CLASSiC

Semantic Role Annotation of a French-English Corpus

Lonneke van der Plas, James Henderson, Paola Merlo

Distribution: Public

CLASSiC
Computational Learning in Adaptive Systems for Spoken Conversation
216594 Deliverable 6.2

Feb 2010

Project funded by the European Community
under the Seventh Framework Programme for
Research and Technological Development

The deliverable identification sheet is to be found on the reverse of this page.

Project ref. no.	216594
Project acronym	CLASSiC
Project full title	Computational Learning in Adaptive Systems for Spoken Conversation
Instrument	STREP
Thematic Priority	Cognitive Systems, Interaction, and Robotics
Start date / duration	01 March 2008 / 36 Months

Security	Public
Contractual date of delivery	M24 = March 2010
Actual date of delivery	Feb 2010
Deliverable number	6.2
Deliverable title	Semantic Role Annotation of a French-English Corpus
Type	Report
Status & version	Draft 1.0
Number of pages	53 (excluding front matter)
Contributing WP	2
WP/Task responsible	UNIGE
Other contributors	
Author(s)	Lonneke van der Plas, James Henderson, Paola Merlo
EC Project Officer	Philippe Gelin
Keywords	semantic role annotation corpus French English

The partners in CLASSiC are:

Heriot-Watt University	HWU
University of Cambridge	UCAM
University of Geneva	GENE
Ecole Superieure d'Electricite	SUPELEC
France Telecom/ Orange Labs	FT
University of Edinburgh HCRC	EDIN

For copies of reports, updates on project activities and other CLASSiC-related information, contact:

The CLASSiC Project Co-ordinator:
Dr. Oliver Lemon
School of Mathematical and Computer Sciences (MACS)
Heriot-Watt University
Edinburgh
EH14 4AS
United Kingdom
O.Lemon@hw.ac.uk
Phone +44 (131) 451 3782 - Fax +44 (0)131 451 3327

Copies of reports and other material can also be accessed via the project's administration homepage, `http://www.classic-project.org`

©2010, The Individual Authors.

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Contents

Executive summary

This document describes deliverable 6.2, due at month 24 of the CLASSIC project: the semantic role annotation of a French-English parallel corpus. The French-English corpus constructed is composed of three parts: first, 276-thousand English sentences automatically annotated using a syntactic-semantic parser trained on merged Penn Treebank/PropBank data, second, 276-thousand French sentences automatically annotated by using automatic transfer of semantic annotation from English to French, and third, 1000 French sentences manually annotated with semantic annotation. This report describes how these three corpora have been constructed and reports on the the quality of these resources as a result of several evaluations we performed.

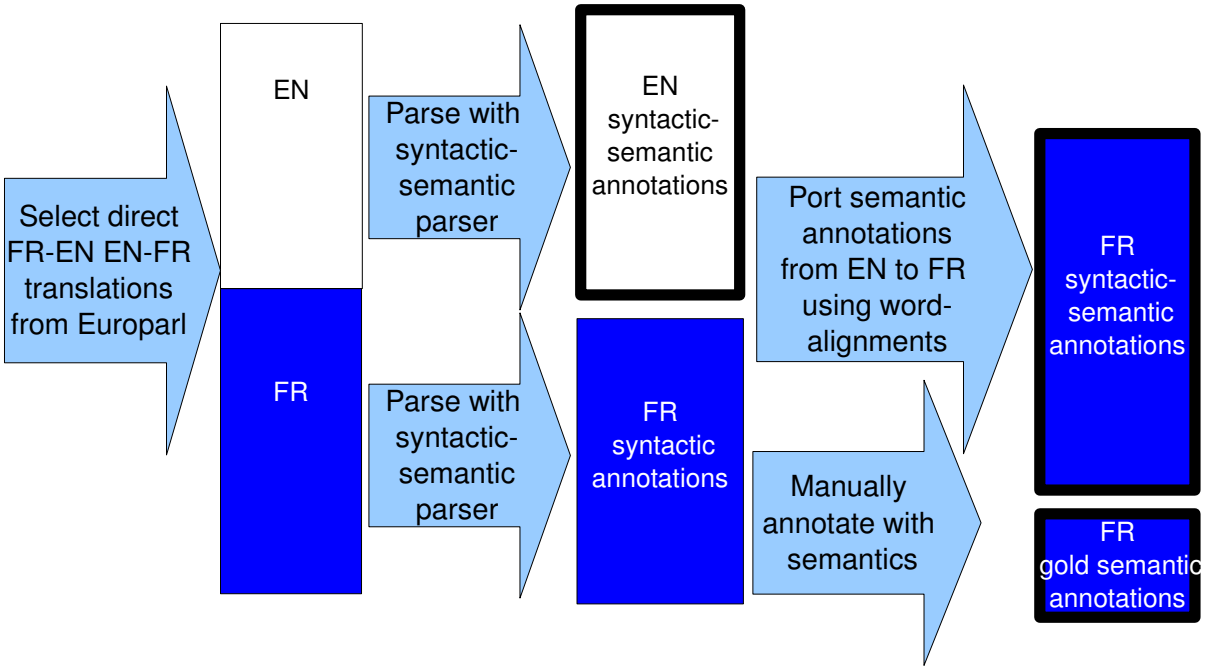

Figure 1: The three resources and how they are acquired

1 Overview

In this report we describe the annotation of an English-French parallel corpus with semantic information. The annotation effort is comprised of three subparts: the automatic annotation of an English corpus of 276-thousand sentences using a syntactic-semantic parser trained on merged Penn Treebank/PropBank data, the automatic annotation of a French corpus of 276-thousand sentences using automatic transfer of semantic annotation from English to French, and the manual annotation of 1000 French sentences. In Figure 1 we see the three resources marked with black solid lines.

In Deliverable 2.3 we described Prototype 2.3. The prototype is a syntactic-semantic parser for French. To be able to train a French syntactic-semantic parser, we need both syntactic and semantic annotations. In this document we describe the data used to train and to evaluate this parser.

In Section 3 we describe how we constructed the first resource: a large collection of English sentences annotated with syntactic-semantic annotations (upper part of Figure 1). We trained an English parser on the merged Penn Treebank/PropBank and parsed the English sentences of the parallel corpus.

The second resource we describe is the large collection of French sentences annotated with syntactic-semantic annotations (middle right of Figure 1). There are no semantic annotations available for French. In line with our previous work [1] we present a technique that consists in generating semantically annotated data from French automatically. We will transfer semantically annotated data from English to French

using parallel word-aligned corpora. This document describes the automatic semantic annotation of the French sentences using cross-lingual transfer in Section 4.

In Section 5 we describe the last resource: a small collection of hand-annotated sentences (bottom right of Figure 1). These sentences constitute a test set to evaluate our work on the transfer of semantic roles across languages.

Before we describe the annotations acquired in Sections 3, 4, and 5 we will describe how we selected a parallel corpus of English-French sentences and how we chose the annotation scheme in the next section: Section 2.

2 Corpora and annotation scheme

To transfer semantic annotation from English to French, we used the Europarl corpus [2]. Because translation shift are known to pose problems for the automatic projection of semantic roles across languages [3], and these are more likely to appear in indirect translations, we decided to select only those parallel sentences in Europarl for which we know they are direct translations from English to French, or from French to English. In a recent paper by Ozdowska and Way [4] the selection of direct translations proved beneficial for machine translation from French to English and vice versa. As usual practice in preprocessing for automatic alignment, the datasets were tokenised and lowercased and only sentence pairs corresponding to a 1-to-1 alignment with lengths ranging from 1 to 40 tokens on both French and English sides were considered. After this pre-processing step, we are left with 104-thousand sentence pairs with French as source and 172-thousand sentence pairs with English as source. In total that is 276-thousand parallel English-French sentences.

Then, we needed to decide on the framework for semantic annotation to choose, because there are several. In Merlo and Van der Plas [5], we proposed a task-independent, general method to analyse two annotation schemes: PropBank and VerbNet. The method is based on information-theoretic measures and comparison with attested linguistic generalisations, to evaluate how well semantic role inventories and annotations capture grammaticalised aspects of meaning. We show that VerbNet is more verb-specific and better able to generalise to new semantic roles, while PropBank, because of its relation to syntax, better captures some of the hierarchical and co-occurrence constraints among roles. We decided to use the PropBank framework, because we are working in a joint syntactic-semantic setting and PropBank is more appropriate when syntactic and semantic representations are calculated jointly.

3 The automatic (syntactic)-semantic annotation of English sentences[1]

For our syntactic-semantic parsing model we use the parser we developed in 2009 for English in the CONLL 2008 shared task. This syntactic-semantic dependency parser was extended in 2009 to several languages with very good results [6, 7]. The good performance on multiple languages with a single model is particularly important in our task because we need a semantic parser that will also work well for French.

[1] This section is taken from D2.3, but added for the sake of completeness

MODEL	CONLL MEASURES			CROSSING ARCS		
	Syntactic Labelled Acc.	Semantic F_1	Macro F_1	Semantics P	R	F_1
Johansson and Nugues 2008	89.3	81.6	85.5	67.0	44.5	53.5
UNIGE IJCAI 2009	87.5	76.1	81.8	62.1	29.4	39.9
UNIGE CONLL 2008	87.6	73.1	80.5	72.6	1.7	3.3

Table 1: Scores on the test set.

3.1 Joint syntactic-semantic model

The crucial intuition behind the treatment of both syntax and semantic in a single model is that these two levels of information are related but not identical. We propose a solution that uses a generative history-based model to predict the most likely derivation of a synchronous dependency parser for both syntactic and semantic dependencies. Our probabilistic model is based on Incremental Sigmoid Belief Networks (ISBNs), a recently proposed latent variable model for syntactic structure prediction, which has shown very good behaviour for both constituency [8] and dependency parsing [9]. The ability of ISBNs to induce their features automatically enables us to extend this architecture to learning a synchronous parse of syntax and semantics without modification of the main architecture. By solving the problem with synchronous parsing, a probabilistic model is learnt which maximises the joint probability of the syntactic and semantic dependencies and thereby guarantees that the output structure is globally coherent, while at the same time building the two structures separately.

