

# Global Methods for Cross-lingual Semantic Role and Predicate Labelling

**Lonneke van der Plas**

Institute for NLP

Pfaffenwaldring 5B

70569 Stuttgart, Germany

vdplasme@ims.uni-stuttgart.de

**Marianna Apidianaki**

LIMSI-CNRS

Rue John von Neumann

91405 Orsay Cedex, France

marianna@limsi.fr

**Chenhua Chen**

Institute for NLP

Pfaffenwaldring 5B

70569 Stuttgart, Germany

cch.chenhua@googlemail.com

## Abstract

We address the problem of transferring semantic annotations to new languages using parallel corpora. Previous work has transferred these annotations on a token-to-token basis, an approach that is sensitive to alignment errors and translation shifts. We present a global approach to transfer that aggregates information across the whole parallel corpus and leads to more robust labellers. We build two global models, one for predicate labelling and one for role labelling, each tailored to the task at hand. We show that the combination of direct and global methods outperforms previous results.

## 1 Introduction

With the proliferation of the Internet in non-English speaking countries, the need for multilingual processing becomes more and more pressing. Various efforts have focused on developing language-independent NLP tools and extending to other languages tools that had been exclusive to English. Furthermore, several annotation efforts have been devoted to developing resources for different languages, needed for supervised learning (Hajič et al., 2009). However, there is still a large number of languages for which corpora with semantic annotations do not exist. Since manual annotation is a costly and time-consuming approach to resource development, cross-lingual annotation transfer offers an alternative.

Semantic parsing or semantic role labelling (SRL) is the task of automatically labelling predicates and arguments with predicate-argument structure. This level of analysis provides a more stable semantic representation across syntactically different sentences. The example sentences (1a) and (1b) illustrate how the semantic annotation remains stable across the locative alternation of the verb *load*.

- (1) a. [AGENT Jessica] [REL-LOAD.01 loaded] [THEME boxes] [DESTINATION into the wagon].  
b. [AGENT Jessica] [REL-LOAD.01 loaded] [DESTINATION the wagon] [THEME with boxes].

Also in the cross-lingual setting, the predicate-argument structure of a sentence is considered to be more stable than its syntactic form as the English sentence in (2a) and its French translation in (2b) show:

- (2) a. [EXPERIENCER Mary] [REL-LIKE.01 liked] [CONTENT the idea]. (English)  
b. [CONTENT L'idée] a [REL-LIKE.01 plu] [EXPERIENCER à Marie]. (French)

This is why several pieces of work have transferred semantic annotations from a source language, for which semantic annotations exist, to a target language using parallel corpora (Padó, 2007; Basili et al., 2009; Annesi and Basili, 2010). These transfer methods rely on the assumption of semantic equivalence of the original and the translated sentences, but also on correct and complete alignments between words

---

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

or constituents in those sentences. We will refer to these traditional methods as direct transfer because the semantic annotations are transferred directly from token to token. Although direct transfer methods are straightforward and easy to implement, they are vulnerable to missing or incorrect alignments which lead to missing and erroneous annotations in the target language. Consequently, non-literal translations and translation shifts present major problems for these methods.

In this paper we propose a global approach to the cross-lingual transfer of PropBank (Palmer et al., 2005) semantic annotations that aggregates information at the corpus level and, as a consequence, is more robust to non-literal translations and alignment errors. Our global approach involves two steps: in the learning step, two global models are learned on the basis of role and predicate annotations in the source language (English). In the labelling step, these models assign labels to verbs and their arguments in the target language (French) without consulting any parallel data. Contrary to previous work, we build separate models for the transfer of semantic role and predicate annotation because predictors for the two models are different in nature. We model cross-lingual transfer of predicate labels as a cross-lingual word sense disambiguation (WSD) task because this fits well with the lexical nature of the task: annotating French verbs with English predicate labels. Our approach to predicate labelling needs word alignments but instead of relying on local (token-to-token) correspondences like the direct method, it exploits alignment information gathered from the whole corpus thus avoiding transfer errors caused by local misalignments. Our model for cross-lingual semantic role labelling<sup>1</sup> is based on syntactic-semantic mappings learned from a gold annotated monolingual corpus. The SRL method does not need aligned data. Our methods are knowledge-lean as our predicate labelling method only needs a part of speech (PoS) tagger in the two languages and no syntactic information on either side, in contrast to previous work. For SRL, a syntactic parser for the target language is needed, but no joint semantic-syntactic parsing framework as was the case in previous work (van der Plas et al., 2011). The requirements of the global annotation transfer methods in terms of data and annotations, and their differences from direct transfer, are illustrated in Figure 1.

Our contribution is three-fold. First, we present a global approach to semantic annotation transfer that corrects token-level mistakes found in traditional direct transfer methods. We show the strengths and limitations of global vs. direct transfer and explain how the two can be combined. Second, in contrast to previous work, we address the two tasks of cross-lingual predicate labelling and cross-lingual semantic role labelling by building two separate models tailored to the task at hand. We show how the predicate labels produced by our high-coverage and knowledge-lean model for cross-lingual predicate labelling are successfully used as predictors for semantic role labelling. Third, due to its knowledge-lean and flexible character, our method adapts relatively easily to other language pairs without requiring semantic lexicons in the target language.

