Cross-lingual Word Sense Disambiguation for Predicate Labelling of French

Lonneke van Der Plas, Marianna Apidianaki

Cross-lingual Word Sense Disambiguation for Predicate Labelling of French

Lonneke van der Plas[1] Marianna Apidianaki[2]
(1) IMS, Pfaffenwaldring 5B, 70569 Stuttgart, Germany
(2) LIMSI-CNRS, Rue John von Neumann, Campus Universitaire d'Orsay
Bât 508, 91405 Orsay Cedex, France
Lonneke.vanderPlas@ims.uni-stuttgart.de, Marianna.Apidianaki@limsi.fr

Résumé. Nous abordons la question du transfert d'annotations sémantiques, et plus spécifiquement d'étiquettes sur les prédicats, d'une langue à l'autre sur la base de corpus parallèles. Des travaux antérieurs ont transféré ces annotations directement au niveau des tokens, conduisant à un faible rappel. Nous présentons une approche globale de transfert qui agrège des informations repérées dans l'ensemble du corpus parallèle. Nous montrons que la performance de la méthode globale est supérieure aux résultats antérieurs en termes de rappel sans trop affecter la précision.

Abstract. We address the problem of transferring semantic annotations, more specifically predicate labellings, from one language to another using parallel corpora. Previous work has transferred these annotations directly at the token level, leading to low recall. We present a global approach to annotation transfer that aggregates information across the whole parallel corpus. We show that this global method outperforms previous results in terms of recall without sacrificing precision too much.

Mots-clés : transfert inter-langue, annotation sémantique automatique, prédicats, désambiguïsation lexicale, corpus parallèles.

Keywords: cross-lingual transfer, automatic semantic annotation, predicates, Word Sense Disambiguation, parallel corpora.

1 Introduction

There has recently been a large interest in multilingual natural language processing. Several annotation efforts have been devoted to developing resources for different languages, needed for supervised learning (Hajič *et al.*, 2009). However, there is a large number of languages that still lack linguistically annotated resources. For example for French, one of the most important European languages, there exists no corpus with predicate-argument annotations.

Predicate-argument annotations are the representation of the grammatically relevant aspects of a sentence meaning. This level of analysis provides a means of expressing a relation between syntactically different sentences, such as the sentence with the transitive verb in (1a) and the one with the intransitive verb in (1b). The semantic label *theme* expresses the fact that the object in (1a) has the same conceptual relation with the verb as the subject in (1b).

(1) a. [AGENT Mary] [REL-STOP.01 stopped] [THEME the car].
b. [THEME The car] [REL-STOP.01 stopped].

Semantic parsing or semantic role labelling refers to the task of automatically labelling predicates and arguments with predicate-argument structure. This task can be divided in two parts. One part refers to the labelling of the predicates with a predicate sense, and the other to the labelling of the arguments with semantic roles. Semantic parsers and the results of cross-lingual annotation transfer are therefore evaluated on both tasks : predicate labelling and role labelling. We focus on predicate labelling in this paper.

Cross-linguistically, the predicate-argument structure of a sentence is considered to be more stable than its syntactic form. The English sentence in (2a) can be considered as equivalent to the French sentence in (2b), despite the fact that the

position of their syntactic subject is occupied by different kinds of lexical elements and that the complements of the verbs differ both syntactically and semantically.

(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)

Since manual annotation is a costly and time-consuming approach to resource development, cross-lingual annotation transfer offers an attractive alternative. In this approach, predicate-argument annotations on a source language, for which there exists semantic annotations, are transferred to a target language using parallel corpora (Padó, 2007; Basili *et al.*, 2009; Annesi & Basili, 2010; van der Plas *et al.*, 2011). The increased stability of predicate-argument structures across languages makes their cross-lingual transfer attractive when compared with, for example, syntactic structures.

Traditional methods for cross-lingual transfer rely on the semantic equivalence of the original and the translated sentences, and on correct and complete alignments between words or constituents in those sentences. Since the semantic annotations are transferred directly from token to token, we will refer to these traditional methods as direct cross-lingual transfer. Alignment errors and translation shifts represent major sources of mistakes in the direct transfer approach which result in incorrect and incomplete annotations in the target language.

