Deeper syntax for better semantic parsing

Olivier Michalon[⋆] **Corentin Ribeyre**[†∗] **Marie Candito**[⋄] **Alexis Nasr**[⋆]
⋆ Aix Marseille Univ, CNRS, Centrale Marseille, LIF, Marseille, France
† Faculty of Humanities, University of Geneva, Switzerland
⋄ Alpage, Université Paris Diderot, Sorbonne Paris Cité, INRIA, Paris, France
⋆ firstname.lastname@lif.univ-mrs.fr
† firstname.lastname@unige.ch
⋄ firstname.lastname@inria.fr

Abstract

Syntax plays an important role in the task of predicting the semantic structure of a sentence. But syntactic phenomena such as alternations, control and raising tend to obfuscate the relation between syntax and semantics. In this paper we predict the semantic structure of a sentence using a deeper syntax than what is usually done. This deep syntactic representation abstracts away from purely syntactic phenomena and proposes a structural organization of the sentence that is closer to the semantic representation. Experiments conducted on a French corpus annotated with semantic frames showed that a semantic parser reaches better performances with such a deep syntactic input.

1 Introduction

FrameNet (Baker et al., 1998) is an English resource containing a set of inter-related semantic frames, each frame containing a set of semantic roles (*frame elements* in FrameNet's terminology). Frames offer semantic generalizations over individual predicates, since different lexical units can evoke the same frame, and semantic roles offer generalizations over syntactic arguments. Hence FrameNet parsing can be viewed as mixing predicate disambiguation and semantic role labelling.[1]

Although FrameNet is more semantically-oriented than other semantic role labeling resources such as PropBank (Palmer et al., 2005), syntactic information has been shown to be decisive for predicting (FrameNet) semantic roles since the early days of the task (Gildea and Jurafsky, 2002). Linking regularities provide the theoretical justification of this result: there exist regularities in how semantic arguments are realized in syntax. Yet it is well known that the mapping from syntactic arguments to semantic ones is not straightforward. First, lexical idiosyncrasies can come into play, for instance the *Addressee* of communication verbs may correspond to the indirect object for verbs like *to say* and to the direct object for a verb like *to inform*. Second, it is also well known that surface syntax exhibits variation that can obfuscate regularities. For instance though the *Speaker* is generally the subject of communication verbs, this does not hold when the verb is passivized. This difference disappears if syntactic alternations are neutralized, and the "canonical" diathesis of a verb is made explicit: the *Speaker* is the canonical subject in both active and passive voices.

In this paper, we investigate the syntax-semantic interface in FrameNet annotated data, and study the impact of using "deeper" syntactic features to predict semantic frames and roles. More precisely, we take advantage of a deep syntactic dependency graphbank for French (Candito et al., 2014b; Ribeyre et al., 2014), which provides a level of representation that abstracts away from purely syntactic variation. The main contributions of the paper are (i) a comparison of the syntax/semantic regularities observed when using plain "surface" syntax to those observed when using deep syntax and (ii) a study of how and why the switch from surface to deep syntax impacts FrameNet semantic parsing. In the remaining of the

[∗] All of his work has been done during his PhD at Alpage.

[1] In the following, we will use shorter terms than those of FrameNet terminology : we use the term *trigger* for a lexical unit that can evoke a frame, the term *role* for frame element, and *role filler* for the sequence of words that instantiates a role.

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 409–420, Osaka, Japan, December 11-17 2016.

paper we will use the terms Surface Syntactic Representations (SSR) and Deep Syntactic Representations (DSR) to refer to surface syntactic trees and deep syntactic graphs.

Using abstract syntactic representations as an intermediate representation level between syntax and semantics has been proposed in different theoretical frameworks, such as derived trees of Tree Adjoining Grammars (Joshi and Schabes, 1997) or deep syntactic structures of the Meaning Text Theory (Mel'čuk, 1988). But we only found few works showing, empirically, that using such representations can effectively help predict the semantic roles of predicates.Two of them concern PropBank semantic role labeling. The early (Gildea and Hockenmaier, 2003) work shows that using CCG-derived predicate-argument features predicted by a CCG parser improves the identification of core PropBank arguments. Vickrey and Koller (2008) investigate the use of simplified syntactic paths and report a slight improvement when applying transformation rules to simplify phrase-structure parses.

As far as FrameNet parsing is concerned, we don't know of any work using more abstract syntactic input than plain "surface" syntactic trees, whether phrase-structure (Gildea and Jurafsky, 2002) or dependency trees (Johansson and Nugues, 2007; Das et al., 2014). We focus on French, first because of the availability of the afore-mentioned DSR, and second because in the French FrameNet corpus (Djemaa et al., 2016) the annotated semantic roles are restricted to essential arguments. On the contrary, both essential ("core") and non essential participants are annotated in the English FrameNet, including modifiers such as time, location, purpose etc... But syntactic variation such as syntactic alternations, VP coordination, control etc... does concern primarily the most salient grammatical functions (subject, direct object, indirect object etc...), which are typically the ones that essential arguments bear. Hence, neutralizing syntactic variation is expected to have an impact primarily on essential semantic roles.

The structure of the paper is the following: in section 2, we present (i) the French FrameNet corpus that we use, (ii) the deep syntactic representations whose impact for FrameNet parsing we wish to investigate, (iii) we compare the syntax/semantic interface when using surface dependency trees and deep dependency graphs and (iv) we compare such deep representations to other deep representations proposed mainly for English. Section 3 and 4 are devoted to the frame-semantic parser and the deep-syntactic parsing architecture we used. We present and discuss the frame-semantic parsing experiments in section 5, and conclude in section 6.