We devise separate derivations $D_d^1,...,D_d^{m_d}$ and $D_s^1,...,D_s^{m_s}$ for the syntactic and semantic dependency structures, respectively, and then divide each derivation into the chunks between shifting each word onto the stack, $c_d^t = D_d^{b_d^t},...,D_d^{e_d^t}$ and $c_s^t = D_s^{b_s^t},...,D_s^{e_s^t}$, where $D_d^{b_d^t-1} = D_s^{b_s^t-1} = shift_{t-1}$ and $D_d^{e_d^t+1} = D_s^{e_s^t+1} = shift_t$. The actions of the synchronous derivations consist of quadruples $C^t = (c_d^t, switch, c_s^t, shift_t)$, where *switch* means switching from syntactic to semantic mode. This gives us the following joint probability model, where n is the number of words in the input.

$$\begin{aligned} P(T_d,T_s) &= P(C^1,\ldots,C^n) \\ &= \textstyle\prod_t P(C^t|C^1,\ldots,C^{t-1}) \end{aligned} \tag{1}$$

The probability of each synchronous derivation chunk C^t is the product of four factors, related to the syntactic level, the semantic level and the two synchronising steps.

$$\begin{aligned} &P(C^t|C^1,\ldots,C^{t-1}) = \\ &\quad P(c_d^t|C^1,\ldots,C^{t-1})\times \\ &\quad P(switch|c_d^t,C^1,\ldots,C^{t-1})\times \\ &\quad P(c_s^t|switch,c_d^t,C^1,\ldots,C^{t-1})\times \\ &\quad P(shift_t|c_d^t,c_s^t,C^1,\ldots,C^{t-1}) \end{aligned} \tag{2}$$

3.2 Results on syntactic-semantic parsing of English

This model results in the performance shown in Table 1. We report both the official CoNLL 2008 shared task numbers [6] and the performances reported in [10], which are greatly improved through a better

treatment of crossing arcs as described in [10]. We compare the two implementations to the best performing model for English. Macro F_1 is the overall score, and the other columns provide more detailed breakdowns, in particular precision and recall on the subset of semantic arcs which cross each other.

4 The automatic semantic annotation of French sentences[2]

In this section we describe a technique to generate semantically annotated data automatically in a language for which these data are not available (French) by relying on a large corpus of semantically unannotated data in this language aligned to a large automatically annotated corpus in a source language (English).

We formalise our method for artificial data generation as a three-component method, as proposed in Van der Plas et al. [1]. The first component of our method takes the form of a procedure for constructing a data point for the desired target resource (French corpus labelled with semantic roles) given a data point in the source resource (English syntactic-semantic parsed corpus). The second and third components define our distributional model. The source distribution describes the distribution of semantic roles and predicates in the source language, which we want to be reflected in the target corpus. The target distribution represents other aspects of the target corpus, in our case syntactic analyses of French.

We argue that these distributional components are crucial to automatically generating a corpus that is useful for training models. In our case, we use it to train a model of syntactic-semantic structure prediction. Previous work on the transfer of semantic annotations cross-lingually [3, 11] has not paid much attention to the distributions in the transferred annotations, focussing rather on achieving high recall and precision when transferring semantic annotations for carefully selected, but therefore possibly biased, test sets.

4.1 A three-component method for cross-lingual semantic transfer

Data-driven induction of semantic annotation based on parallel corpora is a well-defined and feasible task, and it has been argued to be particularly suitable to semantic role label annotation because cross-lingual parallelism improves as one moves to more abstract linguistic levels of representation. While Hwa et al. [12] and [13] find that direct syntactic dependency parallelism between English and Spanish concerns 37% of dependency links, Padó [3] reports an upper-bound mapping correspondence calculated on gold data of 88% F-measure for individual semantic roles, and 69% F-measure for whole scenario-like semantic frames. Recently, Wu and Fung [14] and [15] also show that semantic roles help in statistical machine translation, capitalising on a study of the correspondence between English and Chinese which indicates that 84% of roles transfer directly, for PropBank-style annotations. These results indicate high correspondence across languages at a shallow semantic level.

Based on these results, our initial method for cross-lingual transfer is based on a strong hypothesis of lexical semantic parallelism, the Direct Semantic Correspondence Assumption (DSCA). This hypothesis characterises correspondences across languages in terms of automatically generated word-alignments. Our second method relaxes this assumption for some problematic cases. We will then describe the distributional components for the source language and the target language respectively.

[2]The methods described here are partly described in D2.3, but we made several changes in the methodology.

Models of cross-lingual transfer

The first component of our artificial data generation method maps semantic role labels from English sentences to their French translations. We base our methods on a very strong mapping hypothesis, adapted from the Direct Correspondence Assumption for syntactic dependency trees by Hwa et al. [13].

> **Direct Semantic Correspondence Assumption (DSCA)**: Let a pair of sentences E and F that are (literal) translations of each other be given, with trees T_E and T_F. If vertices x_E and $y_E \in T_E$ are aligned with vertices x_F and $y_F \in T_F$ and if semantic relationship $R(x_E, y_E)$ holds in T_E, then $R(x_F, y_F)$ holds in T_F.

We apply this hypothesis to semantic role dependency graphs, where the vertices of the graph are words and the edges are labelled semantic role dependencies. We also apply it to predicate sense labels, which refer to individual words.

Following the Direct Semantic Correspondence Assumption, our first method for transferring semantic annotation from English to French is based directly on automatically generated word-alignments in the following way:

> **Direct Semantic Transfer (DST)** For any pair of sentences E and F that are translations of each other, we transfer the semantic relationship $R(x_E, y_E)$ to $R(x_F, y_F)$ if and only if there exists a word-alignment between x_E and x_F and between y_E and y_F, and we transfer the semantic property $P(x_E)$ to $P(x_F)$ if and only if there exists a word-alignment between x_E and x_F.

The relationships which we transfer are semantic role dependencies and the properties are predicate senses. The word-alignments are those output by word alignment methods developed for statistical machine translation (e.g. the intersective alignments of [16]).

Upon inspection of the training data, we found one main case in which lexical correspondence is violated. Semantic role dependencies often specify closed-class words as arguments, and this syntactic mechanism is not always mirrored in the translation. In particular, French does not use complementisers or prepositions in some frequent cases where English does. For example *I want to go* is translated as *Je veux [] partir* and *I will sing on Thursday* corresponds to *Je chanterai [] jeudi*. In these cases, automatic word-alignment aligns the open-class dependents of the complementisers or prepositions, leaving the complementisers and prepositions unaligned, and their semantic roles un-transferred.

To address this problem, our second method for transferring semantic annotation uses the following transfer rule, which we call the "D-rule":

> For any pair of translated sentences E and F and a semantic relationship $R(x_E, y_E)$ in E, if there exists a word-alignment between predicates x_E and x_F but not between roles y_E and any y_F, and if the PoS of y_E is 'IN' or 'TO', then we find the dependent y'_E of y_E and if there exists a word-alignment between y'_E and some y'_F, we transfer the semantic relationship $R(x_E, y_E)$ to $R(x_F, y'_F)$.

The two distributional models

Because we want to use the annotated corpus generated by our cross-lingual transfer method to train parsers, we want the resulting corpus not only to have accurate annotations, but also to have accurate distributional properties.

The second component of our artificial data generation method specifies the distributional properties of the source corpus which should be reflected in the target corpus, namely the distribution over semantic role structures. Thus, we assume that the distribution of semantic role structures for English is also applicable to French. We sample from this distribution by starting with a parallel corpus of English-French sentence pairs, and annotating the English sentences with a joint syntactic-semantic parser which has been trained on a separate hand-annotated corpus of English.

To try to ensure that the distributional properties of English are preserved after the transfer of annotation to French, we need to compensate for the incompleteness of the transfer method. As reported by Hwa et al. [13], the DSCA is a strong hypothesis that is useful to trigger a process of role projection, but will not work correctly for several cases. Also, the general notion of semantic correspondence only applies to those sentence pairs in a parallel corpus that can be loosely characterised as literal translations, translations where the lexical, structural and propositional correspondence is direct. For many sentence pairs, the transfer of annotation under the Direct Semantic Transfer approach is incomplete, for example, because for a certain English argument there is no corresponding argument in the French sentence. It therefore does not result in the same semantic structure on the source and target sentences. These incomplete annotations make the resulting target distribution systematically different from the source distribution of semantic structures.

To compensate for this incompleteness, we investigate two filters which select sentences for inclusion in the target corpus. We compare the results to the unfiltered case. The stricter filter, which we call the "complete filter", requires that all predicates with all their roles are transferred to the target sentence. This corresponds to one possible definition of literal translation, although note that we do not require isomorphism of the syntactic structure, as is sometimes required for literal translations. In addition, this filter removes sentences which may be literal translations, but for which errors or mismatches in the automatic word alignment or semantic roles lead to incompletely transfered predicates or roles.

Because the complete filter removes many sentences, we also experiment with a less strict "partial filter". This filter selects those sentences for which all predicates are transferred, but does not require that all the roles are transferred.

The third component of our artificial data generation method specifies the distributional properties to be taken from the target domain. As well as the semantic structures transferred from English, our target corpus is supposed to include syntactic dependency structures. The distribution of these structures are taken from French data. We assume that the distribution of sentences in the parallel corpus is representative of French. These sentences are automatically annotated with syntactic dependency structures using a syntactic parser which has been trained on a separate hand-annotated corpus of French.

As an example of how information about the joint distribution of the syntax and semantics in the target distribution can be exploited, we investigate one additional constraint on the target distribution. This constraint comes from the knowledge that the gold semantic role annotations are only given for verbal predicates. It filters any sentences where a predicate has been transferred to a word which the syntactic parser has not tagged as a verb. We call this constraint the "V-rule". Note that the English sentences already only include verbal predicates, since we are using PropBank. However, verbs are sometimes translated into other parts of speech, as in deverbal nominalisations.

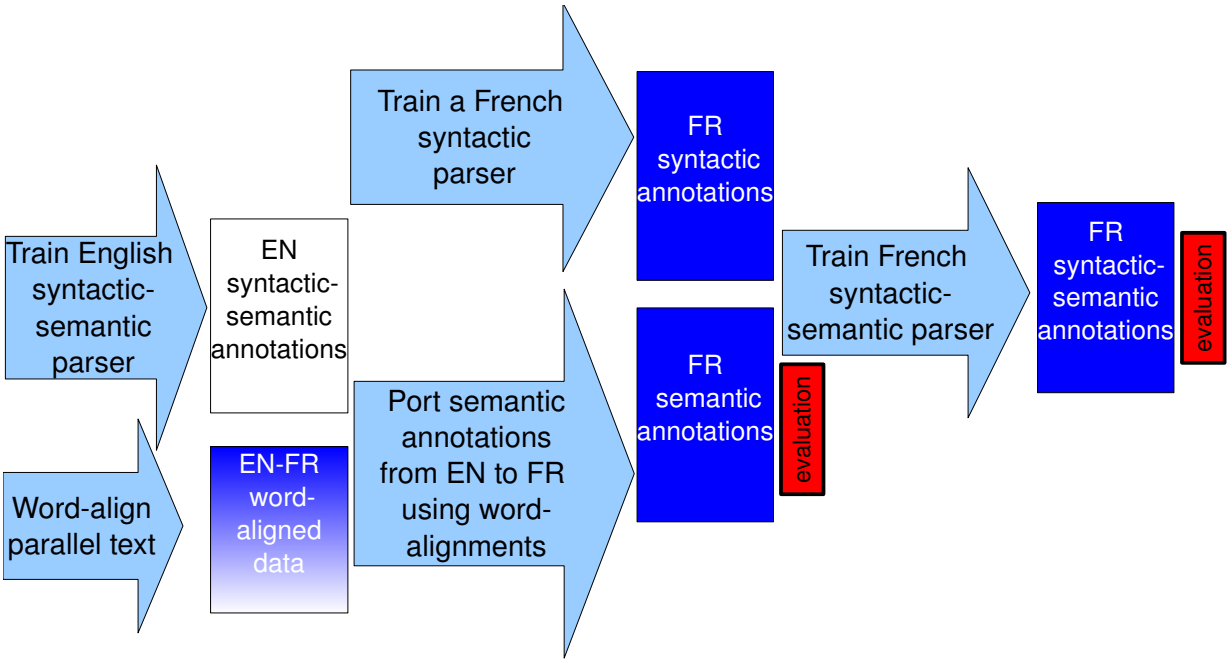

Figure 2: Overview of the evaluation pipeline

4.2 Experiments

We evaluate our methods for artificial data generation using cross-lingual transfer of semantic roles in two ways. First, we evaluate the resulting annotation directly by comparison to a hand-annotated test set. Second, we evaluate the resulting corpus in a task-dependent way, as to its usefulness as data for training a syntactic-semantic parser. This parser is then evaluated on a separate hand-annotated test set. In this subsection, we describe the existing tools and resources used to perform these two evaluations.