In the next section, we present related work on cross-lingual annotation transfer. In Section 3 we present the data used in our experiments and in Section 4 we briefly discuss direct transfer. The two global methods proposed in this paper are presented in Section 5. We report and discuss our results in Section 6, before concluding.

## 2 Related work

Transferring annotation from one language to another in order to train monolingual tools for new languages was first introduced by Yarowsky and Ngai (2001). In their approach, token-level part-of-speech (PoS) and noun phrase bracketing information was projected across word-aligned bitext and this partial annotation served to estimate the parameters of a model that generalized from the noisy projection in a robust way. In more recent work, Das and Petrov (2011) propose a graph-based framework for projecting syntactic information across languages. They create type-level tag dictionaries by aggregating over projected token-level information extracted from bi-text and use label propagation on a similarity graph to smooth and expand the label distributions. A different approach to cross-lingual PoS tagging is proposed

---

<sup>1</sup>Most unsupervised approaches consider argument identification as a separate task that is omitted (Lang and Lapata, 2010) or performed heuristically (Lang and Lapata, 2011). We focus on semantic role labelling in this paper and consider argument identification as given.

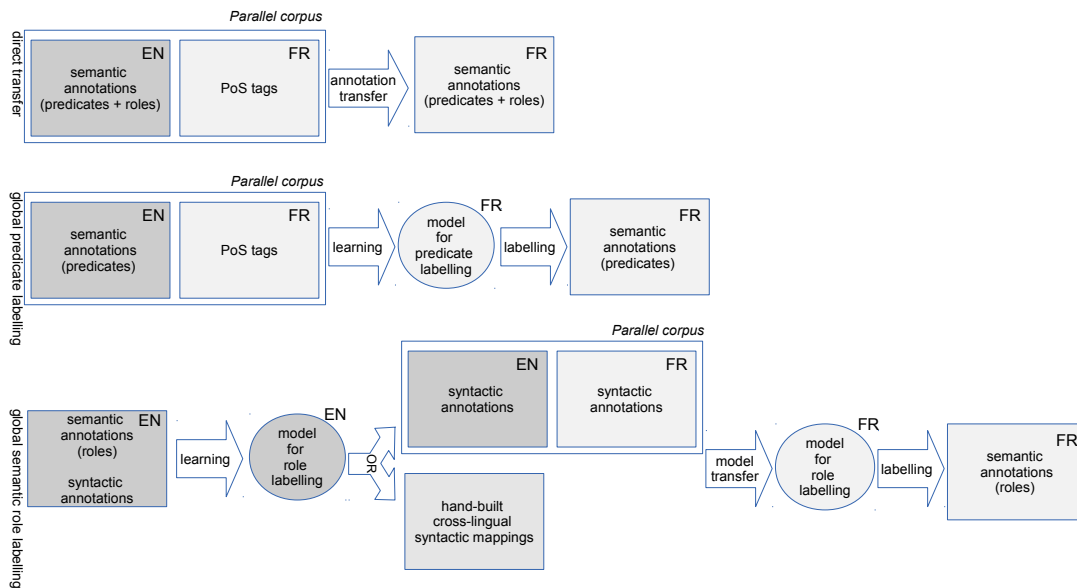


Figure 1: Direct vs. global cross-lingual transfer of semantic annotations.

by Täckström et al. (2013) who couple token and type constraints to guide learning. Our approach to cross-lingual semantic role labelling follows this vein. Instead of solely relying on token-level information acquired from word-alignments, we combine this with type-level information captured by our global methods which are trained on the entire corpus. We however are concerned with semantic annotations and not PoS.

Transfer of semantic annotation has started off with direct transfer of FrameNet semantic annotations (Padó, 2007; Basili et al., 2009; Annesi and Basili, 2010). With the addition of a learning step and the use of PropBank data, Van der Plas et al. (2011) have scaled up previous efforts. They show that a joint semantic-syntactic parser trained on the output of direct transfer produces better parses than the input it received by aggregating information across multiple examples. In their work, transfer of predicate labels and semantic role labels is done in one step. The model needs an aggressive filter to compensate for missing annotations on the predicate level after direct transfer. This filter successively leads to drops in performance for the role labellings. Here, we build two separate global models that complement direct transfer instead of relying on it.

The same emphasis on learning is found in cross-lingual model transfer where source language models are adapted to work on the target language directly. For semantic role labelling, Kozhevnikov and Titov (2013) use shared feature representations (syntactic and lexical) to adapt a source model to a target-language model. The ideas behind their cross-lingual model adaptation resemble the ideas behind our global method for semantic role labelling. However, in contrast to their work we do not consider the predicate labelling as given because, as manual annotations show (van der Plas et al., 2010), this task is not trivial. We first build a tailored global model for cross-lingual predicate labelling and then use the predicted predicate labels for semantic role labelling.