2 Deep syntax and frame semantics

2.1 French FrameNet corpus

The French FrameNet annotated corpus (Djemaa et al., 2016) was produced within the ASFALDA ANR project on French shallow semantic parsing[2]. Two corpora have been annotated with frames and roles: the French Treebank (Abeillé and Barrier, 2004) (hereafter FTB) and the Sequoia Treebank (Candito and Seddah, 2012b). The first one contains $18,535$ sentences from the *Le Monde* newspaper. The second one is much smaller and was originally created for domain adaptation experiments for statistical parsing. It contains $3,099$ sentences from a regional newspaper, from Europarl, from the European Medicine Agency and from the French Wikipedia.

The French FrameNet corpus annotation is restricted to four semantic domains: commercial transactions, cognitive stances, verbal communication and causality. For all lexical items of the lexicon, associated with frames pertaining to these domains, the first 100 occurrences have been annotated. For each occurrence to annotate, annotators were proposed the pertaining frames, plus a special null frame for the cases in which the occurrence evoked a sense not pertaining to the four domains. We provide quantitative characteristics of the corpus in Table 1. The semantic annotations cover 105 frames, and the lexicon extracted from the annotations contains 1112 frame/lemma pairs (i.e. senses). The corpus contains $15,990$ annotated frame occurrences (plus 8727 occurrences of the null frame[3]), 56.2% of which correspond to verbal triggers and 33.0% to noun triggers.

[2] Version 1.0, https://sites.google.com/site/anrasfalda/

[3] The null frame is used to annotate words that would trigger a frame that has not been defined yet. Note that trigger occurrences ahead of the first 100 occurrences do not bear any frame at all, and are not to be considered.

Nb. Sentences	21 634	Nb. annot. frame occurrences	15 990
Nb. Tokens	625 951	Nb. annot. role occurrences	24 147
Nb. distinct annot. frames	105	Mean nb. annot. frames per lemma	18.3
Nb. distinct annot. lemma/frame pairs	1112	Median nb. annot. frames per lemma	6

Table 1: Quantitative characteristics of the French FrameNet annotated corpus (excluding the null frame).

2.2 Deep syntactic representations

We now turn to the deep syntactic graphbank that we use as an alternative syntactic representation for FrameNet parsing. DSRs are available for the two corpora that were annotated with frames and roles (the Sequoia corpus and the French Treebank). The development set of the Sequoia corpus was used to set up the deep syntactic annotation scheme, as well as a surface-to-deep syntax conversion module (Ribeyre et al., 2014) based on a graph-rewriting tool (Ribeyre et al., 2012). While the DSRs were manually validated for the full Sequoia corpus, those for the FTB sentences were automatically obtained using this surface-to-deep syntax conversion module, described in section 4. The quality of the resulting DSRs is high enough to use them as a reference for evaluation[4].

Candito et al. (2014b) define DSRs as dependency graphs which abstract away from purely syntactic variations, as far as verbal and adjectival predicates are concerned, making explicit their predicate-argument structure. SSR and DSR differ on three aspects:

- **Saturation:** The predicate-argument structure of all verbs is saturated for verbs that are not the head of a saturated clause (e.g. coordinated verbs, infinitival verbs). Any element that does not locally depend on the verb but that would do so if the verb were the head of a clause is added as (deep) dependent of the verb. First, this means that arguments shared by several verbs, e.g. in elliptic coordinations or control verb constructions, are attached to all their deep governors. For instance in *Paul loves to eat pies*, *Paul* is the subject of both *loves* and *eat*, and in *Paul loves and often eats pies*, the two coordinated predicates *loves* and *eats* share the same subject *Paul* and direct object *pies*. Second, noun-modifying verbs get the noun as deep syntactic dependents. For instance in *People born before 1969 fear the moon*, the verb *born* gets *People* as subject.
- **Syntactic alternations:** Productive syntactic alternations are neutralized. Syntactic arguments of verbs get their *canonical* grammatical function, which may differ from the observed grammatical function. The most frequent alternations are the passive alternation, then middle and neuter alternations, each marked with a *se* clitic. Other more marginal alternations are impersonal, impersonal passive and causatives. Note that alternations frequently interact with predicate-argument structure saturation. For instance, in *Paul would like to get an interview and then be hired*, *Paul* is added as canonical subject of *get* but canonical object of *hired*. In noun-modifying participial clauses, if the verb is transitive, the past participle is analyzed as a passive. For instance in *People hired after march are few*, the verb *hired* gets *People* as canonical direct object (see also the verb *poussée (urged)* in figure 1).
- **Abstraction:** Most grammatical markers are discarded. Auxiliaries in particular are replaced by deep features on the lexical verb. Empty prepositions and complementizers are bypassed For instance in *Le chat sourit à la souris (The cat smiles to the mouse)*, the preposition *à* is discarded, and the indirect object of the verb is the NP *la souris (the mouse)*.

By extension, the subjects[5] of adjectives are made explicit in the DSRs.

The DSRs are closer to predicate-argument structures than SSRs are, yet predicates are not disambiguated, and thus canonical grammatical functions are used and not semantic roles.