The specific system of artificial data generation and use which we evaluate is illustrated in Figure 2. Several steps are required to build the French syntactic-semantic parser using resources in English. French components and resources are colour-coded in dark blue, and English components and resources in white.

We can see that we need both French syntactic annotation and French semantic annotation to train a syntactic-semantic parser for French. The syntactic annotation is acquired through the upper path, based on a parser trained on the dependency version of the French Paris treebank, as described Deliverable 2.3. To compute French semantic annotations we have to follow the path indicated in the lower part of the figure. The English syntactic-semantic dependency parser described below is used to annotate the English side of the parallel corpus, also discussed below. We word-align the English sentences to the French sentences automatically using GIZA++ [16]. The intersection of the word-alignments found in the source to target alignment and in the target to source alignment is called an intersective alignment. We have chosen to include only intersective alignments, because they have been reported to give the best results in

		Predicates						Arguments (given predicate)					
		Labelled			Unlabelled			Labelled			Unlabelled		
		Prec	Rec	F	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F
1	No Filter	67	54	60	79	63	70	60	48	53	73	58	64
2	Partial filter	68	63	**65**	80	74	**77**	61	48	54	75	59	66
3	Complete filter	67	55	61	80	66	73	62	53	**57**	76	65	**70**
4	No Filter+V-rule+D-rule	79	54	64	93	63	75	58	49	53	70	59	64
5	Partial filter+V-rule+D-rule	80	63	**70**	94	74	**83**	58	50	54	72	62	67
6	Complete filter+V-rule+D-rule	78	58	67	94	70	80	59	51	**55**	73	63	**68**

Table 2: Percent recall, precision, and F-measure for predicates and for arguments given the predicate, for the three **transfer** models with/without V-rule and D-rule

machine translation and are therefore likely to better support our cross-linguistic task. We then transfer the semantic annotation from the English side to the French side, according to the model described above. Finally, the syntactic and semantic French annotations are merged and the same domain-independent parsing model used to parse the English is trained to parse this merged syntactic-semantic French corpus. Both the sub-component which leads to the semantic annotation and the entire pipeline are evaluated as can be seen in in Figure 2. We report the results of these separate evaluations in Subsection 4.3.

4.3 Results

We evaluate the transfer step and the final parser output separately on two separate test sets. Both data sets are constructed following the same methodology described in detail in the next section: Section 5[3].

Evaluation of transferred data

Table 2 shows the results for transferring semantic annotation from English onto French sentences as evaluated on the test set of 100 sentences annotated manually. The first three lines refer to simple filters and the second group of three lines refers to filters augmented with improved alignments (D-rule) and restriction to verbal predicates (V-rule). The first set of columns of results (labelled predicates) reports labelling and identification of predicates and the latter set of columns reports labelling and identification of arguments for the predicates that are identified, respectively.

Note that our evaluations are strict. Despite the fact that we are using English frame files, we require exact match for labelling predicates. As explained in more detail in Subsection 5.3 the annotators used the English frame files to attach a predicate label on the French verbs. In case of ambiguity the annotator might choose a label that is not found in the English translation. For example, the French verb 'reussir à' can be translated as 'manage' or 'succeed', so the annotator could opt for a label that is not found in the

[3]At the time these experiments were run, only 220 sentences had been annotated. These were split into 100 sentences to evaluate the transfer step and 120 to evaluate the output of the final parser.

		Predicates						Arguments (given predicate)					
		Labelled			Unlabelled			Labelled			Unlabelled		
		Prec	Rec	F	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F
1	No Filter	66	23	34	89	30	45	78	52	62	86	58	69
2	Partial filter+V-rule+D-rule	71	29	**41**	94	38	**54**	68	56	61	80	66	**72**
3	Complete filter+V-rule+D-rule	76	24	36	92	29	44	70	57	**63**	80	66	**72**

Table 3: Percent recall, precision, and F-measure for predicates and for arguments given the predicate, for three **parsing** models

English sentence and therefore not part of the automatically transferred data.

The results show that the transfer method without any filters is the worst, especially for recall. The partial filter improves predicate recall, while the complete filter improves argument recall once the predicate is identified. We expect the partial filter to improve the recall on predicates, since it requires that all predicates from the English sentences are transferred. Therefore, the chances of finding unidentified predicates in the resulting French annotations is lower. The complete filter, which requires complete propositions (the predicate and all the accompanying arguments), improves the recall on arguments, as expected.

Adding the D-rule and V-rule particularly helps in getting better precision for predicates. This is mainly due to the V-rule that requires that predicates are verbs.

To compare with previous work we have to first note that the languages and annotations adopted are not directly comparable. However, results are not markedly disparate. Recall that our best result for labelled predicate identification is an F-measure of 70%, and for argument labelling given the predicate is an F-measure of 54%. For German, Padó [3] reports a 56% F-measure on transferring FrameNet roles, knowing the predicate, from an automatically parsed and semantically annotated English corpus. For French, Padó and Pitel [17] report a best result of 57% F-measure for argument labelling given the predicate. For Italian, Basili et al. [11] report 42% recall in identifying predicates and 73% global recall of identifying predicates and roles given these predicates.

Evaluation of the parser trained on transferred data

To evaluate the usefulness of our method for artificial data generation in the task of training a parser, we compare the semantic annotations output by the trained parser to the hand-annotated dataset. This evaluation tests the entire pipeline of our system, starting with the corpus used to train the English syntactic-semantic parser and the corpus used to train the French syntactic parser and ending with the annotations produced by the French syntactic-semantic parser.

The results on syntactic-semantic parsing of dependencies for French are shown in Table 3. We see that the parsing results follow the same pattern as the transfer results. Filtering to remove incompletely transferred annotations, the D-rule, and the V-rule result in a more accurate parser. There is no significant difference in role identification and labelling between the partial filter and the complete filter when using the V-rule and D-rule. The partial filter is better than the complete filter in predicate identification recall.

Unlike for the transfer results, the predicate identification recall (Predicates–Unlabelled–Rec in Table 3) is very low for the output of the two best trained parsers. In other words, the parsers do not identify very many words as predicates. However, those which they do identify are very likely to be correct, with a predicate identification precision above 90%. This reluctance to specify predicates is probably in part due to the need to identify a specific sense for each predicate, which are lexically specific and therefore sparse. Also, a more detailed breakdown of the errors indicates that this problem is largely due to difficulty in identifying past participles (VPP tag) as predicates, which also have a high error rate (19%) in the Part-of-Speech tagging. Recall errors are known to be a particular weakness of the parsing method we use [6], so alternative parsing methods may address this imbalance between precision and recall.

It should be borne in mind that many current results on joint syntactic parsing and semantic role labelling (CoNLL 2009 shared task results, for example) are calculated on data where the predicates are given, eliminating the need to do predicate identification. It is therefore more appropriate to compare these previous results to those in the right half of Table 3.

Once predicates are identified, the performance on arguments is quite good, reaching the same level as some results for parsers trained on hand-annotated data. Interestingly, across all measures, these results actually exceed those for the data that the parsers were trained on (see Table 2). The main piece of information which the parsers use but the transfer method did not use is the syntactic annotation, suggesting that training a joint syntactic-semantic parser on the output of semantic transfer can lead to correcting transfer errors.

There have only been a few authors that trained systems on the annotations acquired through cross-lingual transfer of annotations. Hwa et al. [13] bootstrap a Spanish and a Chinese dependency parser from projected annotations. The English dependency structures are projected on the non-English side using the Direct Transfer algorithm we described in Section 4.1, refined by some hand-built language specific post-projection rules, followed by an aggressive filtering strategy to prune projected trees of poor quality. The Spanish parser reaches performance that is comparable to a state-of-the-art commercial parser, while the Chinese parser reaches a performance equal to that of a parser trained on 3000 sentences annotated with noise-free dependency structures.

With respect to the task of training a semantic role labeller on projected annotations, Johansson and Nugues [18] have trained a FrameNet-based semantic role labeller for Swedish on annotations transferred cross-lingually from English parallel data. They apply the semantic role labelling system to the English part of the Europarl parallel corpus and then project the bracketing and frame elements to the Swedish side using word alignment. Evaluation is done on 150 translated example sentences. They report 55% F-measure for argument labelling given the frame. The authors argue that results are above the baseline, but not as high as for English systems due to the quality of the data, among other factors. Although not directly comparable, our best result for argument labelling given the predicate is 63% F-measure.

5 The manual annotation of 1000 French sentences

To evaluate the transferred information described in Section 4 we need gold annotation. We set up an annotation effort that is comprised of several stages, as we will describe in Subsection 5.4. We first describe how we selected the test sentences and how we chose the annotation tool. In addition, we discuss the fact that we are annotating French text, for which there are fewer resources available, and how we solved that problem.

5.1 Selection of test sentences

For our test set, we extracted 1000 French sentences from the set of French-English parallel Europarl sentences described above in Section 2. These sentences were selected randomly without any constraints on the semantic parallelism of the sentences, unlike much previous work. They were parsed using the syntactic dependency parser trained on French syntactic data and described in Deliverable 2.3.

5.2 Selection of the annotation tool

For the annotation of sentences we need an annotation tool that shows the syntactic dependency tree annotations present in the parsed French data and that allows the user to add semantic arcs and labels. We decided to use the Tree Editor (TrEd, [19]) because it has all needed functionalities and has a good interface. The tool is freely available and support is provided by the developers.