### 3 Data

In our experiments, we use the English-French part of the Europarl corpus (Koehn, 2005). The dataset is tokenised and lowercased and only sentence pairs corresponding to a one-to-one sentence alignment with lengths ranging from one to 40 tokens on both French and English sides are considered. Furthermore, because translation shifts are known to pose problems for the automatic projection of semantic roles across languages (Padó, 2007), we select only those parallel sentences in Europarl that are direct

translations from English to French or vice versa. In the end, we have a parallel corpus of 276-thousand sentence pairs.

The English part of the parallel corpus is annotated by a freely-available syntactic-semantic parser (Henderson et al., 2008; Titov et al., 2009) trained on the CoNLL 2009 training set (the Penn Treebank corpus (Marcus et al., 1993) merged with PropBank labels (Palmer et al., 2005) and NomBank labels (Meyers, 2007)). The probabilistic model is a joint generative model of syntactic and semantic dependencies that maximises the joint probability of the dependencies while building two separate structures.

The WSD classifier used for predicate labelling is trained on the parallel training corpus tagged with semantic roles on the English side. The candidate predicate labels that are considered by the classifier for each French verb are the labels of its English translations in the training corpus. Verbs on the English side are replaced by the corresponding predicate label where available. Then both parts of the corpus are lemmatized and tagged by part of speech (Schmid, 1994) and the parallel files are rebuilt (one sentence per line) by replacing words on both sides by the corresponding ‘lemma\_PoS tag’ pair. The corpus is then word aligned in both directions using GIZA++ (Och and Ney, 2003) and a lexicon is built from intersecting alignments. Lexicon entries for French verbs contain the English predicate labels to which they were aligned in the training corpus. The entry for the verb *encourager*, for instance, contains seven predicate labels: {urge.01, foster.01, stimulate.01, promote.02, encourage.01, encourage.02, renew.01}, two of which correspond to the same English verb (*encourage*). We keep labels with an alignment confidence score above 0.01 according to GIZA++.

Contrary to our predicate labelling model, the role labelling model needs syntactic information in the target language. For parsing French, we use the dependency parser described in Titov and Henderson (2007). We train the parser on the dependency version of the French Paris 7 treebank (Candito et al., 2009), achieving 87.2% labelled accuracy on this data set. The French Treebank (Abeillé et al., 2003) is a treebank of 21,564 sentences annotated with constituency annotation. We use the automatic dependency conversion of the French Treebank into dependency format (Candito et al., 2009) to train the French syntactic parser that is used to annotate the French part of the parallel corpus.

For testing, we use the hand-annotated data described in Van der Plas et al. (2010). We randomly split those 1000 sentences into test and development set containing 500 sentences each. We use the development set for the current experiments, which contains 1,917 core roles in total. We limit our experiments to verbal predicates because the semantic annotations on French test sentences are limited to verbal predicates.

## 4 Direct cross-lingual transfer

Before explaining the global methods, we present the direct semantic transfer (DST) method proposed by Van der Plas et al. (2011) that we use for comparisons and combinations throughout this paper. The method is based on the Direct Correspondence Assumption for syntactic dependency trees proposed by Hwa et al. (2005). The transfer proceeds as follows: For any pair of sentences E and F that are translations of each other in the parallel corpus, 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 that are transferred are semantic role dependencies and the properties are predicate senses. These are transferred from the English part of the parallel training corpus that is automatically annotated with syntactic-semantic analyses, as explained in the previous section.

## 5 Global cross-lingual transfer of semantic annotations

In contrast to direct transfer where annotations are transferred on a token-to-token basis in word-aligned sentences, we propose two global methods for cross-lingual transfer, one for predicates and one for semantic roles, that both consist of a learning and a labelling step. Our methods are globally defined and as a consequence rely less on local translation correspondences than previous methods, which makes them less vulnerable to missing and incorrect alignment links.

## 5.1 Global cross-lingual predicate labelling

In cross-lingual predicate labelling, our aim is to put predicate labels that originate from the English side of the parallel corpus on the French verbs in the other side of the corpus. The predicate labels contain the English verb and its sense. For example, “give.01” stands for the first sense of the verb *give*. As the predicate label contains a lot of lexical information, putting the correct English predicate label on a French verb is very close to Word Sense Disambiguation (WSD), the task of automatically identifying the meaning of words in context (Navigli, 2009). In the cross-lingual variant of this task, the candidate senses are the words’ translations in other languages and WSD aims at predicting semantically correct translations for words in context (Resnik and Yarowsky, 2000; Ng et al., 2003; Carpuat and Wu, 2007; Apidianaki, 2009). The main difference between cross-lingual WSD and our cross-lingual transfer of predicate labels is that we do not search for correct translations of French words but for the most appropriate predicate labels in context (i.e. verbs disambiguated with a predicate sense).