In this paper, we propose a different strategy to predicate labelling that is less sensitive to alignment errors and translation shifts. Instead of transferring semantic annotations on a token-to-token basis, we aggregate information across the whole parallel corpus to correct token-level mistakes resulting from direct cross-lingual transfer. Our approach consists of two steps : In the learning step, a global model is learned on the basis of source language (English) predicate annotations in a word-aligned English-French training corpus. In the labelling step, this model assigns predicate labels to verbs in target language texts (here, French). 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 contributions are two-fold. First, we present a global approach to semantic annotation transfer that corrects token-level mistakes as found in traditional direct transfer methods. Second, we show the strengths and limitations of global vs direct transfer.

In the next section, we present related work on cross-lingual annotation transfer. We then briefly discuss the semantic annotation framework we are using and explain why we decided to use an English semantic framework for annotating French. In Section 4, we briefly present the direct transfer method. In Section 5, we describe the tools and data we use and explain the adopted evaluation framework. In Section 6, we explain the global method proposed in this paper. The results are presented in Section 7, before concluding.

2 Related work

Transferring annotations from one language to another in order to train monolingual tools for new languages was first introduced by Yarowsky & 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 & Petrov (2011) propose a graph-based framework for projecting syntactic information across language boundaries. 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 by Täckström *et al.* (2013) who couple token and type constraints in order to guide learning. These two types of information are viewed as complementary : token-level projections offer precise constraints for tagging in a particular context while broad coverage type-level dictionaries help to filter noise in token-level projections. Our approach to cross-lingual predicate 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 method which is trained on the entire corpus. We, however, are concerned with semantic annotations and not PoS tags.

Transfer of semantic annotation has started off with direct transfer of FrameNet semantic annotations (Padó, 2007; Basili *et al.*, 2009; Annesi & 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 and additional syntactic annotations produces better parses than the input it received by aggregating information

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 *pay*.

across multiple examples. Our method is more resource-light, as we do not need syntactic annotations neither on the target nor on the source side.

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 & Titov (2013) use shared feature representations (syntactic and lexical) to adapt a source model to a target-language model. They, however, do not consider the task of predicate labelling but only semantic role labelling.

In this work, we address predicate labelling in languages other than English as a cross-lingual WSD task. Word sense disambiguation is the task of automatically identifying the meaning of words in context (Navigli, 2009). In its cross-lingual variant, the candidate senses are the words' translations in other languages and WSD aims at predicting semantically correct translations for instances of the words in context (Resnik & Yarowsky, 2000; Ng *et al.*, 2003; Carpuat & Wu, 2007; Apidianaki, 2009). In our experiments, we apply the cross-lingual WSD method employed by Apidianaki *et al.* (2012) for improving the quality of Machine Translation, in a different setting : instead of assigning semantically appropriate translations to words in context, the WSD classifier serves for selecting the most adequate English predicate label for verbs in a new language. More details on the adaptation of the method to predicate labelling are given below, in Section 6 of the paper.

3 The semantic annotation framework

There exist three frameworks for annotating corpora with predicate-argument structure : FrameNet (Fillmore *et al.*, 2003), VerbNet (Kipper, 2005) and PropBank (Palmer *et al.*, 2005). We chose PropBank for applying our predicate labelling method. In the following, we describe PropBank in more detail as well as the criteria that led to choosing this framework as well as how the English PropBank can be used to annotate French.

3.1 The Proposition Bank

The Proposition Bank (PropBank) is a linguistic resource that contains information on the semantic structure of sentences (Palmer *et al.*, 2005). It consists of a one-million-word corpus of naturally occurring sentences annotated with semantic structures and a lexicon (the PropBank frame files) that lists all the predicates (verbs) that can be found in the annotated sentences and the sets of semantic roles they introduce.

Predicates are marked with labels that specify the sense of a verb in a particular context. Each lemma described in the frame files (3300 verbs) contains one or more lexemes (4500 verb senses), which are used as predicate labels. The PropBank frame files specify the interpretation of the roles for each verb in its different senses. The interpretation of the numbered roles is given for each lexeme separately. Table 1 illustrates the entry for the verb *pay* in the PropBank frame files.