5.3 Using English frame files to annotate French verbs

Frame	Semantic roles
pay.01	A0: payer or buyer A1: money or attention A2: person being paid, destination of attention A3: commodity, paid for what
pay.02 (pay off)	A0: payer A1: debt A2: owed to whom, person paid
pay.03 (pay out)	A0: payer or buyer A1: money or attention A2: person being paid, destination of attention A3: commodity, paid for what
pay.04	A1: thing succeeding or working out
pay.05 (pay off)	A1: thing succeeding or working out
pay.06 (pay down)	A0: payer A1: debt

Table 4: The PropBank lexicon entry for the verb *pay*.

(3) [$_{A0}$ The Latin American nation] has [$_{REL-PAY.01}$ paid] [$_{A1}$ very little] [$_{A3}$ on its debt] [$_{AM-TMP}$ since early last year].

The annotations in the PropBank corpus consist of predicate-argument structures. Predicates (verbs) are marked with labels that specify the sense of the verb in the particular sentence. (See the label for the verb *paid* in (3) and the verb senses of *pay* in Table 4.) Verb senses are used as labels for predicates since each verb sense introduces a different set of semantic roles.

Arguments are marked with the labels A0 to A5. The labels A0 and A1 have approximately the same value with all verbs. They are used to mark instances of typical AGENT (A0) and PATIENT (A1). The value of other numbers varies across verbs. The numbers are assigned according to the roles' prominence in the sentence. More prominent are the roles that are more closely related to the verb. In the appendix, the annotation of arguments is explained in more detail.

The argument labels for each verb are specified in the PropBank frame files. The frame files provide verb-specific description of all possible semantic roles and illustrate these roles with examples as shown for the verb *paid* in 3) and the verb senses of *pay* in Table 4. Annotators need to look up each verb in the frame files to be able to label it with the right verb sense and to be able to allocate the arguments consistently according to the framefiles.

There are no frame files for French. Constructing a complete set of frame files for French would be too time-consuming. We decided to annotate the sentences in a way that is inspired by the work of Monachesi et al. [20]. In their annotation scheme for Dutch, the English PropBank labels and the PropBank frame files were used to annotate both the verb senses and the role labels. Since semantic structures are generally preserved across languages, we can assume that French verbs introduce the same frames as English verbs.

Because there are often several potential translation of a French predicate in an English verb sense, annotators were asked to carefully select the verb sense that is closest to the meaning of the French predicate *and* that has a set of semantic roles most similar to the set found in the French sentence. In some cases it is not possible to select one best verb sense label, for example in case of synonymy in the English verb senses. Also, since the PropBank lexicon is not an exhaustive list of all possible French frames, it can happen that no corresponding frame can be found for a French predicate. In this case, annotators were asked to mark the French predicate with the label DUMMY and mark arguments according to their own intuition. For the annotators, choosing the predicate sense was the hardest part of the annotation. We will see in Section 5.5 that this is where most of the initial disagreements between annotators were introduced.

5.4 The annotation process

In the previous sections we described the most important decisions we took regarding the annotation framework. More details on the annotation can be found in the appendix entitled 'Guidelines for the semantic annotation of French sentences from Europarl'. We will now describe the several phases in the annotation process.

The training procedure described in Figure 3 is inspired by the methodology indicated in Padó [3]. The sentences were annotated manually by four annotators for the training and the calibration phase. The remaining 1000 sentences are annotated by one annotator (out of those four) with very good proficiency in both French and English.

During the training phase, all 4 annotators, among which were two specialists of semantic role annotation, annotated 10 sentences together. After this initial training (TrainingA) 3 annotators annotated the same

- Training phase
  - TrainingA (10 sentences) all annotators together
  - TrainingB and C (15 + 15 sentences) individually
  - Measure inter-annotator agreement TrainingB and C
  - Discussion of difficulties encountered for TrainingB
  - Correct TrainingB and TrainingC individually
  - Measure inter-annotator agreement TrainingB and TrainingC
  - Discussion of difficulties encountered for TrainingC
  - Correct TrainingC individually
  - Measure inter-annotator agreement
  - Adjusting guidelines (if needed)
- Calibration phase
  - 100 sentences individually
  - Measure inter-annotator agreement
  - Adjusting guidelines (if needed)
- Main annotation phase
  - 1000 sentences

Figure 3: The annotation process

two sets of 15 sentences (Training B and C). Inter-annotator agreement was measured and difficulties were discussed for the first 15 sentences. After the discussions, annotators went back to the sentences and corrected mistakes in sentences 1 to 15 and tried to apply the conclusions from the discussions to the last 15 sentences. Inter-annotator agreement was measured again after these corrections both for the sentences discussed (TrainingB) and those not discussed yet (TrainingC). After that the sentences of TrainingC were discussed and corrected and the inter-annotator agreement was again measured. The guidelines were adjusted.

During the calibration phase, the main annotator annotated 100 sentences, the other three annotators each annotated one third of the 100 sentences. Again inter-annotator agreement was measured before and after consensus. There was no need to adjust the guidelines. After that, the main annotator annotated the 1000 sentences in the main annotation phase. The inter-annotator agreement in all stages is discussed in Subsection 5.5.

		Predicates		Arguments	
		Labelled F	Unlabelled F	Labelled F	Unlabelled F
1	TrainingB+C	60	85	62	75
2	TrainingB(adjud.)	91	95	89	93
3	TrainingC	61	88	67	79
4	TrainingB+C(adjud.)	97	97	91	95
5	Calibration	76	93	78	82

Table 5: Percent inter-annotator agreement (F-measure) for labelled/unlabelled predicates and for labelled/unlabelled arguments

5.5 Evaluation of annotations

To get an idea of the quality of the manual annotations in several stages of the annotation process, we measured the average F-measure of all pairs of annotators after each phase.

In Table 5 we can see the agreement between annotators at several stages of the annotation process. The first set of columns of results (labelled 'Predicates') reports labelling and identification of predicates and the latter set of columns reports labelling and identification of arguments, respectively.

The first row shows agreement on the 30 sentences of TrainingB, that is after the annotators have seen 10 sentences together. The second row shows their results on the first 15 sentences after they have seen the 15 sentences together and reached a consensus. The third row shows agreement for the last 15 sentences after a consensus has been reached for the first 15 sentences. The fourth row shows agreement after all 30 sentences have been adjudicated by all annotators. Row five shows the results of inter-annotator agreement on the Calibration sentences (100).

The difference between the first row and the second row shows that there were many differences between annotators that they could reach a consensus on. After adjudication, the scores are all around 90%. The score for labelled argument is the lowest. This is probably due to the fact that the annotators focused on the differences in predicate labelling in their discussions, a task that was deemed rather difficult, and not so much on argument labelling. The next row shows how they are able to project what they have learnt from TrainingB onto the 15 sentences of TrainingC. The improvement are not as large as we would have thought, though they improve on all identification and labelling tasks by a few points. Row 4 shows the inter-annotator agreement on TrainingB and TrainingC combined after adjudication. Again, we can see that the annotators are able to reach a consensus on almost all annotations. Row 5 shows that the agreement in the calibration phase made a large jump from the last training phase (row 3). It seems that the annotators, who seemed at first not to be able to learn from the training phase (small difference between row 1 and row 3), have in fact learnt a lot by the time they started the calibration phase. This might in part be due to the fact that the guidelines were adjusted by the end of the training phase, but seems largely due to the fact that the annotators are getting more and more acquainted to the task and the software. In the training phase some mistakes were introduced by using the annotation tool incorrectly.

The acquaintance with the task and tools is reflected in the annotation times reported. In training phase B annotators spent 1,5 hours up to two hours to annotate 10 sentences, while by the beginning of the

calibration phase they usually finished 10 sentences within the hour.

Recall from the previous section (Subsection 5.3) that annotators need to use the English PropBank frame files to annotate French verbs. As expected, the task of labelling predicates is the most difficult and results in the lowest agreement scores overall. On the other hand, predicate identification without labelling has high agreement (93% F-measure).

6 Conclusions

This document describes the semantic annotation of a French-English parallel corpus in three subparts: the automatic annotation of an English corpus of 276-thousand sentences using a syntactic-semantic parser trained on merged Penn Treebank/PropBank data, the automatic annotation of a French corpus of 276-thousand sentences sentences by using automatic transfer of semantic annotation from English to French, and the manual annotation of 1000 French sentences with semantic annotation. The quality of all three annotated corpora has been evaluated in separate evaluation frameworks. The English parser used to parse the corpus of 276-thousand sentences is a state-of-the-art parser, with very good results on many languages. The automatically generated annotation for the French sentences is comparable to previous work. The manual annotations for the 1000 French sentences resulted in good inter-annotator scores, which leads us to believe that the quality of the annotations is of high quality.

In future work, we will keep improving the method for automatic transfer of semantic annotations from English to French parallel sentences. In particular, we would like the transfer to exploit the syntactic annotation on the French side. This information seems to have allowed a parser trained on the joint syntactic-semantic data to improve the semantic annotation. Also, syntactic information could be used in a previous step to improve the word alignments, especially in terms of recall. Adding syntactic dependency features into a syntax-enhanced alignment model has been shown to yield a 5.57% relative decrease in alignment error rate [21].

The results of this research also have implications for models of Machine Translation (MT). The degree of parallelism found between translated sentences at the level of semantic roles suggests that this may be a useful transfer level for MT. It may be possible to extend current work on syntax-based MT to some form of semantic-role-based MT.

Bibliography

[1] L. van der Plas, J. Henderson, and P. Merlo. Domain adaptation with artificial data for semantic parsing of speech. In *Proceedings of NAACL*, 2009.

[2] P. Koehn. Europarl: A multilingual corpus for evaluation of machine translation. 2003.

[3] S. Padó. *Cross-lingual Annotation Projection Models for Role-Semantic Information*. PhD thesis, Saarland University, 2007.

[4] S. Ozdowska and A. Way. Optimal bilingual data for french-english pb-smt. In *Proceedings of the 13th Annual Conference of the European Association for Machine Translation (EAMT'09)*, 2009.

[5] P. Merlo and L. van der Plas. Abstraction and generalisation in semantic role labels: PropBank, VerbNet or both? In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2009.

[6] J. Henderson, P. Merlo, G. Musillo, and I. Titov. A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of CONLL 2008*, pages 178–182, 2008.

[7] A. Gesmundo, J. Henderson, P. Merlo, and I. Titov. A latent variable model of synchronous syntactic-semantic parsing for multiple languages. In *Proceedings of the thirteenth conference on computational natural language learning (CoNLL)*, 2009.

[8] I. Titov and J. Henderson. Constituent parsing with Incremental Sigmoid Belief Networks. In *Procs. ACL*, 2007.

[9] I. Titov and J. Henderson. A latent variable model for generative dependency parsing. In *Proceedings of the International Conference on Parsing Technologies (IWPT-07).*, 2007.

[10] I. Titov, J. Henderson, P. Merlo, and G. Musillo. Online graph planarisation for synchronous parsing of semantic and syntactic dependencies. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09),*, 2009.