The global predicate labelling method consists of a learning step and a labelling step. During learning, we compute estimates for annotation transfer on the basis of the word alignments between English and French predicates over the entire parallel training corpus. At labelling time, we label French verbs with English predicate labels without the need for parallel data or alignments. The method is language-independent and only requires minimal linguistic resources (PoS information).

In terms of coverage, a predicate label is provided for all French verbs in the test set for which information was retained during training and not only for aligned ones, in contrast to direct transfer. We expect to augment the recall when using global estimates and hope that the effect on precision is not too negative.

### Learning

For each French verb ( $v$ ) in the lexicon built as described in Section 3, we want to be able to identify its correct predicate label in a new context by choosing one among its candidate labels ( $L$ ) retained from the training corpus. A feature vector is built for each candidate label  $L_i$  ( $1 \leq i \leq |L|$ ) found for the verb  $v$  in the lexicon, following the procedure described in Apidianaki et al. (2012). For each candidate label, we extract the content word co-occurrences of the verb  $v$  in the French sentences where it translates an English verb tagged with this label in the training corpus. The retained French words constitute the features of the vector built for that label. Let  $N$  be the number of features retained for each label  $L_i$  of the verb  $v$  from the corresponding French contexts. Each feature  $F_j$  ( $1 \leq j \leq N$ ) receives a total weight with the label ( $\text{tw}(F_j, L_i)$ ) which is learned from the data and defined as the product of the feature’s global weight ( $\text{gw}(F_j)$ ) and its local weight with that label ( $\text{lw}(F_j, L_i)$ ). The global weight of a feature  $F_j$  is a function of the number  $n$  of candidate labels of  $v$  to which  $F_j$  is related, and of the probabilities ( $p_{ij}$ ) that  $F_j$  co-occurs with instances of the verb  $v$  corresponding to each of the labels:

$$\text{gw}(F_j) = 1 - \frac{\sum_{i=1}^n p_{ij} \log(p_{ij})}{n} \quad (1)$$

Each  $p_{ij}$  is computed as the ratio of the co-occurrence counts of  $F_j$  with  $v$  when it is aligned to a label  $L_i$  to the total number of features ( $N$ ) seen with this candidate label:

$$p_{ij} = \frac{\text{cooc\_count}(F_j, L_i)}{N} \quad (2)$$

The local weight between feature  $F_j$  and label  $L_i$  ( $\text{lw}(F_j, L_i)$ ) directly depends on the number of times they occur together:

$$\text{lw}(F_j, L_i) = \log(\text{cooc\_count}(F_j, L_i)) \quad (3)$$

The intuition underlying this weighting scheme is that if an interesting semantic relation exists between a feature  $F_j$  and a specific predicate label  $L_i$  of a verb  $v$ , then we expect the probability ( $p_{ij}$ ) of the feature  $F_j$  occurring in the contexts where  $v$  is translated by this label to be larger than if they were independent. In other words, a feature gets a high total weight ( $\text{tw}$ ) with a label when it appears frequently in the corresponding French contexts and rarely in the contexts of the other labels.

## Labelling

Predicate identification on the French side is done by selecting verbs based on the PoS labels provided by the tagger and subsequently filtering out modals and instances of the verb *être* (*be*).<sup>2</sup> The most suitable predicate labels are then assigned to the retained French verbs by the disambiguation classifier. The context of a new instance of a French verb is compared to the weighted feature vectors ( $V_i$ 's) built for its candidate labels as described above, and an association score is assigned to each label. To facilitate comparison with the vectors, the new contexts (sentences) are lemmatised and PoS tagged on the fly (with TreeTagger) and the content word co-occurrences of the French verb are gathered in a bag of words. If common features ( $CF$ s) are found between the new context and the vector of a label ( $L_i$ ), their association score corresponds to the mean of the weights of their shared features with  $L_i$  found in the corresponding vector. In Equation 4,  $(CF_j)_{j=1}^{|CF|}$  is the set of common features between a label vector  $V_i$  and the new context  $C$  and  $tw$  is the total weight of a  $CF$  with label  $L_i$ , computed as explained in the previous section.

$$\text{assoc\_score}(V_i, C) = \frac{\sum_{j=1}^{|CF|} \text{tw}(CF_j, L_i)}{|CF|} \quad (4)$$

The label that receives the highest association score with the new context is returned and serves to annotate the corresponding French verb.

## 5.2 Global cross-lingual role labelling

For role labelling, we adopt a different strategy. Whereas predicate labels include a lot of lexical information, role labels do not. However, for role labels there is another source of information that helps to define global estimates: the correlation between syntax and semantics.

Previous work in monolingual unsupervised semantic role induction (Grenager and Manning, 2006; Lang and Lapata, 2010; Lang and Lapata, 2011) showed that mapping rules that assign semantic roles to arguments of a verb based on the syntactic functions of these arguments, represent a baseline that is very hard to beat. This strong correlation between syntactic labels and semantic role labels in the PropBank annotation has been shown in detail by Merlo and Van der Plas (2009). In contrast to previous work on monolingual unsupervised semantic role induction, we add the predicate label as a predictor. The core arguments of the verb, that are the numbered labels in PropBank, are known to be verb-specific. We have access to predicate labels assigned by the cross-lingual predicate labelling method described in the previous section and exploit them for role labelling.

