constituent and dependency annotations at the surfacic and deep levels, both parsers need vertical (spines) and horizontal (paths) contexts combined with the functional label provided by the dependencies to be able to accurately predict argument structures. This seems to corroborate the hypothesis that when going further into abstracting away from syntactic divergences, argument-structure retrieval on French and English benefits from the same topological extra information, regardless of the language. Extensive cross-language experiments would be of course required to explore this potentially interesting point.

6. Conclusion

In this paper, we investigated deep syntactic parsing for French. We showed that mixing topologically different sort of syntactic features provides contextual information that improves the prediction of deep syntactic graphs. We also observed that the best models in French are the same for DM in English, regardless of the difference of language and annotation scheme (both at the surfacic and deep levels). This is coherent with the intuitive belief that language differences diminish when abstracting away from morphological and (surface) syntax variation.

[5]The parser also uses noop transitions, allowed on final items, in order to compensate for paths of various lengths (avoiding to favor either longest or shortest paths).

[6]With performance (LAS/UAS) on the FTB test set of 80.45/84.42 and 83.60/84.43 on the dev set.

[7]Using DYALOG-SR as a basis, this architecture was used to annotate the DEEPFTB, see (Ribeyre et al., 2014a) for details.

Acknowledgment

We thank Andre T. Martins for making his parser available and for kindly answering our questions, our anonymous reviewers for their comments and Marie Candito for her useful suggestions and insightful comments. This work was carried out while the first author was affiliated to the *Inria's Alpage project* and partly funded by the Program "Investissements d'avenir" managed by the *Agence Nationale de la Recherche* ANR-10-LABX-0083 (Labex EFL).

Bibliographical References

Abeillé, A. and Barrier, N. (2004). Enriching a French treebank. In *Proc. of LREC*.

Abeillé, A., Clément, L., and Toussenel, F., (2003). *Building a Treebank for French*. Kluwer, Dordrecht.

Artzi, Y., Lee, K., and Zettlemoyer, L. (2015). Broad-coverage CCG Semantic Parsing with AMR. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1699–1710, Lisbon, Portugal, September. Association for Computational Linguistics.

Ballesteros, M., Bohnet, B., Mille, S., and Wanner, L. (2014). Deep-Syntactic Parsing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1402–1413, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.

Berant, J. and Liang, P. (2014). Semantic parsing via paraphrasing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1415–1425, Baltimore, Maryland, June. Association for Computational Linguistics.

Berant, J., Chou, A., Frostig, R., and Liang, P. (2013). Semantic Parsing on Freebase from Question-Answer Pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1544, Seattle, Washington, USA, October. Association for Computational Linguistics.

Beschke, S., Liu, Y., and Menzel, W. (2014). Large-scale CCG Induction from the Groningen Meaning Bank. In *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, pages 12–16, Baltimore, MD, June. Association for Computational Linguistics.

Candito, M., Crabbé, B., and Denis, P. (2010). Statistical french dependency parsing: Treebank conversion and first results. In *Proc. of LREC*.

Candito, M., Perrier, G., Guillaume, B., Ribeyre, C., Fort, K., Seddah, D., and De La Clergerie, É. (2014). Deep Syntax Annotation of the Sequoia French Treebank. In *International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Islande, May.

Carreras, X. and Màrquez, L. (2005). Introduction to the

CoNLL-2005 Shared Task: Semantic Role Labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, CONLL '05, pages 152–164, Stroudsburg, PA, USA. Association for Computational Linguistics.

Chen, J. and Rambow, O. (2003). Use of deep linguistic features for the recognition and labeling of semantic arguments. In *Proc. of the 2003 conference on Empirical methods in natural language processing*, pages 41–48. Association for Computational Linguistics.

de La Clergerie, É. (2010). Convertir des dérivations TAG en dépendances . In *Proc. of TALN*.

Dybro-Johansen, A. (2004). Extraction Automatique De Grammaires D'Arbres Adjoints à Partir D'Un Corpus Arboré Du Français. Master's thesis, Université Paris 7.

Farkas, R., Bohnet, B., and Schmid, H. (2011). Features for phrase-structure reranking from dependency parses. In *Proc. of the 12th International Conference on Parsing Technologies*, pages 209–214. Association for Computational Linguistics.

Flanigan, J., Thomson, S., Carbonell, J., Dyer, C., and Smith, N. A. (2014). A Discriminative Graph-Based Parser for the Abstract Meaning Representation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland, June. Association for Computational Linguistics.