Figure 1 shows the SSR, DSR and FrameNet annotations for one sentence (the role fillers are reduced to their syntactic head, cf. section 2.3). It can be seen, for instance, that the past participle *poussée*

[4]Ribeyre et al. (2014) report a 98.4 Fscore evaluated on manually validated DSRs for 200 sentences from the FTB.

[5]The subject of the adjective is either the noun it modifies in case of an attributive adjective, or the subject of the copular verb in case of a predicative adjective.

(urged) modifies the proper noun *EDF* in both syntactic representations, but the noun is its canonical direct object in the deep representation.

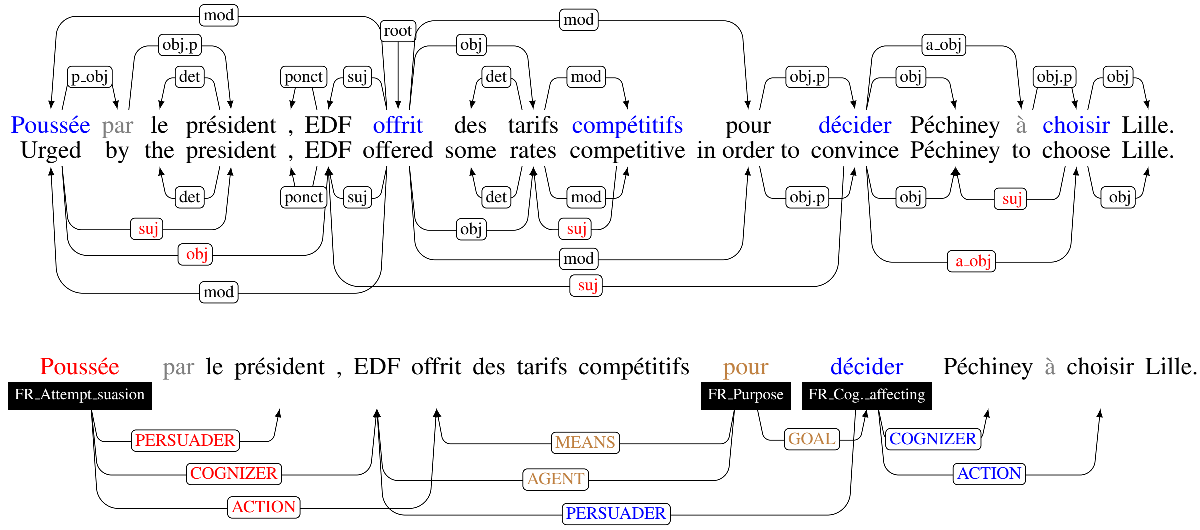

Figure 1: Example of syntactic and semantic annotations for a sentence. **Top**: Surface and deep syntactic representations (edges above: SSR, edges below: DSR). Verbs and adjectives are in blue. Tokens discarded in the DSR are in gray. Grammatical functions added and/or normalized when switching from SSR to DSR are in red. **Bottom**: frame and role annotations (from trigger to syntactic head of role fillers), for three triggers (2 verbs and 1 preposition).

2.3 Syntax semantics interface

As already mentionned in the introduction, syntax is a major feature when predicting semantic roles. Reducing the variety of syntactic features might therefore help fighting against data sparsity and improve this prediction task. Because they are meant to neutralize syntactic variations, DSRs are good candidates for such a reduction. In all the following, we will use as syntactic features the syntactic paths that link a frame trigger to the syntactic head of each of its role fillers (the head is taken as the leftmost root of the subtrees composing the role filler). In this section we will measure how much the use of DSRs helps to reduce the variety of syntactic paths. In order to do so, we will compute the entropy of the probability distribution of the syntactic paths that correspond to a role. Two entropies will be compared: the *surface* syntactic entropy and the *deep* syntactic entropy.

Before defining surface and deep syntactic entropy, we need to define precisely the notions of semantic path: deep syntactic path (DSP) and surface syntactic path (SSP). Given sentence S that contains an occurrence of frame F having word t as trigger and which role R is filled by a sequence of tokens W (the role filler, which may be discontinuous). We will call the tuple (t, R, W) a semantic path of S.

We associate to every semantic path $p = (t, R, W)$ of sentence S a surface syntactic path $SSP(p)$ and a deep syntactic path $DSP(p)$, which link the trigger to the head of the role filler W, noted $h(W)$.

$SSP(p)$ is the shortest path linking t and $h(W)$ in the SSR of S. The SSR being a tree, such a path exists and is unique, it is the sequence of dependencies that link t to $h(W)$. We represent it formally as a sequence of tuples *(direction,label)*, where *direction* is + if a dependency is traversed from the governor to the dependent and - otherwise.

Defining DSP is not as straigtforward: the DSR being a graph, there can be several shortest paths[6] from t to $h(W)$. We select a unique shortest path using the following hierarchy of grammatical functions to rank paths of length one: $suj > obj > ats/ato > a\_obj > de\_obj > p\_obj > mod$.[7] The left part of Table 2 shows the five most frequent syntactic paths, when the trigger is a verb, using either surface or

[6]Actually, since deep syntax can form non connected oriented graphs, there can be no path at all between t and $h(W)$ (due to errors in deep syntax or in semantic annotations). We use the special tag *__no_path__* in such cases.