For a given predicate, diathesis alternations are the major source of variation in propositions. They give rise to different syntactic structures, while the semantic roles remain stable. For example, the sentence “I gave the book to Jean” is syntactically different from “I gave Jean the book”, but semantic roles on the three arguments stay the same. We will show in a feasibility study that the effect of diathesis alternations on the correlation between syntax and semantics is limited. In a cross-lingual setting, structural divergences (Dorr, 1994) are expected to reduce the correlation between syntax and semantics. An example is the difference in syntactic structure between the sentences “Tu me manques” vs. “I miss you”, which are translations of each other, however the semantic roles are the same across languages.

As our global method is not restricted to alignments at labelling time, we are able to classify all given arguments<sup>3</sup> and not just those that are aligned in a parallel corpus. In this way, we believe that the negative effect of structural divergences and diathesis alternations is limited. Moreover, we show how mild supervision from the partial annotations that result from the direct transfer can potentially remedy these difficulties.

## Learning syntactic-semantic mappings

The syntactic-semantic mapping rules that are exploited by our model for role labelling are extracted from gold-annotated monolingual data. As a consequence, the extracted rules are of high quality which

<sup>2</sup>We exclude the verb *être* because its English counterpart (*be*) is not annotated in the CoNLL-2009 data used in our experiments.

<sup>3</sup>We focus on the classification of core semantic roles because diathesis alternations and cross-lingual divergences mainly involve these roles.

would not be the case if parallel data was used. Manually annotated parallel corpora are very sparse and automatic parsing introduces errors which might be propagated by the direct transfer methods and result in noisy annotations. Using gold-standard monolingual data thus ensures the quality of the mappings exploited by our global model.

We build a model that determines the most suitable semantic role label  $r$  for a given argument of a given predicate  $p$ , based on its syntactic dependency label  $d$ .<sup>4</sup> We simply compute the maximum likelihood estimates (MLE) and count occurrences of the following triples  $\langle p, d, r \rangle$  in a large body of English gold semantically and syntactically annotated data.

$$P_{MLE}(r|p, d) = \frac{\text{count}(p, d, r)}{\text{count}(p, d)} \quad (5)$$

In the cross-lingual setting, the mapping rules extracted from the English training data are applied to French. We learn the correspondences between English and French syntactic labels in a data-driven way by syntactically annotating both sides of our parallel training corpus. We base the cross-lingual syntactic mapping on alignment counts between syntactic labels in the parallel corpus parsed syntactically both on the English and the French side (cf. Section 3). An alternative that needs no parallel data is to study the annotation guidelines for the two languages and determine the cross-lingual correspondences between syntactic labels by hand.

The syntactic label set used for French (Candito et al., 2009) is less fine-grained than the English labels (20 versus 36). As a consequence, the mapping from English syntactic labels to French treebank labels is for the most part a many-to-one mapping, which leads to information loss but suffices for our purpose as will be shown in the next section.

Once the correspondences between the syntactic labels of the two syntactic annotation frameworks are discovered, the cross-lingual transfer of syntactic-semantic mappings consists in substituting the English syntactic labels with their French counterparts to adapt the model described above.

### Labelling

For role labelling, we use estimates derived from the training data (see Equation 5) to determine the most suitable role of a given argument. Because a particular triple in the test set might not have been seen during training, we backoff to 2-tuples that discard the predicate label, and backoff to A1 if neither the dependency label nor the predicate has been seen in training.

To treat the R-suffix, which takes care of anaphoric arguments, we use the following simple rule: for the monolingual setting all arguments with PoS-tags “WDT”, “WP”, and “WRB” receive the R-suffix. In the cross-lingual setting, we translate the PoS tags to the single French PoS tag “PROREL”. We do not treat the C-prefix, which takes care of discontinuous arguments, because there were only a few examples.

We do not accept duplicate semantic roles, a constraint that leads to valid role configurations in general (Punyakanok et al., 2008). We expect the more prominent semantic roles, such as A0 and A1, to appear earlier in the sentence than semantic roles with higher numbers. We therefore attribute semantic roles of a predicate from left to right.