The semantic role annotation is based on Dowty's theory of Proto-Roles (Dowty, 1991). Arguments are marked with the labels A0 to A5, which represent semantic roles of a very general kind. Only the labels A0 and A1 have approximately the same value with all verbs : they are used to mark instances of proto-agent (A0) and proto-patient (A1). The meaning of other numbered arguments is verb-dependent. It depends on the meaning of the verb, on the type of the constituent they are assigned to, and on the number of roles present in a particular sentence. A3, for example, can mark purpose as is the case in (3), or it can mark direction or some other role with other verbs. The indices are assigned according to the roles' prominence in the sentence. More prominent are the roles that are more closely related to the verb. The AM-* labels can be specified further as : location, cause, extent, time, discourse connectives, purpose, general purpose, manner, direction. The labels for adjuncts are more specific but less verb-specific than the labels for arguments, and they do not depend on the presence of other roles in the sentence.

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

Although PropBank is considered as the most language-specific of the three resources, Samardžić *et al.* (2010) motivate the use of PropBank for annotation transfer with two reasons. First, the lexicon in this resource is corpus-driven. It is built by extracting and describing all the predicates that occur in a predefined sample of naturally occurring sentences. Since the aim is to annotate a corpus of naturally occurring sentences exhaustively, we expect that such a lexicon can provide a better coverage than the lexicon in FrameNet and VerbNet, which are not corpus-driven. Second, the labels used in PropBank both for predicates and arguments involve fewer theoretical assumptions than the labels in FrameNet. While the FrameNet labels capture mostly linguistic intuition at the targeted level of lexical semantics and the relations between the lexical items, the PropBank labels rely strongly on the observable behaviour of words. The distinction between the different verb senses, for instance, is made taking into account the different sets of arguments and other observable differences, such as the presence of the particle that distinguishes pay.04 from pay.05 in Table 1. This approach can be expected to provide more tangible criteria for annotators in deciding how to annotate each instance of the predicate-argument structure found in the corpus, ensuring a more reliable and more consistent annotation. Also, it enables a more direct comparison of the structures across languages, since the representation of the structures does not include any hypothesized levels of abstraction.

3.2 Using the English PropBank to annotate French

Adapting a semantic framework to a new language is a time-consuming process. In order to generate broad-coverage annotations for a target language in limited time, we transfer semantic annotations from the source language directly to the target language without adapting the semantic annotation framework to the target language. This means that French verbs will be annotated with English predicate senses. For this to work, we need to show that PropBank is cross-lingually valid. Van der Plas et al. (2010) did this in a manual annotation effort. For a complete description of the annotation procedure, that involved four evaluators and several stages, we point the reader to Van der Plas et al. (2010). In this section, we summarize the procedure and briefly discuss the main outcomes.

In this manual annotation study, annotators used the English PropBank frame files to annotate French sentences. This means that for every predicate found in a French sentence, they needed to translate it and find an English verb sense that was applicable to the French verb. If an appropriate entry could not be found in the frame files for a given predicate, the annotators were instructed to use the dummy label for the predicate and fill in the roles according to their own insights.

Some of the differences in annotation observed between annotators were due to lexical variation in English. For example, if one annotator put the label 'demonstrate.01' whereas the other used 'show.01' on an instance of the verb *montrer*, this should not be counted as a disagreement as the two senses are linked. Therefore, agreement scores were provided both on the basis of the predicate sense labels and on the basis of the verb class, using the verb classifications from VerbNet (Kipper, 2005) and the mapping to PropBank labels as given in the type mappings of the SemLink project [1] (Loper *et al.*, 2007). VerbNet is a hierarchically organised verb lexicon of English verbs (Kipper, 2005). It is organized into verb classes extending Levin (1993) classes through refinement and addition of subclasses to achieve syntactic and semantic coherence among members of a class. The senses 'demonstrate.01' and 'show.01' appear in the same verb class according to the type mappings of the SemLink project and were thus considered as correct.

The average inter-annotator agreement reported in these experiments was relatively low when the annotations on the PropBank verb sense level were compared : 59%. However, at the level of verb classes, the inter-annotator agreement

1. http://verbs.colorado.edu/semlink/

increased to 81%. The authors identify collocations and idiomatic expressions as the main sources of disagreement in predicate labellings among annotators, as is also shown in studies on other language pairs (Burchardt *et al.*, 2009).