Joshi, A. K. and Schabes, Y. (1997). Handbook of formal languages, vol. 3. chapter Tree-adjoining Grammars, pages 69–123. Springer-Verlag New York, Inc., New York, NY, USA.

Martins, A. F. T. and Almeida, M. S. C. (2014). Priberam: A turbo semantic parser with second order features. In *Proc. of the 8th International Workshop on Semantic Evaluation*, pages 471–476.

Martins, A., Smith, N., Figueiredo, M., and Aguiar, P. (2011). Dual Decomposition with Many Overlapping Components. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 238–249. Association for Computational Linguistics.

Moschitti, A., Pighin, D., and Basili, R. (2008). Tree kernels for semantic role labeling. *Computational Linguistics*, 34(2):193–224.

Oepen, S., Kuhlmann, M., Miyao, Y., Zeman, D., Flickinger, D., Hajic, J., Ivanova, A., and Zhang, Y. (2014). Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In *Proc. of the 8th International Workshop on Semantic Evaluation*, pages 63–72.

Øvrelid, L., Kuhn, J., and Spreyer, K. (2009). Improving Data-driven Dependency Parsing Using Large-scale LFG Grammars. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 37–40, Stroudsburg, PA, USA. Association for Computational Linguistics.

Perrier, G., Candito, M., Guillaume, B., Ribeyre, C., Fort, K., and Seddah, D. (2014). Un schéma d'annotation en dépendances syntaxiques profondes pour le français. In *Proc. of TALN 2014*, Marseille, France.

Petrov, S., Barrett, L., Thibaux, R., and Klein, D. (2006). Learning accurate, compact, and interpretable tree annotation. In *Proc. of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*.

Ribeyre, C., Seddah, D., and Villemonte De La Clergerie, É. (2012). A Linguistically-motivated 2-stage Tree to Graph Transformation. In Chung-Hye Han et al., editors, *Proc. of TAG+11*, Paris, France. INRIA.

Ribeyre, C., Candito, M., and Seddah, D. (2014a). Semi-Automatic Deep Syntactic Annotations of the French Treebank. In *The 13th International Workshop on Treebanks and Linguistic Theories*, In Proceedings of the 13th International Workshop on Treebanks and Linguistic Theories, Tübingen, Germany, December. Tübingen Universität.

Ribeyre, C., Villemonte De La Clergerie, É., and Seddah, D. (2014b). Alpage: Transition-based Semantic Graph Parsing with Syntactic Features. In *International Workshop on Semantic Evaluation*, Dublin, Irlande, August.

Ribeyre, C., Villemonte De La Clergerie, É., and Seddah, D. (2015). Because Syntax does Matter: Improving Predicate-Argument Structures Parsing Using Syntactic Features. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, USA, June.

Ribeyre, C. (2016). *Méthodes d'Analyse Supervisée pour l'Interface Syntaxe-Sémantique: De la Réécriture de Graphes á l'Analyse par Transitions*. Ph.D. thesis, Université Paris Diderot, Paris, France.

Sagae, K. and Tsujii, J. (2008). Shift-reduce dependency DAG parsing. In *Proc. of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 753–760.

Seddah, D., Tsarfaty, R., Kübler, S., Candito, M., Choi, J. D., Farkas, R., Foster, J., Goenaga, I., Gojenola Galletebeitia, K., Goldberg, Y., Green, S., Habash, N., Kuhlmann, M., Maier, W., Nivre, J., Przepiórkowski, A., Roth, R., Seeker, W., Versley, Y., Vincze, V., Woliński, M., Wróblewska, A., and de la Clergerie, E. V. (2013). Overview of the SPMRL 2013 shared task: A cross-framework evaluation of parsing morphologically rich languages. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 146–182, Seattle, Washington, USA, October. Association for Computational Linguistics.

Seddah, D. (2010). Exploring the spinal-stig model for parsing french. In *Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10)*.

Villemonte De La Clergerie, É. (2013a). Exploring beam-based shift-reduce dependency parsing with DyALog: Results from the SPMRL 2013 shared task. In *4th Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL'2013)*.

Villemonte De La Clergerie, É. (2013b). Improving a symbolic parser through partially supervised learning. In *The 13th International Conference on Parsing Technologies*

(IWPT), Nara, Japon.

Villemonte de la Clergerie, É. (2014). Jouer avec des analyseurs syntaxiques. In *Proceedings of TALN 2014 (Volume 1: Long Papers)*, pages 67–78, Marseille, France, July. Association pour le Traitement Automatique des Langues.