[11] R. Basili, D. De Cao, D. Croce, B. Coppola, and A. Moschitti. *Computational Linguistics and Intelligent Text Processing*, chapter Cross-Language Frame Semantics Transfer in Bilingual Corpora, pages 332–345. Springer Berlin / Heidelberg, 2009.

[12] R. Hwa, P. Resnik, A. Weinberg, and O. Kolak. Evaluating translational correspondence using annotation projection. In *Proceedings of the 40th Annual Meeting of the ACL*, 2002.

[13] R. Hwa, P. Resnik, A.Weinberg, C. Cabezas, and O. Kolak. Bootstrapping parsers via syntactic projection accross parallel texts. *Natural language engineering*, 11:311–325, 2005.

[14] D. Wu and P. Fung. Can semantic role labeling improve SMT? In *Proceedings of the Annual Conference of European Association of Machine Translation*, 2009.

[15] D. Wu and P. Fung. Semantic roles for smt: A hybrid two-pass model. In *Proceedings of the Joint Conference of the North American Chapter of ACL/Human Language Technology*, 2009.

[16] F.J. Och. GIZA++: Training of statistical translation models. Available from `http://www.isi.edu/~och/GIZA++.html`, 2003.

[17] S. Padó and G. Pitel. Annotation précise du fran cais en sémantique de rôles par projection cross-linguistique. In *Proceedings of TALN*, 2007.

[18] R. Johansson and P. Nugues. A FrameNet-based semantic role labeler for Swedish. In *Proceedings of the annual Meeting of the Association for Computational Linguistics (ACL)*, 2006.

[19] P. Pajas and J. Štěpánek. Recent advances in a feature-rich framework for treebank annotation. In *Proceedings of the 22nd International Conference on Computational Linguistics*, 2008.

[20] P. Monachesi, G. Stevens, and J. Trapman. Adding semantic role annotation to a corpus of written dutch. In *Proceedings of the Linguistic Annotation Workshop (LAW)*, 2007.

[21] Y. Ma, S. Ozdowska, Y. Sun, and A. Way. Improving word alignment using syntactic dependencies. In *Proceedings of the Second Workshop on Syntax and Structure in Statistical Translation (SSST-2)*, 2008.

1 Appendix

Guidelines for the semantic annotation of French sentences from Europarl*

Lonneke van der Plas, Tanja Samardzic, and Paola Merlo

March 10, 2010

Contents

*Sections 2.1 and 2.2 are adapted from the English PropBank guidelines (Babko-Malaya, 2005).

1 Introduction

The aim of this project is to create a corpus of French sentences with annotated semantic structure, which can be used in developing systems for automatic semantic analysis.

Semantic structure is analysed at the level of a clause. It involves identifying and labelling two kinds of elements: **predicates** and their **arguments**. The predicate is the semantic head of a clause. It is a word whose meaning is relational, typically a verb. Arguments are words or phrases which denote the elements which are related by the predicate. Both predicates and arguments can be assigned labels that describe their meaning. For example, the verb *washed* is the predicate in (1a). Its meaning is described by the label GROOMING. This predicate takes two arguments: an AGENT, the element that performs the activity (*Mary*), and a THEME, the element that undergoes the activity (*the car*). The labels that are assigned to the arguments are called **semantic roles**. Different predicates introduce different set of semantic roles in a clause. The verb in (1a), for example, introduces two semantic roles (AGENT and THEME), and the verb in (1b) introduces only one (THEME). Note also the different semantic roles introduced by the verbs in (1a) and (2a).

Semantic analysis is similar to syntactic analysis in some respects. Semantic heads are usually syntactic heads too, and arguments are syntactic dependants. However, there are important differences between them. Constituents with different syntactic functions can have the same semantic role. For example, the constituent *the car* is the syntactic object in (1a) and the subject in (1b). However, this constituent denotes the element that undergoes the event in both sentences, which is why it bears the same semantic role in both cases. On the other hand, constituents with the same syntactic function can bear different semantic roles. For example, syntactic subjects of (1a) and (1b), *Mary* and *the car* respectively, have different semantic roles, because the kind of meaning of these sentences is different (sentence (1a) describes an activity, while sentence (1b) describes a change of state).

(1) a [$_{AGENT}$ Mary] [$_{GROOMING}$ washed] [$_{THEME}$ the car].

b [$_{THEME}$ The car] [$_{DEPARTING}$ disappeared].

Semantic structure is better preserved across languages than syntactic structure. Sentences that have different syntactic representation in different languages can still have the same semantic representation, as it can be seen in

(2a-b). Even though *Mary* is the subject of the English sentence (2a) and *Marie* is the indirect object of the French sentence (2b), they bear the same semantic role in both cases. The same applies to *the idea* and *l'idée*.

(2) a [$_{EXPERIENCER}$ Mary] liked [$_{CONTENT}$ the idea]. (English)

b *[$_{CONTENT}$ L'idée]* *a* *plu* *[$_{EXPERIENCER}$* *à* *Marie].*
The idea AUX liked to Mary
(French)

The range of labels that can be assigned to predicates and arguments depend on the theoretical framework that is adopted. In this project, we will use the framework that has been developed for annotating the **Proposition Bank** (**PropBank**, Palmer et al., 2005).

2 The PropBank annotation framework

PropBank is a linguistic resource that contains information on the semantic structure of sentences. It consists of a one-million-word **corpus** of naturally occurring sentences with manually annotated semantic structure and a lexicon (the PropBank **Frame files**) listing all the predicates (verbs) that can be found in the annotated sentences with the sets of semantic roles that they introduce.

In the annotated sentences of the **PropBank corpus**, predicates (verbs) are marked with labels that consist of two parts. The first part, REL, stands for *relation* and it is the same for all the verbs. The second part specifies the sense of the verb in the particular sentence. (See the label for the verb *paid* in (3) and the verb senses of *pay* in Table 1.) Verb senses are used as labels for predicates since each verb sense introduces a different conceptual **frame** involving a different **role set**.

Arguments are marked with the labels A0 to A5. The numbered labels represent semantic roles that are more general and more abstract than the commonly used roles such as those in (1) and (2). The labels A0 and A1 have approximately the same value with all verbs. They are used to mark instances of typical AGENT (A0) and PATIENT (A1) (see previous section). The value of other numbers varies across verbs. It depends on the meaning of the verb, on the type of the constituent that they are attached to, and on the number of roles present in a particular sentence. A3, for example, can mark the PURPOSE, as it is the case in (3), or it can mark a DIRECTION or some other role with other verbs. The numbers are assigned according to the roles' prominence in the sentence. More prominent are the roles that are more closely related to the verb. In Subsection 2.1 the annotation of arguments is explained in more detail.

Modifiers are annotated in PropBank with the label AM. This label can have different extensions depending on the semantic type of the constituent: LOC denoting *locatives*, CAU for *cause*, EXT for *extent*, TMP for *temporal markers*, DIS for *discourse connectives*, PNC for *purpose*, PRD for *secondary predication*, REC for *reciprocals*, ADV for *adverbials*, MNR for *manner*, DIR for *directionals*, NEG for *negation marker*, and MOD for *modal verb*. The last two labels do not mark modifiers, but they are added to the set of labels for semantic annotation nevertheless, so that all the constituents that surround the verb could have a semantic label. The semantic roles of modifiers are more specific than the roles of arguments. They do not belong to the set of roles introduced

by a particular frame and they do not depend on the presence of other roles in the sentence. In Subsection 2.2 all types of modifiers are explained in more detail.

Frame	Semantic roles
pay.01	A0: payer or buyer A1: money or attention A2: person being paid, destination of attention A3: commodity, paid for what
pay.02 (pay off)	A0: payer A1: debt A2: owed to whom, person paid
pay.03 (pay out)	A0: payer or buyer A1: money or attention A2: person being paid, destination of attention A3: commodity, paid for what
pay.04	A1: thing succeeding or working out
pay.05 (pay off)	A1: thing succeeding or working out
pay.06 (pay down)	A0: payer A1: debt

Table 1: The PropBank lexicon entry for the verb *pay*.

(3) [$_{A0}$ The Latin American nation] has [$_{REL-PAY.01}$ paid] [$_{A1}$ very little] [$_{A3}$ on its debt] [$_{AM-TMP}$ since early last year].

For example, the verb *pay* in 3 assigns three semantic roles to its arguments and one to a modifier. A0 is attached to the noun phrase that is the subject of the sentence (*The Latin American nation*) and it represents the AGENT. A1 is attached to the direct object (*very little*). A3 is attached to the prepositional phrase that denotes the purpose of the payment (*on its debt*). The label for the modifier (*since early last year*), AM-TMP, is mapped from the syntactic label for the corresponding phrase.

2.1 Annotation of arguments (A0-A5)

2.1.1 The PropBank frame files

The argument labels for each verb are specified in **the frame files**. The frame files provide verb-specific description of all possible semantic roles and illustrate these roles with examples.

Frame File for the verb 'expect': (*s'attendre* in French)

Roles:
A0: expecter
A1: thing expected

Example: Transitive, active:
Les gestionnaires de fonds d'investissement s'attendent à de nouvelles baisses des taux d'intérêt.

A0: Les gestionnaires de fonds d'investissement
REL: s'attendent
A1: à de nouvelles baisses des taux d'intérêt

2.1.2 Choosing A0 versus A1

In most cases, choosing an argument label is straightforward, given the verb specific definition of this label in the frame files. However, in some cases, one needs to choose between the A0 and the A1 label.

The A0 label is assigned to arguments which are understood as agents, causers, or experiencers. The A1 label is usually assigned to the patient argument, i.e. the argument which undergoes the change of state or is being affected by the action. A0 arguments are the subjects of transitive verbs and a class of intransitive verbs called unergatives.

Jean (A0) a chanté la chanson.
Jean (A0) a éternué.

Semantically, A0 arguments have what Dowty (1991) called Proto-Agent properties, such as

- volitional involvement in the event or state
- causing an event or change of state in another participant
- movement relative to the position of another participant

A1 arguments are the objects of transitive verbs and the subjects of intransitive verbs called unaccusatives:

Jean a brisé la fenêtre (A1)
La fenêtre (A1) s'est brisée

These arguments have Proto-Patient properties, which means that these arguments

- undergo change of state
- are causally affected by another participant
- are stationary relative to movement of another participant

Whereas for many verbs, the choice between A0 or A1 does not present any difficulties, there is a class of intransitive verbs (known as verbs of variable behavior), where the argument can be tagged as either A0 or A1.

Une balle (A1) a atterri à ses pieds
Il (A0) a atterri

Arguments which are interpreted as agents should always be marked as A0, independent of whether they are also the ones which undergo the action.

In general, if an argument satisfies two roles, the highest ranked argument label should be selected, where A0 > A1 > A2 > ... > *A*5.

Given this rule, agents are ranked higher than patients. If an argument is both an agent and a patient, then an A0 label should be selected.