For a single annotator, the main measure of cross-lingual validity was the percentage of dummy predicates in the annotation. Van der Plas (2010) found 130 dummy annotations in 1000 sentences. A manual classification of the dummy labels showed that the dummy label was mainly used for French multi-word expressions (82%), most of which could be translated by a single English verb (47%) whereas others could not because they were translated by a combination that included a form of 'be' that was not annotated in PropBank (25%). The 47% of multi-word expressions that received the dummy label showed the annotator's reluctance to put a single verb label on a French multi-word expression. The annotation guidelines could however be adapted to instruct annotators not to hesitate in such cases. Based on these findings, the authors conclude that the annotation framework PropBank is cross-lingually valid.

4 Direct cross-lingual transfer

Before presenting our global method for predicate labelling, we would like to remind the reader of the method for direct cross-lingual transfer which is used in comparisons and combinations throughout this paper. It is taken from Van der Plas et al. (2011), but we give a short summary here for the readers convenience. The method is based on the Direct Correspondence Assumption for syntactic dependency trees by Hwa *et al.* (2005).

> **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 properties that are transferred through DST are predicate senses. The relationships that are transferred are semantic role dependencies, but we are not concerned with them in this paper. The properties are transferred from the English side of a parallel corpus that is automatically annotated with syntactic-semantic analyses to the foreign language side, as described in the following section.

5 Tools and data

The parallel corpus used in our experiments is the English-French part of the Europarl corpus (Koehn, 2005). 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 [2] (Meyers, 2007)). In the experiments presented in this paper, we only use the predicate labels found in the English part of Europarl, omitting the labels assigned to the arguments.

For the direct transfer of semantic annotations from the English to the French side of the parallel corpus, we use the method described in Section 4. As is usual practice in pre-processing for automatic word alignment, both parts of the parallel corpus were 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, were considered. We subsequently word aligned the English and French sentences automatically using GIZA++ (Och & Ney, 2003) in both translation directions and retained only intersecting alignments. Furthermore, because translation shifts are known to pose problems for the automatic projection of semantic annotation across languages (Padó, 2007), we selected only those parallel sentences in Europarl that are direct translations from English to French, or vice versa. In the end, we obtained a word-aligned parallel corpus of 276-thousand sentence pairs.

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 879 predicates.

2. We limit our experiments to verbal predicates only because the semantic annotations on French test sentences are limited to verbal predicates. Even though verbal predicates in the target language can be expressed as non-verbal predicates in the source language, the transfer of Nombank labels to verbal predicates is not straightforward due to difficulties in mapping between the two annotation frameworks.

6 Global cross-lingual predicate labelling

Traditional cross-lingual transfer methods are locally defined. Transfer takes place on a token-to-token basis and, as a consequence, missing or incorrect alignments lead immediately to missing and incorrect annotations in the target language. Our method for cross-lingual predicate labelling is globally defined and relies less on actual alignments.

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, assigning the correct English predicate label to a French verb is a task very close to word sense disambiguation (WSD), which aims at automatically identifying the meaning of words in context (Navigli, 2009). In cross-lingual WSD, the candidate senses of words are their translations in other languages from which the most adequate has to be selected for contextualized instances of the words (Carpuat & 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 that we propose 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 corpus. At labelling time, we label French verbs with English predicate labels without need for parallel data or alignments. The method is language-independent and only requires minimal linguistic resources.

In contrast to direct transfer, we provide a predicate label for all French verbs in the test set, not only aligned ones. We expect to augment the recall when using global estimates and hope that the affect on precision is not too negative.

6.1 Pre-processing

The WSD classifier is trained on the Europarl corpus tagged with PropBank information on the English side, as described in Section 5. To identify the sets of candidate predicate labels for each French verb, we replace the English verbs by the corresponding predicate label wherever this is available. Then we tag both parts of the corpus by part of speech (PoS) using the TreeTagger (Schmid, 1994) and rebuild the parallel files (one sentence per line) by replacing words on both sides by the corresponding 'lemma_PoS tag' pair, and keeping the predicate labels in the place of English verbs. The corpus is then aligned at the word level in both directions using GIZA++ (Och & Ney, 2003) and a lexicon is built from intersecting alignments. Lexicon entries for French verbs contain the 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 a high alignment confidence score according to GIZA++ and experiment with two thresholds, 0.01 and 0.001. Naturally, the second threshold retains a higher number of candidate labels.