### 5.3 Combining direct and global cross-lingual transfer

Direct transfer methods generally have low recall, we however expect them to be more precise than the global methods. In our combined method, we use the annotations assigned by direct transfer as the backbone and fill missing labels by the global methods. The annotations from direct transfer restrict the possible roles the global method adds. We expect, as an additional benefit of this combination, that the partial annotations from direct transfer together with the no-duplicate-role constraint described above will remedy problems related to diathesis alternations. Although the probabilities computed will favour the canonical alternation in general, the partial annotations may prevent a canonical analysis in a particular proposition. Consider the following alternation example: *Mary presented the flowers to John* vs. the less canonical alternation *Mary presented John with the flowers*. Although the most probable role

<sup>4</sup>We chose not to include the complete dependency path from predicate to argument because of data sparseness. We select the dependency label on the arc that points to the argument under discussion.

| Predicate identification and labelling |           |          |     |           |            |     |           |
|--|-----------|----------|-----|-----------|------------|-----|-----------|
|  |           | Labelled |     |           | Unlabelled |     |           |
|  |           | Prec     | Rec | F         | Prec       | Rec | F         |
| 1                                      | Direct    | 51       | 29  | 37        | 93         | 57  | 71        |
| 2                                      | Global    | 45       | 39  | 42        | 95         | 83  | 89        |
| 3                                      | Combined  | 45       | 45  | <b>45</b> | 92         | 91  | <b>91</b> |
| 4                                      | Plas11    | 68       | 25  | 37        | 98         | 36  | 53        |
| 5                                      | Plas11(f) | 56       | 46  | <b>51</b> | 97         | 80  | <b>87</b> |
| 6                                      | Manual    | 61       | 57  | 59        | 97         | 89  | 93        |

Table 1: Percent recall, precision and F-measure for predicate identification and labelling.

| Cross-lingual semantic role labelling |                             |    |
|---------------------------------------|-----------------------------|----|
| 1                                     | Direct                      | 35 |
| 2                                     | Global                      | 68 |
| 3                                     | Combined                    | 73 |
| 4                                     | Most frequent semantic role | 48 |

Table 2: Percent accuracy for semantic role labelling

for the *prep* relation would be A2, based on the canonical alternation, partial annotations on *Mary* (A0) and *John* (A2) in combination with the no-duplicate-role constraint would rule that out and the next most probable label would be put on *with*: A1.

## 6 Results and discussion

We ran experiments using the two global methods described in Section 5 separately and combined with direct transfer. In this section, we present the results and compare to several baselines and upper bounds from manual annotations and previous work.

### 6.1 Cross-lingual predicate labelling

Table 1 shows the results of cross-lingual predicate labelling (Labelled) and identification (Unlabelled). The first row shows the results from using the traditional direct transfer method. The second row presents results from the global method where we use cross-lingual WSD to label predicates. The third row combines direct and global transfer, as explained in Section 5.3. For comparison, we present results when using the parser from Van der Plas et al. (2011) on our test data: the fourth row contains results when using all (unfiltered) data, the fifth row when using data filtered for incomplete predicate labellings. We show an upper bound in the last row which corresponds to the inter-annotator agreement for manual annotation on a random set of 100 sentences (van der Plas et al., 2010).

Overall the figures, including the upper bound from manual annotations, are not very high. Annotating French verbs with English predicate labels is a hard task. When we look at the differences between the three automatic methods, we see that recall is very low (29%) for the direct method. From the recall figures for unlabelled predicates, we see that the direct method leaves many predicates without a label.

The global method has a much better recall, 39%, and a slightly lower precision. The best results are however attained when the two methods are combined, that is, when global transfer is used to fill in missing predicates from direct transfer. We get an F-measure of 45% which is a big improvement over the baseline of direct transfer, which attained 37%. These results show that the global method for predicate labelling improves recall without sacrificing precision too much.

We compare these results also to the results obtained by Van der Plas et al. (2011)’s three step model, where a parser trained on transferred annotations annotates in turn the test sentences. We see that the current method gives better results (recall and F-measure) when the parser is trained on unfiltered data. An aggressive filter, that removes more than half of the data and leads to a big drop in performance for argument labelling (recall that argument and predicate labelling is done in parallel in this model) finally leads to a result that outperforms ours. This result is not surprising because the parser has access to much more expressive syntax. Note that our global method only needs a PoS tagger in the source language and no syntactic information nor joint semantic-syntactic parsing frameworks. It is thus knowledge-lean and easier to apply to languages without a parser, a difference that should be taken into account in the interpretation of the results. However, we can learn from these results that structural information is beneficial. In future work, we plan to include word position information in our cross-lingual WSD method. This will give the method access to structural information while keeping it knowledge-lean.

In Figure 2, we give an example that illustrates the contribution of the global method. In this example,



**English (automatic):** There is in particular one amendment, let [let.01] me point [point.02] out, concerning [concern.01] the energy sector, which, in my capacity as rapporteur, I see [see.01] as particularly important.

**Transfer:** Il y a notamment un amendement, je le souligne, concernant [concern.01] le secteur de l'énergie, qui me paraît en tant que rapporteur particulièrement important.

**CLWSD:** Il y a notamment un amendement, je le souligne [stress.01], concernant [concern.01] le secteur de l'énergie, qui me paraît [seem.01] en tant que rapporteur particulièrement important.

Figure 2: Predicate label addition and correction using CLWSD.

the cross-lingual WSD method annotates more verbs than the direct transfer approach: labels [stress.01] and [seem.01] assigned during disambiguation, are missing from the first sentence after transfer. Even with a high quality word alignment, it would not be possible to get these labels from the English source sentence through direct transfer because they are simply not there, due to the non-literal translation. This example shows the limitations of token-to-token direct transfer and how the global method compensates for that by using information aggregated across the whole parallel corpus.