Not all A0s are agentive, however. There are many inanimate as well as clausal arguments which are being marked as A0s. These arguments are usually the ones which cause an action or a change of state.

A notion which might be useful for selecting A0 arguments is the notion of 'internally caused' as opposed to 'externally caused' eventualities, as defined in Levin and Rapapport (1995). In internally-caused eventualities, "some property inherent to the argument of the verb is responsible for bringing about the eventuality. For agentive verbs such as *play, speak*, or *work* (*jouer,parler, travailler*), the inherent property responsible for the eventuality is the will or volition of the agent who performs the activity. However, an internally caused eventuality need not be agentive. For example, the verbs *blush* (*rougir*) and *tremble* (*trembler*) are not agentive, but they, nevertheless, can be considered to denote internally caused eventualities, because these eventualities arise from internal properties of the arguments, typically an emotional reaction. In contrast to internally caused verbs, verbs which are externally caused inherently imply the existence of an external cause with an immediate control over bringing about the eventuality denoted by the verb: an agent, and instrument, a natural force, or a circumstance. Thus something breaks because of the existence of some external cause; something does not break because of its own properties" (Levin and Rappaport, 1995). The difference between internal and external causation is important for distinguishing A0s and A1s: the arguments which are responsible for bringing out the eventuality are A0s, whereas those which undergo an externally caused event are A1s.

To sum up, A0 arguments are the arguments which cause the action denoted by the verb, either agentively or not, as well as those which are traditionally classified as experiencers, i.e. the arguments of stative verbs such as *love, hate, fear* (*aimer,haïr,craindre*. A1 arguments, on the other hand, are those that change due to external causation, as well as other types of 'patient'-like arguments.

2.2 Annotation of modifiers (AMs)

We will give a short description and some examples for all types of modifiers in the following subsections:

2.2.1 Directionals (AM-DIR)

Directional modifiers show motion along some path. Both "source" and "goal" are grouped under "direction." On the other hand, if there is no clear path being followed a "location" marker should be used instead. Thus, "marcher le long de la route (walked along the road)" is a directional, but "marcher dans la campagne (walk around the countryside)" is a location.

Workers dumped large burlap sacks of the imported material into a huge bin, poured in cotton and acetate fibers and mechanically mixed the dry fibers in a process used to make filters.

Les travailleurs déversaient dans un grand bac des grands sacs de toile du matériel importé, coulaient du coton et des fibres d'acétate et mélangeaient mécaniquement les fibres sèches dans un processus utilisé pour la fabrication de filtres

A0:	Workers	/ Les travailleurs
REL:	dumped	/ déversaient
A1:	large burlap sacks of the imported material	/ des grands sacs de toile du matériel importé
AM-DIR:	into a huge bin	/ dans un grand bac

"No one wants the U.S. to pick up its marbles and go home , " Mr. Hormats says.

" Personne ne veut que les Etats-Unis ramassent leurs billes et rentrent à la maison,"dit M. Hormats

A1:	the U.S.	/ les Etats-Unis
REL:	go	/ rentrent
AM-DIR:	home	/ à la maison

2.2.2 Locatives (AM-LOC)

Locative modifiers indicate where some action takes place. The notion of a locative is not restricted to physical locations, but **abstract locations** are being marked as LOC as well, as "dans son discours-LOC, il parlait de ... ([in his speech]-LOC he was talking about ...)".

The percentage of lung cancer deaths among the workers at the West Groton , Mass., paper factory appears to be the highest for any asbestos workers studied in Western industrialized countries , he said .

Le pourcentage des décès par cancer du poumon parmi les travailleurs de fabrique de papier de West Groton, Massachusetts, semble être le plus élevé pour tous les travailleurs de l'amiante qui ont été étudiés dans les pays occidentaux industrialisés, dit-il.

A1:	any asbestos workers	/ les travailleurs de l'amiante
REL:	studied	/ étudiés
AM-LOC:	in Western industrialized countries	/ dans les pays occidentaux industrialisés

Areas of the factory were particularly dusty where the crocidolite was used.

Des parties de l'usine étaient particulièrement poussiéreuses là où la crocidolite a été utilisée.

AM-LOC:	where	/ où
A1:	the crocidolite	/ la crocidolite
REL:	used	/ utilisée

In his ruling , Judge Curry added an additional $ 55 million to the commission's calculations.

Dans sa décision, le juge Curry ajouta un montant supplémentaire de 55 millions de dollars au calculs de la Commission.

AM-LOC:	In his ruling	/ Dans sa décision
A0:	Judge Curry	/ le juge Curry
REL:	added	/ ajouta
A1:	an additional $ 55 million	/ un montant supplémentaire de 55 millions de dollars
A2:	to the commission's calculations	/ au calculs de la Commission

2.2.3 Manner markers (AM-MNR)

Manner adverbs specify how an action is performed. For example, "il travaille bien avec les autres (works well with others)" is a manner. Manner tags should be used when an adverb be an answer to a question starting with 'comment? (how?)'. When in doubt between AM-MNR and AM-TMP, choose AM-TMP.

Among 33 men who worked closely with the substance, 28 have died – more than three times the expected number.

Parmi les 33 hommes qui ont travaillé en contact étroit avec la substance, 28 sont morts - un nombre plus de trois fois supérier que prévu.

A0:	33 men	/ les 33 hommes
R-A0:	who	/ qui
REL:	worked	/ travaillé
AM-MNR:	closely	/ en contact étroit
A1-with:	with the substance	/ avec la substance

Workers dumped large burlap sacks of the imported material into a huge bin, poured in cotton and acetate fibers and mechanically mixed the dry fibers in a process used to make filters.

Les travailleurs déversaient dans un grand bac des grands sacs de toile du matériel importé, coulaient du coton et des fibres d'acétate et mélangeaient mécaniquement les fibres sèches dans un processus utilisé pour la fabrication de filtres.

A0:	Workers	/ Les travailleurs
AM-MNR:	mechanically	/ mécaniquement
REL:	mixed	/ mélangeaient
A1:	the dry fibers	/ les fibres sèches
AM-LOC:	in a process used to make filters	/ dans un processus utilisé pour la fabrication de filtres

The next morning, with a police escort, busloads of executives and their wives raced to the Indianapolis Motor Speedway , unimpeded by traffic or red lights.

Le lendemain matin, avec une escorte de police, des autobus remplis de cadres et de leurs épouses sont accourus vers l'Indianapolis Motor Speedway, sans être gênés ni par le trafic ni par les feux rouges.

AM-TMP:	The next morning	/ Le lendemain matin
AM-MNR:	with a police escort	/ avec une escorte de police
A0:	busloads of executives and their wives	/ des autobus remplis de cadres et de leurs épouses
REL:	raced	/ accourus
A1:	to the Indianapolis Motor Speedway	/ vers l'Indianapolis Motor Speedway
AM-ADV:	unimpeded by traffic or red lights	/ sans être gênés ni par le trafic ni par les feux rouges

2.2.4 Temporal markers (AM-TMP)

Temporal AMs show when an action took place, such as "en 1987 (in 1987)", "mercredi dernier (last Wednesday)", "tout à l'heure (soon)" or "tout de suite, immédiatement (immediately)". Also included in this category are adverbs of frequency (e.g. souvent, toujours, par fois, (often always, sometimes)

with the exception of 'jamais (never)', see NEG below), adverbs of duration "pendant un an (for a year/in an year)", order, e.g. "premier (first)", and repetition, e.g. "de nouveau, encore (again)". When in doubt between AM-MNR and AM-TMP, choose AM-TMP.

A form of asbestos once used to make Kent cigarette filters has caused a high percentage of cancer deaths among a group of workers exposed to it more than 30 years ago , researchers reported.

Une forme d'amiante autrefois utilisée pour la fabrication de filtres de cigarettes Kent a provoqué un pourcentage élevé de décès par cancer chez un groupe de travailleurs exposés il y a plus de 30 ans, ont signalé encore plus des chercheurs.

A1:	A form of asbestos	/ Une forme d'amiante
AM-TMP:	once	/ autrefois
REL:	used	/ utilisée
AM-PNC:	to make Kent cigarette filters	/ pour la fabrication de filtres de cigarettes Kent

Four of the five surviving workers have asbestos-related diseases, including three with recently diagnosed cancer.

Quatre des cinq travailleurs survivants ont des maladies liées à l'amiante, y compris trois avec un cancer diagnostiqué récemment.

AM-TMP:	recently	/ récemment
REL:	diagnosed	/ diagnostiqué
A2:	cancer	/ un cancer

2.2.5 Extent markers (AM-EXT)

AM-EXT indicate the amount of change occurring from an action, and are used mostly for

- numerical modifiers like "(raised prices) by 15%",
- quantifiers such as "a lot"
- and comparatives such as "(he raised prices) more than she did".

"An active 55-year-old in Boca Raton may care more about Senior Olympic games, while a 75-year-old in Panama City may care more about a seminar on health, " she says.

" Une personne active de 55 ans à Boca Raton peut plus se soucier des jeux olyimpiques pour senior, alors que celle de 75 ans dans la ville de Panama peut s'interesser plus à un séminaire sur la santé,"dit-elle.

A0:	An active 55-year-old in Boca Raton	/ Une personne active de 55 ans à Boca Raton
AM-MOD:	may	/ peut
REL:	care	/ soucier
AM-EXT:	more	/ plus
A1:	about Senior Olympic games	/ des jeux olyimpiques pour senior
AM-ADV:	while a 75-year-old in Panama City may care more about a seminar on health	/ alors que celle de 75 ans dans la ville de Panama peut s'interesser plus à un séminaire sur la santé

Rep. Jerry Lewis, a conservative Californian, added a provision of his own, intended to assist Bolivia, and the Senate then broadened the list further by including all countries in the U.S. Caribbean Basin initiate as well as the Philippines - backed by the powerful Hawaii Democrat Sen. Daniel Inouye.

Rep Jerry Lewis, un Californien conservateur, a ajouté une disposition de son propre chef, destiné à aider la Bolivie, et le Sénat a ensuite encore plus élargi la liste en incluant tous les pays du bassin des Caraïbes US ainsi que les Philippines - soutenu par le puissant sénateur démocrate de Hawaï, Daniel Inouye.

A0:	the Senate	/ le Sénat
AM-TMP:	then	/ ensuite
REL:	broadened	/ élargi
A1:	the list	/ la liste
AM-EXT:	further	/ encore
AM-MNR:	by including all countries in the U.S. Caribbean Basin initiate as well as the Philippines	/ en incluant tous les pays du bassin des Caraïbes US ainsi que les Philippines
AM-ADV:	backed by the powerful Hawaii Democrat Sen. Daniel Inouye	/ soutenu par le puissant sénateur démocrate de Hawaï, Daniel Inouye

2.2.6 Reciprocals (AM-REC)

These include constituents such as *himself, itself, themselves, together, each other, jointly, both,* which refer back to arguments.