6.2 Learning

For each French verb (v) in the lexicon, we want to be able to identify its correct predicate label in a new context. A feature vector is built for each candidate label following the procedure described in Apidianaki *et al.* (2012). For each candidate label L_i of a French verb v, we extract the content word co-occurrences of v in the sentences where it translates an English verb tagged with the label L_i. The retained French words constitute the features of the vector built for the label. Let N be the number of features retained for each label L_i of v from the corresponding French contexts. Each feature F_j ($1 \leq j \leq N$) receives a total weight with the label $\text{tw}(F_j, L_i)$ 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 labels (L_i's) to which F_j is related, and of the probabilities (p_{ij}) that F_j co-occurs with instances of v corresponding to each of the L_i's :

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

Each p_{ij} is computed as the ratio of the co-occurrence count of F_j with v when it corresponds to a label L_i to the total number of features (N) seen with L_i in the corpus :

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

The local weight $\text{lw}(F_j, L_i)$ between a feature F_j and a label L_i directly depends on their co-occurrence count :

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

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 (tw) with a label when it appears frequently in the corresponding French contexts and rarely in the contexts of the other labels.

6.3 Labelling

Predicate identification 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*.[3] The most suitable predicate labels are then assigned to the retained French verbs by our disambiguation classifier. The weighted feature vectors built for the candidate labels of a French verb as described in the previous section are compared to the context of a new instance of the verb and an association score is assigned to each candidate label. To facilitate comparison with the vectors, the new contexts (sentences) are lemmatized 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 are found between the new context and the vector of a label, their association score corresponds to the mean of the weights of their shared features with that label (i.e. found in its vector). In Equation 4, $(CF_j)_{j=1}^{|CF|}$ is the set of common features between the label vector V_i and the new context C and tw is the weight of a CF with label L_i, computed as explained in the previous section.

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

The label that receives the highest association score with the new context is returned and serves to annotate the corresponding French verb. For example, among the candidate labels for the verb *encourager* ({urge.01, foster.01, stimulate.01, promote.02, encourage.01, encourage.02, renew.01}), the classifier selects the predicate label *encourage.02* for the following instance :

> D'ailleurs, le rapport von Wogau, que vient de voter le Parlement européen *encourage* [encourage.02] en ce sens.

The label selected in this case corrects the label [support.01] that was assigned through direct transfer.

7 Results and discussion

We run experiments using the global method for predicate labelling described in the previous section and compare the results to the ones obtained through direct transfer. The results are presented in Table 2 where they are also compared to upper bounds from manual annotations and previous work.

The first row of Table 2 shows the results from using the traditional direct transfer method. The second and third rows present the results obtained using the global method, where we use cross-lingual WSD to label predicates. In row 2, we present the results obtained when using an alignment confidence threshold of 0.01 (retaining labels with an alignment score above 0.01, according to GIZA++) and in row 3, the results obtained using a threshold of 0.001. For comparison, we show results when using a parser as in Van der Plas et al. (2011) who use a joint syntactic-semantic parser and

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

	Predicate senses			Verb classes		
	P	R	F	P	R	F
Direct Transfer	51	29	37	75	38	50
CLWSD(0.01)	45	39	42	73	61	67
CLWSD(0.001)	42	40	41	70	65	67
Parser	56	46	51	83	63	72
Manual	61	57	59	85	76	81

TABLE 2 – Percent recall, precision and F-measure for predicate labelling.

syntactic annotations on French to do predicate and semantic role labelling. [4] We show an upper bound in the last row. This represents the inter-annotator agreement for manual annotation on a random set of 100 sentences taken from data provided by Van der Plas et al. (2010). Two evaluation settings were used in that work, in order to avoid penalizing synonymous verb senses assigned by the annotators : the inter-annotator agreement reached at the verb sense label was compared to the agreement reached using verb classes, as explained in Section 3.2. Because we do not want to penalize the predicate labelling system for selecting verb senses that are synonyms of the verb senses in the gold, we follow the same strategy and take verb classes into account during evaluation. More precisely, we calculate scores based on the exact correspondence between the label proposed by our system and the gold label found in the test data, and we also perform a coarser evaluation taking verb classes into account. The first evaluation is too strict as it penalizes the system when it selects a predicate sense that is synonymous to the gold sense or a predicate that belongs to the same word class. The results headed by verb classes evaluate on a more realistic basis, capturing semantic correspondences beyond surface variations. In this setting, predicate labels are correct if they belong to the same verb class as the predicate in the gold annotations.