## 6.2 Global cross-lingual role labelling

Though already supported by previous work (Grenager and Manning, 2006; Lang and Lapata, 2010; Lang and Lapata, 2011), we tested the hypothesis that syntactic-semantic mappings provide good approximations for semantic role labelling, especially when adding predicate information. We therefore first ran a monolingual feasibility study by collecting counts from the CoNLL 2009 training set and testing on the CoNLL 2009 test set. The accuracy attained with this simple method is 79%. This shows that in a monolingual setting, the predicate label combined with the syntactic label of the argument are good predictors for the semantic role of the argument. This number can serve as a baseline for semantic role labelling given the correct predicate label.

In previous sections, we discussed diathesis alternations as problematic for using syntactic-semantic mapping rules. To measure the importance of diathesis alternations we need to measure the variation in a large corpus. By applying the mapping rules learned from the training data on the same data we get an idea of the amount of variation. We get an accuracy of 86%. Although the 14% probably contains the most interesting examples from a linguistic point of view, these results on monolingual data show that predicate-centered syntactic-semantic mapping rules are a promising direction for improving recall in direct transfer methods.

Table 2 shows the results<sup>5</sup> for semantic role labelling for the three cross-lingual transfer methods and the baseline of applying the most frequent semantic role label 'A1'. For the global and the combined methods we use the predicate labels provided by the cross-lingual WSD method. The numbers in Table 2 provide performance numbers given the predicate from the cross-lingual WSD method. We discussed in Subsection 5.2 that, when applying syntactic-semantic mapping rules cross-lingually, differences in annotation framework and cross-lingual divergences are at play. We see indeed that when applying syntactic-semantic mapping rules cross-lingually the accuracy drops from 79 to 68%. This drop in performance when applying to French the syntactic-semantic mappings that were learned on English data is not too important. This accuracy number is, in any case, much better than the results from direct transfer. This is mainly due to the low recall of direct transfer which results in very few but rather precise (87%) semantic roles. It is therefore very useful to use the direct transfer method as a backbone that restricts the labels we get from global transfer by imposing consistency with the available annotation (no-duplicate-argument-constraint). By combining the two methods, we get 73% accuracy that is not far from the 79% in the monolingual setting. In future work, we would like to investigate whether the drop in performance between the monolingual and cross-lingual setting is larger for languages that are less related.

We also compare our results to previous work on cross-lingual transfer of semantic roles. Kozhevnikov and Titov (2013) evaluate on the full test set described in Subsection 3 (1000 sentences), they use gold predicates instead of predicted predicates and evaluate on both core roles and adjuncts. The authors shared with us their results for core roles only: 74% and 77%, when using original and transferred

<sup>5</sup>As we focus on argument labelling (and not identification) we provide accuracy scores.

syntax, respectively. We use original syntax and should therefore compare to the 74%. When we use gold predicate annotations as Kozhevnikov and Titov (2013) did, instead of the predicate labels obtained through cross-lingual WSD, and test on all 1000 sentences, we attain 75% for the combined method and 71% for the global method. These results compare favourably with their results. This is encouraging because their model uses a larger feature set that includes (cross-lingual) lexical features, the unlabelled dependency graph and PoS information. Interestingly, they attain better scores when they use a transferred syntactic model instead of the original syntax. This result seems in line with our discussion on the loss of information when trying to map the English syntactic label inventory to the French inventory. We keep syntactic model transfer in mind for future work.

Because we consider the arguments as given, while Van der Plas et al. (2011) do both argument identification and labelling for all core roles and adjuncts, and provide precision and recall given the predicate only, we cannot directly compare to their results. We however include their results for the sake of completeness. Their parser results in 65% F-score.

Applying A1 (the most frequent semantic role) to the entire data set gives us 48% accuracy. That is much higher than results from transfer, again due to the low recall of the direct transfer method, but much lower than the results of the combined and global methods.

## 7 Conclusion

We have introduced a global approach to transfer that aggregates information at the corpus level thereby correcting and complementing the annotations from traditional direct transfer methods that suffer from token-level mistakes. We show that the combination of direct transfer (a high-precision method) and global methods (high in recall) outperforms previous results.

In contrast to previous work, we transfer predicate annotations and semantic role annotations by building two separate models tailored to the task at hand. We show how the predicate labels produced by our high-coverage model for cross-lingual predicate labelling are successfully used as predictors for semantic role labelling.

In future work, we would like to feed structural information to the cross-lingual WSD method such as information about word position, which would preserve its knowledge-lean character without need for syntactic parsing. Furthermore, we intend to use cross-lingual WSD for labelling adjuncts (non-core semantic roles) since this task is also rather lexical in nature. Last but not least, we want to add argument identification which will allow to propose a complete SRL annotation framework based on global information.

## Acknowledgements

This research was funded and supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) as part of the SFB 732.