But voters decided that if the stadium was such a good idea someone would build it himself, and rejected it 59% to 41%.

Mais les électeurs ont décidé que si le stade était une si bonne idée alors quelqu'un le construira lui-même, et l'ont rejeté à 59% contre 41%.

AM-ADV:	if the stadium was such a good idea	/ si le stade était une si bonne idée
A0:	someone	/ quelqu'un
AM-MOD:	would	/
REL:	build	/ construira
A1:	it	/ le
AM-REC:	himself	/ lui-même

2.2.7 Markers of secondary predication (AM-PRD)

These are used to show that a modifier of a predicate is in itself capable of carrying some predicate structure. Typical examples include

- depictives
- 'as'-phrases, e.g. 'supplied as security in the transaction' and other cases of secondary predication

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

Pierre Vinken, 61 ans, rejoindra le conseil en tant que directeur non-exécutif le 29 novembre.

A0:	Pierre Vinken, 61 years old	/ Pierre Vinken, 61 ans
AM-MOD:	will	/
REL:	join	/ rejoindra
A1:	the board	/ le conseil
AM-PRD:	as a nonexecutive director	/ en tant que directeur non-exécutif
AM-TMP:	Nov. 29	/ le 29 novembre

Prior to his term, a teacher bled to death in the halls, stabbed by a student.

Avant son mandat, un enseignant a saigné à mort dans les salles, poignardé par un étudiant.

AM-TMP:	Prior to his term	/ Avant son mandat
A1:	a teacher	/ un enseignant
REL:	bled	/ saigné
AM-PRD:	to death	/ à mort
AM-LOC:	in the halls	/ dans les salles
AM-ADV:	stabbed by a student	/ poignardé par un étudiant

2.2.8 Purpose clauses (AM-PNC)

Purpose clauses are used to show the motivation for some action. Clauses beginning with "pour ((in order) to)" are canonical purpose clauses.

In a disputed 1985 ruling, the Commerce Commission said Commonwealth Edison could raise its electricity rates by $ 49 million to pay for the plant.

Dans une décision contestée en 1985, la Commission du Commerce a précisé que Commonwealth Edison pourrait augmenter ses tarifs d'électricité à 49 millions de dollars pour payer les installations.

A0:	Commonwealth Edison	/ Commonwealth Edison
AM-MOD:	could	/ pourrait
REL:	raise	/ augmenter
A1:	its electricity rates	/ ses tarifs d'électricité
A2:	by $ 49 million	/ à 49 millions de dollars
AM-PNC:	to pay for the plant	/ pour payer les installations

2.2.9 Cause clauses (AM-CAU)

Similar to "purpose clauses", these indicate the reason for an action. Clauses beginning with "parce que" or "car" are canonical cause clauses. Also questions starting with ''pourquoi":

Pro-forma balance sheets clearly show why Cray Research favored the spinoff.

Des bilans pro-forma montrent clairement pourquoi Cray Research favorisa la spinoff.

AM-CAU:	why	/ pourquoi
A0:	Cray Research	/ Cray Research
REL:	favored	/ favorisa
A1:	the spinoff	/ la spinoff

However, five other countries – China, Thailand, India, Brazil and Mexico – will remain on that so-called priority watch list as a result of an interim review, U.S. Trade Representative Carla Hills announced.

Cependant, cinq autres pays - Chine, Thaïlande, Inde, Brésil et Mexique – restent sur ce qu'on appelle la liste de surveillance prioritaire après un réexamen intermédiaire, a annoncé la représentante U.S. au Commerce Carla Hills.

AM-DIS:	However	/ Cependant
A1:	five other countries – China, Thailand, India, Brazil and Mexico –	/ cinq autres pays – Chine, Thaïlande, Inde, Brésil et Mexique –
AM-MOD:	will	/
REL:	remain	/ restent
A3:	on that so-called priority watch list	/ sur ce qu'on appelle la liste de surveillance prioritaire
AM-CAU:	as a result of an interim review	/ après un réexamen intermédiaire

2.2.10 Discourse markers (AM-DIS)

These are markers which connect a sentence to a preceding sentence.

Examples of discourse markers are: *aussi, en plus (also, as well), pourtant, cependant, toutefois, néanmoins (however), mais (but), et (and), Comme nous l'avons vu auparavant (as we've seen before), plutôt (instead), d'autre part (on the other hand), par exemple (for instance)*, etc. Note that conjunctions such as *mais* or *et* are only marked in the beginning of the sentence. **Do not mark *and, or, but*, when they connect two clauses in the same sentence.**

But for now, they're looking forward to their winter meeting – Boca in February.

Mais pour l'instant, ils attendent avec impatience leur réunion d'hiver - Boca en Février.

AM-DIS:	But	/ Mais
AM-TMP:	for now	/ pour l'instant
A0:	they	/ ils
REL:	looking forward	/ attendent
A1:	to their winter meeting – Boca in February	/ leur réunion d'hiver - Boca en Février

Another type of discourse markers includes vocatives, which are marked as VOC in Treebank:

Kris, go home.
Kris, rentre à la maison.

AM-DIS:	Kris	/ Kris
Rel:	go	/ rentre
A0:	Kris	/ Kris
AM-DIR:	home	/ à la maison

I ain't kidding you, Vince
Je ne blague pas, Vince.

AM-DIS:	Vince	/ Vince
Rel:	kidding	/ blague
A0:	I	/ I
A1:	you	/
AM-NEG:	n't	/ ne pas

Vocatives are **not labelled** if they are appositives, such as 'Mr. President, you have ....'. They are annotated as discourse markers if htey are used in an imperative form.

And, finally, the class of discourse markers includes interjections such as 'mon Dieu (oh my god)' 'ah', and 'damn'.

2.2.11 Adverbials (AM-ADV)

These are used for syntactic elements which clearly modify the event structure of the verb in question, but which do not fall under any of the headings above.

1. Temporally related (modifiers of events)
   Des trésors sont situées tout près, attendant d'être ramassé.

   Treasures are just lying around, waiting to be picked up.

2. Intensional (modifiers of propositions)
   Probablement (Probably, possibly)

3. Focus-sensitive
   Seulement (only), même (even)

4. Sentential (evaluative, attitudinal, viewpoint, performatives)
   Heuresement, vraiment, frachement (fortunately, really, frankly speaking) clauses beginning with "étant donné que (given that)", " en dépit de (despite)", "à exception de (except for)", "si (if)".

As opposed to AM-MNR, which modify the verb, AM-ADVs usually modify the entire sentence.

2.2.12 Modals(AM-MOD)

English modals correspond in French to two sets of verbs that are called aspectual auxiliaires and modal auxiliaries. We label them all as AM-MOD. Like in English they can be single verbs or "phrasal modals". Aspectual modals indicate the beginiing, progress, ending and recent ending of an action. They are: *aller, être sur le point de, commencer à, se mettre à, être en train de, finir de, cesser de, venir de.* Modals indicating possibility, probability, and obligation are *pouvoir, devoir.* Notice that the verb *avoir* can indicate obligation in a phrasal construction: *Nous n'avons pas à faire cela.*

(Modals in English are: will, may, can, must, shall, might, should, could, would. Phrasal modals are "going (to)", "have (to)" and "used (to)".)

Phrasal modals are included, but, unlike the regular modals, these are also

annotated as verbs in their own right, where they take their own Negation and Adverbial markers, but not any numbered arguments. Thus, in the sentence “Jean n’a pas à courir”, “a” is a modal modifier of “courir”, but “ne pas” is a negation modifier of “a”, and not of “courir”.

2.2.13 Negation (AM-NEG)

This tag is used for elements such as “ne”, “pas”, “jamais”, “ne personne” and other markers of negative sentences.

Negation in French is systematically discontinuous, so we annotate “ne” as AM-NEG and “pas” or the negative polarity items as C-AM-NEG.

Negation is an important notion for Propbank annotation; therefore, all markers which indicate negation should be marked as NEG. For example, when annotating adverbials like “ jamais”, which could be marked as either TMP or NEG, the NEG tag should be used.

3 Annotating the French sentences

In this project, the annotator is asked to assign PropBank labels to the elements of the presented French sentences using the TrEd annotation tool and the PropBank frame files.

The TrEd annotation tool shows one sentence at a time. The sentence is presented in the upper part of the window. Its syntactic analysis presented in the form of a tree can be viewed in the lower part of the window (see Figure 1). The annotation of semantic structure is added to the nodes in the syntactic tree. The following procedure should be followed. (Instruction on how to use TrEd are given in Section 5.)

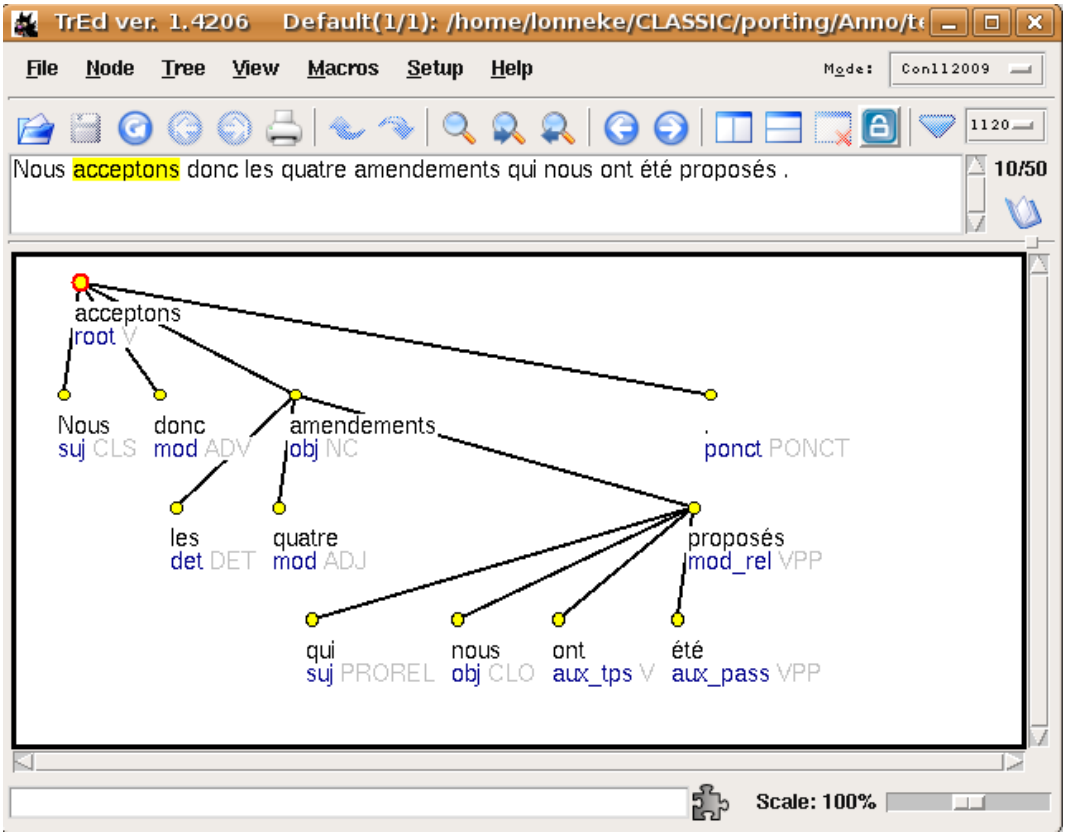

Figure 1: The TrEd interface

3.1 Stepwise procedure

Step 1: Find the predicate.