[7]With $a > b$ meaning a has priority over b and a/b meaning a has the same priority as b.

surface syntax		deep syntax	
(+suj)	25.02%	(+suj)	33.1%
(+obj)	17.01%	(+obj)	32.79%
(-mod)	8.04%	(+a_obj)	4.73%
(+obj,+obj.cpl)	4.42%	(-mod)	3.15%
(+a_obj,+obj.p)	4.09%	(+mod,+obj.p)	2.46%
Total	58.58%	Total	76.23%

role	synt. head	syntax	path
PERSUADER	*EDF*	surface	(-obj.p,-mod,+suj)
		deep	(+suj)
COGNIZER	*Péchiney*	surface	(+obj)
		deep	(+obj)
ACTION	*choisir*	surface	(+a_obj,+obj.p)
		deep	(+a_obj)

Table 2: **Left**: Most frequent gold syntactic paths in training corpus, when the trigger is a verb. **Right**: surface and deep paths for the *FR_cognizer_affecting* frame evoked by *décider* in the sentence of Figure 1.

deep syntax. We can see that the distribution of paths is much more compact when using deep syntax : the first five paths represent more than 76% of deep paths, compared to 58% for surface paths. (obj) and (suj) paths represent 42.03% of SSP but 65.89% of DSP (in order to reach that coverage with SSP, the 8 most frequent SSP are needed).

The right part of Table 2 shows the deep and surface paths corresponding to the roles of the *FR_cognizer_affecting* frame evoked by *décider* in the sentence of Figure 1.

In order to measure the reduction of the variety of the syntactic realization when moving from surface to deep syntax, we have computed the average entropy over all roles R of the probability distributions $P(p|R)$ where p is a path. These distributions have been estimated on the training corpus. The average entropy when computed on surface syntax is equal to 1.65 and to 1.32 when computed on deep syntax. This decrease is a direct measure of the normalizing effect of the deep syntactic frame we used. Note though that an entropy reduction could be artificially obtained by neutralizing meaningful syntactic distinctions. Yet, the DSRs were designed following syntactic principles and experiments in section 5 are intended to check that such a normalization is indeed beneficial for downstream semantic parsing.

2.4 Comparison with other deep representations

There has been various previous works proposing deeper syntactic annotation schemes that can represent information absent in plain constituency or dependency trees, such as long-distance dependencies, subjects of control verbs, subjects of coordinated verbs etc. This additional information is sometimes viewed as pertaining to semantic representations, sometimes retained as still syntactic.

English has been the first focus language, along with Czech thanks to the Prague Dependency Treebank (Hajič et al., 2006). For English, several works automatically convert Penn Treebank constituency trees into deeper representations, based on lexicalized grammar formalisms such as LFG, CCG or HPSG. Cahill et al. (2004) automatically construct LFG f-structures from PTB trees, a work adapted for various other languages including French (Schluter and van Genabith, 2008). Hockenmaier and Steedman (2007) extracted a corpus of CCG derivations and dependency structures from the Penn Treebank. These two kinds of deeper representations do capture long distance dependencies, subjects of non finite verbs, argument sharing between coordinated verbs. When compared to the DSRs we use though, the main missing trait is the neutralization of syntactic alternations, which we believe is a major source for the syntactic path normalization effect described in section 2.3[8].

The Stanford dependencies (SD, De Marneffe and Manning (2008)) constitute another proposal for obtaining dependencies not directly present in surface syntactic trees. The stanford parser comprises a dependency extraction system, which can output several variants of typed word-to-word dependencies, from plain dependency trees to more semantically-oriented graphs. The deepest variant ('collapsed with propagation of conjunct dependencies' variant) does cope with some of the aforementioned phenomena such as subject of infinitival verbs or coordinated verbs. Compared with the DSRs for French, the major differences are that syntactic alternations are not neutralized, and that all prepositions are collapsed and injected in the labels (while only void prepositions are collapsed in the French DSRs).[9]

[8]Passive alternations is by far the most frequent alternation, and also happens to be rather easy to identify, so we hypothesize that using such representations on top of passive neutralization would be an alternative to the DSRs we use.

[9]We actually did some unfruitful experiments on the English FrameNet data, comparing the use of syntactic features ex-

Taking a further step towards semantic representations, predicate-argument structure graphs such as those used for the Broad-Coverage Semantic Dependency Parsing task at SemEval 2014 (Oepen et al., 2014) are also very close to the DSRs we use, with respect to the covered linguistic phenomena. The three datasets used in this shared task are (i) predicate-argument semantic graphs extracted from the HPSG-grounded DeepBank of Flickinger et al. (2012), (ii) predicate-argument structures from the Enju HPSG Treebank [10], and (iii) the Prague Czech-English Dependency Treebank (Hajič et al., 2012). These three datasets differ in how far they differ from syntactic representations. While some traits are common to the DSRs we use, one major difference lies in the more semantically-oriented labelling of the word-word dependencies: the semantic arguments are simply numbered (arg0, arg1, etc...). We believe that in the absence of word sense disambiguation at the level of predicates, this plain numbering obfuscates syntactic clues that are crucial for FrameNet semantic role labelling. If we take a French example, the verb *convenir* has two senses (among others), in which the arguments bear different FrameNet roles, and which can be disambiguated by the canonical subcategorization frame: we have *X(subject) convenir à Y(a-object)* meaning "X suits Y" versus *X(subject) convenir de Y(de-object)* meaning "X admit to Y".