## References

- A. Abeillé, L. Clément, and F. Toussanel. 2003. Building a treebank for French. In *Treebanks: Building and Using Parsed Corpora*. Kluwer Academic Publishers.
- P. Annesi and R. Basili. 2010. Cross-lingual alignment of FrameNet annotations through Hidden Markov Models. In *Proceedings of CICLing*.
- M. Apidianaki, G. Wisniewski, A. Sokolov, A. Max, and F. Yvon. 2012. WSD for n-best reranking and local language modeling in SMT. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 1–9, Jeju, Republic of Korea, July. Association for Computational Linguistics.
- M. Apidianaki. 2009. Data-driven Semantic Analysis for Multilingual WSD and Lexical Selection in Translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*, pages 77–85, Athens, Greece.

- R. Basili, D. De Cao, D. Croce, B. Coppola, and A. Moschitti, 2009. *Computational Linguistics and Intelligent Text Processing*, chapter Cross-Language Frame Semantics Transfer in Bilingual Corpora, pages 332–345. Springer Berlin / Heidelberg.
- M.-H. Candito, B. Crabbé, P. Denis, and F. Guérin. 2009. Analyse syntaxique du français : des constituants aux dépendances. In *Proceedings of la Conférence sur le Traitement Automatique des Langues Naturelles (TALN'09)*, Senlis, France.
- M. Carpuat and D. Wu. 2007. Improving Statistical Machine Translation using Word Sense Disambiguation. In *Proceedings of the Joint EMNLP-CoNLL Conference*, pages 61–72, Prague, Czech Republic.
- D. Das and S. Petrov. 2011. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 600–609, Portland, Oregon, USA, June. Association for Computational Linguistics.
- B. Dorr. 1994. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20(4):597–633.
- T. Grenager and C. Manning. 2006. Unsupervised discovery of a statistical verb lexicon. In *Proceedings of EMNLP*.
- J. Hajič, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štěpánek, P. Straňák, M. Surdeanu, N. Xue, and Y. Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009)*.
- J. Henderson, P. Merlo, G. Musillo, and I. Titov. 2008. A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of CONLL 2008*, pages 178–182.
- R. Hwa, P. Resnik, A. Weinberg, C. Cabezas, and O. Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11:311–325.
- P. Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit X*, pages 79–86, Phuket, Thailand.
- M. Kozhevnikov and I. Titov. 2013. Crosslingual transfer of semantic role models. In *In Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August. Association for Computational Linguistics.
- J. Lang and M. Lapata. 2010. Unsupervised induction of semantic roles. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 939–947, Los Angeles, California, June. Association for Computational Linguistics.
- J. Lang and M. Lapata. 2011. Unsupervised semantic role induction via split-merge clustering. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.
- M. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Comp. Ling.*, 19:313–330.
- P. Merlo and L. van der Plas. 2009. Abstraction and generalisation in semantic role labels: PropBank, VerbNet or both? In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 288–296, Suntec, Singapore.
- A. Meyers. 2007. Annotation guidelines for NomBank - noun argument structure for PropBank. Technical report, New York University.
- R. Navigli. 2009. Word Sense Disambiguation: a Survey. *ACM Computing Surveys*, 41(2):1–69.
- H. T. Ng, B. Wang, and Y. S. Chan. 2003. Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 455–462, Sapporo, Japan.
- F. J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29:19–51.
- S. Padó. 2007. *Cross-lingual Annotation Projection Models for Role-Semantic Information*. Ph.D. thesis, Saarland University.

- M. Palmer, D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31:71–105.
- V. Punyakanok, D. Roth, and W. Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287.
- P. Resnik and D. Yarowsky. 2000. Distinguishing Systems and Distinguishing Senses: New Evaluation Methods for Word Sense Disambiguation. *Natural Language Engineering*, 5(3):113–133.
- H. Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK, September. <http://www.ims.uni-stuttgart.de/~schmid/>.
- O. Täckström, D. Das, S. Petrov, R. McDonald, and J. Nivre. 2013. Token and type constraints for cross-lingual part-of-speech tagging. In *Transactions of the ACL*. Association for Computational Linguistics, March.
- I. Titov and J. Henderson. 2007. A latent variable model for generative dependency parsing. In *Proceedings of the International Conference on Parsing Technologies (IWPT-07)*, pages 144–155, Prague, Czech Republic.
- I. Titov, J. Henderson, P. Merlo, and G. Musillo. 2009. 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)*, Pasadena, California, July.
- L. van der Plas, T. Samardžić, and P. Merlo. 2010. Cross-lingual validity of PropBank in the manual annotation of French. In *In Proceedings of the 4th Linguistic Annotation Workshop (The LAW IV)*, Uppsala, Sweden.
- L. van der Plas, P. Merlo, and J. Henderson. 2011. Scaling up cross-lingual semantic annotation transfer. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics and the Human Language Technologies conference*.
- D. Yarowsky and G. Ngai. 2001. Inducing multilingual pos taggers and np bracketers via robust projection across aligned corpora. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, NAACL '01*, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.