Go through the sentence in the upper part of the TrEd window and find a predicate. This can be a verb which heads a clause (except the verb *être*) or an adjective which is the past participle form of a verb. To distinguish between adjectival and verbal uses of past participles, apply the *très* test. If a past participle can be modified with *très*, then it is an adjective and it is not considered for this annotation. Otherwise, the form is a past participle and it needs to be annotated.

Step 2: Find the arguments of the predicate.

These are the words or phrases that denote the semantic elements related by the predicate, usually the subject of the clause and the complements of the verb. However, the arguments of a verb can be words or phrases that are not found in the same clause as the verb, but need to be looked for elsewhere in the sentence. Annotation of these arguments is explained in more detail in Subsection 4.2 and Subsection 4.3.

Step 3: Find the corresponding frame in the PropBank lexicon.

This is probably the most difficult step and it has to be done very carefully. Since semantic structures are generally preserved across languages, we can assume that French verbs introduce the same frames as English verbs. This is why we will be using the existing English labels that can be found in the PropBank frame files as much as possible. For example, if a French sentence contains a form of the verb *payer*, this predicate can be labelled using one of the frames specified for the English verb *pay*.

Since there will be more than one potential translation of a French predicate, it will be necessary to choose the best one. Look up the entries in the PropBank frame files for all the verbs that are potential translation equivalents of the French predicate. Choose the entry that contains the frame with the meaning most similar to the meaning of the French predicate *and* with the set of semantic roles most similar to the set found in the French sentence. If you are not sure of the English translation for the French predicate, you can look up a bilingual dictionary.

Since the PropBank lexicon is not an exhaustive list of all possible frames, it can still happen that no corresponding frame can be found for a French predicate. In this case, mark the French predicate with the label DUMMY and arguments according to your own intuition.

Step 4: Label the predicate with the frame label and its arguments with the corresponding semantic role labels (A0-5).

Up to this step, you will be using the simple view of the French sentence presented in the upper part of the annotation program window. Once an appropriate frame has been identified, the labels should be added to the syntactic tree presented in the lower part of the window. Note that the syntactic tree is based on the dependency grammar formalism and not on a phrase structure analysis. This means that there are no phrase nodes, which is why labels can only be attached to single words. This is not a problem for the predicate label, since it is attached to a single word anyway (to the verb). However, argument labels usually span over phrases. If a semantic role spans over a phrase (and not just a single word) attach the label to the syntactic head of the phrase.

Since the syntactic trees are produced by an automatic parser, there might be errors resulting in the actual head of the phrase being in the wrong place. In these cases, ignore the error and put the label on the actual head.

The predicate should be labelled with the frame label found in the PropBank frame files, leaving out the first part of the label the PropBank predicate label (REL) since it is same for all the predicates. If the verb form is in a compound tense, involving an auxiliary verb form, the label is attached only to the lexical form (main verb). The arguments should be labelled with the roles that are defined for that frame. The interpretation given for each semantic role should help you find the right labels.

Not all semantic roles that are listed in the PropBank frame set have to occur in the analysed sentence. For examples, the role A2 in the **pay.01** list is missing in (3). However, the roles that do occur should be labelled as specified in the entry.

Step 5: Find all the adjuncts in the clause and mark them with the corresponding AM labels. If an adjunct is a global modifier (modifying the whole sentence and not the particular predicate), relate it to the main verb.

Step 6: Repeat steps 1−5 for all the predicates in the sentence.

4 Special Cases

4.1 Clauses as arguments

Arguments of some verbs such as *dire, croire, penser, comprendre* can be clauses. If this is the case, the argument label is attached to the complementizer, marked with boldface in the following example.

Marie a compris ***que*** *le bruit venait de la boîte.*
A0: Marie
REL: compris
A1: **que** le bruit venait de la boîte

4.2 Unexpressed arguments

Some predicates (infinitive forms) can have an argument that is not syntactically realised in the same clause, but it is interpreted as a constituents that can be found in the previous context. These arguments should be identified and labelled, as illustrated in the following example:

Marie aime nager.
A0: Marie
REL: aime
A1: nager

A0: Marie
REL: nager

4.3 Arguments with co-reference

Pronouns that refer to some constituent that can be found in the previous context should be labelled with a special *reference* label, R-A. The corresponding A label should be attached to its antecedent, as shown in the following example.

Marie n'a pas aimé le livre que sa mère lui a donné.
A2: Marie
...
A1: le livre
R-A1: que
A0: sa mère
R-A2: lui
REL: donné

4.4 Coordinated arguments

If a clause contains two co-ordinated nominal arguments of the same predicate, only the first of them should be given the label, as shown in the following example:

Marie et Jean travaillent sur un projet.
A0: Marie
REL: travaillent
A1: sur un projet

Sentential coordinates are annotated normally as independent sentences.

4.5 Discontinuous arguments

Verbs of saying can be inserted in the constituent that denotes one of their arguments. In this case, the part of the argument that precedes the predicate is marked with the A label, and the part of the argument that follows the predicate is marked with the corresponding C-A label, where C stands for *continuation*.

Ces livres, a dit Marie, ont été écrits il y a longtemps.
A1: Ces livres
REL: dit
A0: Marie
C-A1: ont été écrits il y a longtemps

4.6 Expletives

Expletives are constituents that have a syntactic function, but have no semantic content. This is why they cannot be assigned any semantic role. The pronoun *Il* in the following example does not receive any A label, even though it is the the subject of the sentence. French expletives are *il* and *ce*.

Il paraît que le spectacle a provoqué beaucoup de discussions.
REL: paraît
A1: que le spectacle a provoqué beaucoup de discussions

4.7 Fixed expressions: idioms and collocations

Two kinds of fixed expressions can occur in the sentences: **idioms**, phrases with non-compositional meaning (e.g.*se mettre le doigt dans l'œil*, *se casser le nez*, *tirer sur la corde*) and **collocations** with compositional meaning (e.g. *mettre à jour*, *faire attention*, *tenir compte de*, *donner une fête*). If an idiom involves the predicate of the clause, as in the given examples, the clause should not be analysed. However, if an idiom is found inside one of the arguments only (e.g. *savoir quelque chose* ***sur le bout des doigts***), the clause should be analysed. Collocations, should be analysed as regular phrases.

4.8 English quasi-synonyms translated by the same French verb

It will often happen that a given French verb can be translated by many English verbs and therefore be annotated by several frames. We choose in this case, first the frame that best covers all and only the arguments that need to be annotated. If syntax still leaves an option open, we choose the frame that best conveys the French meaning, even in cases if this means choosing the verb that might not be the most commonly used in English, taking care to always stay in the realm of literal or quasi-literal translations. Usually a good way of testing is to try a translation in Google translate.

5 The annotation tool

5.1 Opening the PropBank frame files

The frame files of PropBank predicates annotated in PropBanklist are available on the Internet. Start an Internet browser. Type the following address `http://verbs.colorado.edu/propbank/framesets-english/` in the address bar. This will bring you to the page containing the index of all predicates annotated in PropBank. Keep the page open during the annotation. You will need it to find the appropriate predicate label for the predicates in the sentence and to find examples of the roles the predicate takes. To find the predicate you need, just scroll the list up and down or search for it using `cntrl+F`. Once you find it, click on it to see its framesets with corresponding labels and examples.

5.2 Connecting to the computer that runs the annotation tool

- Switch on a computer that runs Ubuntu.
- Log in
- Open a terminal
- Type `ssh -X your_first_name@129.194.69.238`, hit enter
- You will be prompted for a password (this is given to you beforehand)
- Type your password, hit enter
- You are now working remotely on the computer that runs the annotation tool

5.3 Starting TrEd

- To start TrEd just type in the command line: `start_tred`
- This will open a new working window which will be used in the annotation.

5.4 Finding the file with sentences to annotate

- In the main menu of the TrEd window, click on `File/Open`. This will show all available files.
- Find the file with the name `<filename>` in the directory called `Annotation`, click on it and then click `Open`.
- The first sentence will appear in the upper part and its syntactic analysis in the lower part of the window.
- At this point, the program is ready.

5.5 Labelling predicates

- Once a predicate has been identified in the sentence, find its node in the syntactic graph. Use `Shift+Click` to select the node. The selected node becomes bigger and turns orange.
- Then double click on it. A dialogue box will pop-up.
- Find the matching predicate label in the PropBank frame files.
- Copy and paste the predicate label from the PropBank frame files to the dialogue box of TrEd (without spaces) in the field `Pred` and then click `OK`. The predicate label will be added to the syntactic graph, next to the annotated word.
- Note: a predicate can be an argument of another predicate at the same time. The procedure of labelling a predicate is the same, regardless of whether it has already been labelled as an argument.

5.6 Labelling arguments and modifiers

- A role can be annotated only after the corresponding predicate has been selected.
- Once the word that bears the role has been identified in the sentence, find its node in the syntactic graph.
- Use `Control+Click` to select the node. A dialogue box will pop-up.

- Write down the role label in the window and then click `OK`.
- An arrow will be added to the syntactic graph connecting the predicate and the annotated argument and bearing the assigned semantic role.
- Make sure the arc has a colour and is not gray. If it is gray, the program did not understand the label you gave, so you have probably typed it incorrectly.
- Repeat the same procedure for all the arguments of the predicate.

5.7 Removing or editing annotations

- To edit the predicate label: double click on the predicate label and a dialogue box will pop up that will allow you to change the label.
- To edit the semantic role label: double click on the arc and a dialogue box will pop up that will allow you to make changes.
- To remove a completed arc: mark the predicate (Shift+Click) and Control+Click on the argument. The procedure is the same as for adding an arc. A dialogue box will ask you if you want to remove the arc.

5.8 Starting a new sentence

- Once all the predicates and all the arguments in the sentence have been annotated, save the file (click on the `Save` icon)
- Click on the right arrow in the arrow pair on the right side of the menu to move to the next sentence.

References

[1] David Dowty. 1991 Thematic Proto-Roles and Argument Selection. Language, 67(3), 547–619.

[2] Levin, Beth and Malka Rappaport Hovav. 1995. *Unaccusativity.* MIT Press, Cambridge, MA.

[3] Martha Palmer and Daniel Gildea and Paul Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. Computational Linguistics, 31(1), 71–105.