To sum up, while the various deep representations cited above do capture the topology of predicate-argument structures, by coping with major phenomena such as control verbs or coordinated verbs, the DSRs are appealing for framenet parsing for two reasons: first they are still syntactic in nature (they are thus recoverable deterministically from surface syntax, cf. section 4), while a semantic graph would represent a too sophisticated input for the task. Second, the DSRs use canonical grammatical functions, which are both more abstract than surface grammatical function labels, but do not obfuscate important syntactic clues for predicate and role disambiguation.

3 Semantic parser

The semantic prediction system (FastSem) is a baseline system based on a cascade of linear classifiers[11]. For every word w of a sentence, we proceed in two steps. A *frame* identification step (which frame (if any) does w trigger?) followed by a *role* identification step (which role (if any) is w the head of?). This architecture is based on two strong independence hypotheses: frames are independent from one another in a sentence and roles inside a frame are independent[12].

We chose to use a simple architecture as our focus here is to assess whether normalized syntactic paths help semantic parsing. It remains to be proved, although it can be easily supposed, that it would also help with less naive hypotheses.

In the first step we use for each lexical unit a different linear classifier, each using the following features: the fine- and coarse-grained PoS of the target word t, and for each word w of the sentence, its lemma, its PoS (fine and coarse) and the syntactic path that links t to w. The classifier used for the second step is frame specific. To predict the role of word f, we use as features the lemma and PoS (coarse and fine) of f, t's lemma and fine-grained PoS, the syntactic path between t and f, plus the combination of the syntactic path and the lemma of t.

4 Predicting deep syntax

In order to evaluate the impact of deep syntax on semantic parsing in realistic conditions, we need to obtain predicted deep syntactic representations. Although directly training a graph parser would be an option (as in (Ribeyre et al., 2016)), we retain the rule-based architecture that was used to bootstrap the deep syntactic annotations. Our motivation is to be able to apply the surface-to-deep rewriting rules step-by-step, in order to study the impact of each phenomenon.

tracted from two variants of SD (basic versus collapsed with propagation of conjunct dependencies). We concluded that the collapsed dependencies are not adapted for our purpose: they no not neutralize syntactic alternations, and multiply labels by collapsing all prepositions. We could measure that this has the result of actually increasing the entropy of the syntactic paths that correspond to a role. Preposition collapsing has a negative impact on predicting non essential semantic roles, such as temporal or locative modifiers.

[10] See http://kmcs.nii.ac.jp/enju

[11] The classifier library used is LIBLINEAR (Fan et al., 2008).

[12] These hypothesis are known to be too strong. For instance Das et al. (2014) show that collectively predicting all role fillers of a given frame occurrence improves performance.

	Surface dependency parser		Conversion to deep syntax			
			on predicted surf. parses		on gold surf. parses	
	UAS	LAS	UF1	LF1	UF1	LF1
trainjk	86.9	83.5	83.5	80.4	99.7	99.5
dev	87.5	84.1	84.1	81.0	99.7	99.4
test	86.6	83.3	83.5	80.5	99.7	99.5

Table 3: Parsing performance. Columns 2 and 3: unlabeled and labeled attachment scores of the (surface) dependency parser. Last four columns: unlabeled and labeled F-scores after classification of *il/se* clitics and conversion to deep syntax, applied either on the predicted surface dependency parses (columns 4 and 5) or on the gold dependency parses (last 2 columns). Results on the training set are obtained using a 10-fold jackknifing. Results on the dev and test set are obtained using training on the full training set.Punctuation tokens are taken into account.

The surface-to-deep syntax conversion module of Ribeyre et al. (2014) takes as input surface dependency trees in which a few linguistic phenomena have already been made explicit, because they were considered difficult to capture by a rule-based approach. This is in particular the case for the status of the *il* and *se* clitics, which results from complex syntactic and lexical factors. In order to do so, we designed two classifiers that predict the status of these clitics. We omit to describe here these classifiers as well as their evaluation, for reason of lack of space.

The architecture of our deep syntactic parser is to apply sequentially **(i)** part-of-speech tagging and lemmatization, **(ii)** surface dependency parsing and **(iii)** surface-to-deep syntax rewriting rules.

Tagging and syntactic parsing were performed with MACAON (Nasr et al., 2011), a tool suite for standard NLP tasks. The tagging is based on a CRF model whereas the dependency parser is a second order graph-based parser, with standard features. We report parsing performance in Table 3 (first two colums). The scores are comparable to the baseline scores obtained by the SPMRL shared task participants on French (Björkelund et al., 2013), without any special handling of multi-word expressions.

The last step consists in applying the surface-to-deep syntax conversion module (Ribeyre et al., 2014). This module uses OGRE (Ribeyre et al., 2012), a deterministic two-stage graph rewriting system.

The first stage follows the Single Pushout Approach (SPO) (Rozenberg, 1997), a widely used method when dealing with graph rewriting system. This stage identifies graph patterns and applies rewriting operations such as adding an edge, removing an edge, changing a label, and so on. This is done in one pass and contrary to the SPO approach, the first stage is executed only once.

The second stage is a propagation step. During the first stage, the rewriting rules may have left what we call *triggers* on edges. Those are special actions that, given a specific edge configuration, apply a serie of rewriting steps using a fixed-point algorithm: when all possible rewritings have been done, the algorithm terminates. It is especially helpful in case of linguistic phenomena interacting with each other. In the SSR of the sentence *John seems to want to give a book to Mary*, for example, *John* is the subject of *seems* and *want* is a dependent of *seems* and *give* a dependent of *want*. Ultimately, in the DSR, *John* is the subject of both *want* and *give*. The interaction between raising and control verbs is obtained through the propagation of rules of the form "if V_1 taking V_2 as complement has or gets a final subject X then add X as final subject of V_2". Moreover, the two-stage rewriting system ensures that the algorithm terminates and the system is confluent. See (Ribeyre, 2016) for more details and proofs.

The surface-to-deep syntax module applies sequentially 5 sets of rewriting rules:

1. The first set converts tense auxiliaries into mood and tense features on the lexical verb.
2. The second set distributes dependents of coordinated predicates and identifies the final subject of non finite verbs and by extension, of adjectives also, whether used as predicative complements or noun modifiers.
3. Syntactic alternations are mainly handled in the third set, which identifies the canonical grammatical functions for arguments of verbs (whether already present in the surface tree, or added by the second

	Prec.		Recall		F-measure		Prec.		Recall		F-measure	
Input	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR
trigger detec.	89.3	89.4	88.7	88.7	89	89	88.6	88.8	88.4	88.3	88.5	88.6
frame selec.	81.1	81.2	80.6	80.6	80.8	80.9	80.3	80.5	80.2	80	80.3	80.2
role detec.	85.1	86.2	59.2	62.6	69.8	72.5	79.6	81.3	51.7	55.7	62.7	66.1
role selec.	77.9	80.9	54.2	58.7	**63.9**	**68.1**	72	75.9	46.7	52	**56.7**	**61.7**

Table 4: FastSem results for **all triggers**, using **gold** (left) and **predicted** (right) SSR and DSR.

module).

4. The fourth set handles comparative and superlative constructions mostly.
5. The last set exclusively deals with bypassing the semantically empty words.

We provide the performance evaluation of DSR prediction step in Table 3. Columns 4 and 5 show the result of the whole parsing architecture, where steps (i), (ii) and (iii) are predicted. The last two columns show the result of applying step (iii) on gold SSR. Not surprisingly, the DSR built from gold SSR are almost perfect. This is due to the fact that the deep syntactic corpus contains gold DSR for the small Sequoia part only, the other part, which corresponds to the FTB, is made of pseudo-gold DSR obtained by the application of step (iii) on gold SSR ! The table shows the sharp drop in quality when DSR are computed on predicted SSR.

5 Experiments and discussion

We now turn to FrameNet parsing experiments, meant primarily to compare the use of surface versus deep syntactic paths as features. All experiments were used using the same split.[13] Feature engineering was performed on the development set.

5.1 Evaluation metrics

The train, dev and test examples are made of the set of annotated frame occurrences of the train, dev and test sets, including the null frame cases. For each setting, we trained word specific classifiers for the frame selection step and frame specific classifiers for the role selection step. But, since selecting the null frame is a rather easy task, we chose to evaluate each of the two steps using two different metrics. For frame selection, we first evaluate the task of deciding whether a word triggers an actual frame or the null frame. The results are reported in lines "trigger detection" of Table 4. The "frame selection" lines report the precision, recall and F-scores of choosing a frame, computed when setting aside the triggers whose gold frame is the null frame.

For role labeling, prediction and evaluation is made on heads of role fillers only. It is also broken in two: we first evaluate the task of deciding whether a word plays a role or not with respect to the trigger (reported in the "role detection" lines in the result tables). Then, for words that are actually head of role fillers in gold data, we compute precision, recall and F-score of the head and role pairs predicted by our semantic parser (reported in the "role selection" lines in the tables). Note that in both cases, the role is counted as incorrect if the frame was not predicted correctly.

5.2 Results and discussion

The experiments conducted vary according to two dimensions: the use of surface vs. deep syntactic paths (SSP or DSP) and whether they are predicted or gold. The predicted SSP are obtained using predicted PoS, lemmas, morphological features and surface dependency syntax. The predicted DSP are obtained by applying *il/se* classification and rewriting rules on predicted surface dependency trees (cf. section 4). All results are computed on the test set.

The left part of Table 4 shows results using gold syntactic structure, whether surface or deep. As can be seen, the results for the first three metrics slightly increase when switching from SSP to DSP, but

[13]The training set is the concatenation of the usual training sets of the Sequoia and FTB corpus. Same for the development and test sets.

	Prec.		Recall		F-measure		Prec.		Recall		F-measure	
Input	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR	SSR	DSR
frame selec.	80.1	80.7	80.1	80.7	80.1	80.7	80	80.5	80.8	80.9	80.4	80.7
role selec.	81.4	86.4	59.1	66.1	**68.5**	**74.9**	75.7	80.3	51.6	59	**61.3**	**68**

Table 5: FastSem results for **verbs**, using **gold** (left) and **predicted** (right) SSR and DSR.

	SSP	DSP				
		all	alt	byp	subj	coo
gold	68.5	74.3	70.6	69.1	69.3	70.2
predicted	61.3	68	63.3	63.1	62.4	63.1

Table 6: FastSem F-measure for role selection with application of deep rewriting rule sets in isolation, for verbal triggers. Rules are applied on SSP that are either gold (first row) or predicted (last row). The first column reports the results when using SSP. The second when using DSP with all rules applied. See text for description of the rule sets (alt) to (coo).

we obtain a 4.2 point improvement for the overall result of role selection when using DSP instead of SSP (63.9 to 68.1). Because our DSR focus on the predicate argument structure of verbs and adjectives, and because the number of adjectival triggers is marginal in the French FrameNet corpus, we chose to provide, in Table 5, the same metrics as in Table 4, computed on verbal triggers only. As one can see, using deep syntax provides substantial help for predicting roles: we obtain a 6.4 point improvement for role selection for verbal triggers (68.5 to 74.9).

We now turn to a more realistic setting in which all features for the semantic parser are predicted: lemmas, PoS, SSP and DSP. Not surprisingly, the results shown in Table 4 (all triggers) and 5 (verbal triggers) are overall lower than when using gold features. However, switching from surface to deep syntax leads to higher gain for predicted data than for gold data: 5.1 points (56.7 to 61.7) for all trigger, instead of 4.2 for gold data and 6.7 points (61.3 to 68) instead of 6.4 on gold data for verbal triggers. These results clearly show the benefit of using deep syntactic features.

The differences between SSP and DSP are of various kinds, as seen in section 2.2. We propose to study the impact of each phenomenon, by applying in isolation each set of graph-rewriting rules of the surface-to-deep syntax conversion module. More precisely, we applied in isolation (alt) the rules for syntactic alternations, (byp) the bypassing of empty prepositions and complementizers, (subj) the addition of subjects for non finite verbs and adjectives and (coo) the distribution of dependents to coordinated predicates. We provide the results in Table 6, for the role selection task, computed on verbal triggers only. It shows that every rule set contributes to a better prediction of the semantic structure.

5.3 Error analysis

In order to perform error analysis, we analyzed the changes in role selection when switching from SSP to DSP (table 7). The number of corrected errors (W→C) is more than four times the number of introduced errors (C→W). We reproduce below three cases of errors that were corrected when switching from surface to deep syntax. They correspond to syntactic alternation (1), coordination of VPs (2) and control verb (3). The trigger is in capital letters, and the (head of) role fillers we focus on are in bold:

	C→C	C→W	W→C	W→W
predicted	1163	47	218	481
gold	1362	48	203	316

Table 7: Improvements and degradations for role selection when switching from SSP to DSP, using either gold syntax (first row) or predicted syntax (second row). Break-down of the non-null gold roles of the dev set, when frames are correctly identified by both systems. C stands for correct, W stands for wrong.

freq. range	G1 (frequent) $> 10\%$		G2 (medium) $< 10\%$ and $> 1\%$		G3 (rare) $< 1\%$	
	Prop.	F1	Prop.	F1	Prop.	F1
SSR	42.1%	89.2	33.1%	78.3	24.8%	38.3
DSR	65.9%	92.3	14.3%	75	19.8%	23.5

Table 8: Role selection task results on the dev set, using gold frames triggered by verbs: break-down by frequency (in the training set) of the gold syntactic path. "Prop." columns provide the proportion of each sub-group.

1. *Cette **thérapie** a été DÉCIDÉE par le **gouvernement***
   (This therapy has been decided by the government.)
   thérapie: DSP=(+obj) SSP=(+subj)   **gouvernement**: DSP=(+subj) SSP=(+p_obj,+obj.p)
2. ***Grandier** avait publié un pamphlet et S'OPPOSAIT fermement à la destruction des murailles.*
   Grandier had published a pamphlet and was firmly opposed to the destruction of the walls.
   Grandier: DSP=(+subj) SSP=(-dep.coord,-coord,+subj)
3. ***Ils** ont essayé de les PERSUADER de bouleverser le calendrier.*
   They have tried to them persuade to change the schedule.
   Ils: DSP=(+subj) SSP=(-obj.p, -de_obj, +subj)

We also took a closer look at the introduced errors. They mostly correspond to cases in which the role filler has same surface and deep syntactic path, the path being rather unusual for the role filler. This may indicate that increased regularity of the DSP makes role fillers with unusual syntactic path more difficult to detect. We tried to assess this hypothesis by breaking-down the performance of the role selection task by frequency of the syntactic paths between the head of the role filler and the trigger. Results are shown in table 8. The frequent paths (G1) lead to better role prediction than the other two groups, and this is even more true when using DSRs than SSRs (92.3 versus 89.2). This explains most of the improvements, since this group represents a higher proportion when using DSRs than SSRs (65.9 versus 42.1). For less frequent paths (G2 and G3), results are either slightly (G2) or much (G3) better when using SSRs than DSRs, but these two groups represent a much lower proportion in the DSR paths than in the SSR paths. To sum up, frequent paths are even more frequent when using DSRs, and thus lead to better role prediction, whereas the non frequent paths exhibit the opposite trend.

6 Conclusion

In this paper we showed that frame semantic structure prediction can benefit from a deeper syntactic representation, in which the syntactic paths between a verb and its arguments are normalized. This reduces the variety of the syntactic realization of semantic roles, which we assessed by measuring a decrease of the entropy of the syntactic paths of a given role. We then showed that a FrameNet semantic parser can take advantage of this simpler syntax/semantic interface and reach better performance when switching from surface syntax to deep syntax.

Acknowledgments

This work was funded by the French National Research Agency (ASFALDA project ANR-12-CORD-023), and supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

References

Anne Abeillé and Nicolas Barrier. 2004. Enriching a french treebank. In *Proceedings of LREC 2004*, Lisbon, Portugal.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference*, pages 86–90, Montreal, Canada.

Anders Björkelund, Ozlem Cetinoglu, Richárd Farkas, Thomas Mueller, and Wolfgang Seeker. 2013. (re)ranking meets morphosyntax: State-of-the-art results from the SPMRL 2013 shared task. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 135–145, October.

A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Padó, and M. Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of LREC 2006*.

Aoife Cahill, Michael Burke, Ruth O'Donovan, Josef van Genabith, and Andy Way. 2004. Long-distance dependency resolution in automatically acquired wide-coverage pcfg-based lfg approximations. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Marie Candito and Djamé Seddah. 2012a. Effectively long-distance dependencies in French: annotation and parsing evaluation. In *Proceedings of TLT 11*.

Marie Candito and Djamé Seddah. 2012b. Le corpus sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Proceedings of TALN 2012 (in French)*, Grenoble, France, June.

Marie Candito, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot, and Laure Vieu. 2014a. Developing a French FrameNet: Methodology and first results. In *Proceedings of LREC 2014*, Reykjavik, Iceland, May.

Marie Candito, Guy Perrier, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah, and Éric De La Clergerie. 2014b. Deep Syntax Annotation of the Sequoia French Treebank. In *Proceedings of LREC 2014*, Reykjavik, Islande, May.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56.

Marie-Catherine De Marneffe and Christopher D Manning. 2008. Stanford typed dependencies manual. Technical report, Stanford University.

Marianne Djemaa, Marie Candito, Philippe Muller, and Laure Vieu. 2016. Corpus annotation within the french framenet: methodology and results. In *Proceedings of LREC 2016*, Portoroz, Slovenia, May.

Marianne Djemaa. 2014. Traitement framenet des constructions à attribut de l'objet. In *Proceedings of the 16e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2014)*, pages 13–24, Marseille, France, July.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Daniel Flickinger, Yi Zhang, and Valia Kordoni. 2012. Deepbank: A dynamically annotated treebank of the wall street journal. In *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories (TLT-11)*, pages 85–96, Lisbon, Portugal. Edições Colibri, Lisbon.

Association for Computing Machinery. 1983. In *Computing Reviews*, volume 24, pages 503–512.

Daniel Gildea and Julia Hockenmaier. 2003. Identifying semantic roles using combinatory categorial grammar. In Michael Collins and Mark Steedman, editors, *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 57–64.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288, September.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing prague czech-english dependency treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160, İstanbul, Turkey. ELRA, European Language Resources Association.

Jan Hajič, Jarmila Panevová, Eva Hajicová, Petr Sgall, Petr Pajas, Jan Štepánek, Jiří Havelka, Marie Mikulová, Zdenek Zabokrtskỳ, and Magda Ševcıková Razımová. 2006. Prague dependency treebank 2.0. *CD-ROM, Linguistic Data Consortium, LDC Catalog No.: LDC2006T01, Philadelphia*, 98.

Julia Hockenmaier and Mark Steedman. 2007. Ccgbank: A corpus of ccg derivations and dependency structures extracted from the penn treebank. *Comput. Linguist.*, 33(3):355–396, September.

Richard Johansson and Pierre Nugues. 2007. Lth: Semantic structure extraction using nonprojective dependency trees. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 227–230, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aravind K Joshi and Yves Schabes. 1997. Tree-adjoining grammars. In *Handbook of formal languages*, pages 69–123. Springer.

Igor Mel'čuk. 1988. *Dependency syntax: theory and practice*. SUNY press.

Alexis Nasr, Frédéric Béchet, and Jean-François Rey. 2010. Macaon : Une chaîne linguistique pour le traitement de graphes de mots. In *Traitement Automatique des Langues Naturelles - session de démonstrations*, Montréal.

Alexis Nasr, Frederic Bechet, Jean-Francois Rey, Benoit Favre, and Joseph Le Roux. 2011. Macaon: An nlp tool suite for processing word lattices. In *The 49th Annual Meeting of the Association for Computational Linguistics: demonstration session*.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Yi Zhang. 2014. Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 63–72, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March.

Corentin Ribeyre, Djamé Seddah, and Éric Villemonte De La Clergerie. 2012. A Linguistically-motivated 2-stage Tree to Graph Transformation. In *Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms*, Paris, France.

Corentin Ribeyre, Marie Candito, and Djamé Seddah. 2014. Semi-Automatic Deep Syntactic Annotations of the French Treebank. In *Proceedings of the 13th International Workshop on Treebanks and Linguistic Theories*, Tübingen, Germany, December.

Corentin Ribeyre, Eric Villemonte de la Clergerie, and Djamé Seddah. 2016. Accurate deep syntactic parsing of graphs: The case of french. In *Proceedings of LREC 2016*, may.

Corentin Ribeyre. 2016. *Data-driven methods for syntax-semantic interface*. Theses, Université Paris Diderot, January.

Grzegorz Rozenberg, editor. 1997. *Handbook of Graph Grammars and Computing by Graph Transformation: Volume I. Foundations*. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Natalie Schluter and Josef van Genabith. 2008. Treebank-based acquisition of lfg parsing resources for french. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may.

Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola Galletebeitia, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Joakim Nivre, Adam Przepiórkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński, Alina Wróblewska, and Eric Villemonte de la Clergerie. 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.

David Vickrey and Daphne Koller. 2008. Sentence simplification for semantic role labeling. In *Proceedings of ACL-08: HLT*, pages 344–352, Columbus, Ohio, June. Association for Computational Linguistics.