When we look at the differences between the three automatic methods for the evaluation on predicate senses, we see that for the direct transfer method especially recall is very low, 29%. The global method (alignment score threshold = 0.01) has a much better recall, 39%. Precision is lower but the F-score increases by 5 percentage points. When the alignment confidence threshold is lowered to 0.001, which means that more candidate labels are retained, recall increases and precision goes down, as expected. As explained above, the parser from Van der Plas et al. (2011) (shown in row 4) has access to both PoS and syntactic information on the target side and uses a joint syntactic-semantic framework. When we take its dependence on two extra resources into consideration, its performance is not that impressive. However, we can learn from these results that structural information is beneficial. Nevertheless, our results show that the cross-lingual WSD method which relies on much less external knowledge, outperforms the direct method on both senses and verb classes. 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 staying knowledge-lean.

When we look at the results using verb classes which permit to abstract from surface variations and capture semantic correspondences, the overall performance numbers are higher as expected. More importantly, the difference in performance between the cross-lingual WSD method using the more restrained set of labels (alignment score threshold = 0.01), which performs best, and the direct transfer method are now much larger (three times as important, from 5 to 17 percentage points) whereas the difference between our method and the parser is further reduced (from 10 to 5 percentage points). The differences between the parser and our method can be mainly attributed to arbitrary variations between predicate labels that belong to the same verb class. When we look at the precision and recall scores we see that the cross-lingual WSD method with the threshold set at '0.01' improves the recall of the direct transfer method by 23 percentage points, whereas precision only drops by 2 points. The cross-lingual WSD method that needs only a PoS tagger on the target and no syntactic annotation nor parsing frameworks, results in much better scores than direct transfer that is equally knowledge-light. All automatic methods are still quite far from the results from manual annotation.

In summary, these results show that the global cross-lingual WSD method for predicate labelling improves recall of direct transfer methods without sacrificing precision too much. In future work, we plan to combine the direct transfer method and the global cross-lingual WSD method, because the two are complementary in terms of recall and precision.

An example of predicate labelling might help the reader to get an idea of the contribution of the global method. In Table 3, we present an example where cross-lingual WSD annotates more verbs than the direct transfer : labels [stress.01] and [seem.01], assigned during disambiguation, are missing from the first sentence after transfer. Moreover, it would be impossible to get these labels through direct transfer from the English source sentence because they are simply not there,

4. The results are different from the results reported in Van der Plas et al. (2011) because we used the development set in our evaluations.

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.

TABLE 3 – Predicate label addition and correction using CLWSD.

due to the non-literal translation. This example shows the limitations of token-to token, direct transfer and how the global method is able to compensate for that by using information aggregated across the whole parallel corpus.

8 Conclusion

In this paper, we present a knowledge-light global approach to the cross-lingual transfer of semantic annnotation that aggregates information across the whole parallel corpus. Previous work has transferred annotations directly from token to token in parallel sentences leading to low recall and token-level mistakes. We show how the global method, based on cross-lingual word sense disambiguation, improves recall by a large margin without sacrificing precision too much.

Given the knowledge-lean character of the proposed method, in future work we plan to apply it for cross-lingual predicate labelling in other language pairs. Furthermore, we would like to include target side structural information (e.g. word position information) in the cross-lingual WSD method. Last but not least, we intend to work towards global methods for role identification and labelling which will allow to propose a complete SRL annotation framework based on global information. In that respect, we would also like to try and combine global and direct methods because the two seem complementary in terms of recall and precision.

Références

ANNESI P. & BASILI R. (2010). Cross-lingual alignment of FrameNet annotations through Hidden Markov Models. In *Proceedings of CICLing*.

APIDIANAKI M. (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)*, p. 77–85, Athens, Greece.

APIDIANAKI M., WISNIEWSKI G., SOKOLOV A., MAX A. & YVON F. (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*, p. 1–9, Jeju, Republic of Korea : Association for Computational Linguistics.

BASILI R., CAO D. D., CROCE D., COPPOLA B. & MOSCHITTI A. (2009). *Computational Linguistics and Intelligent Text Processing*, chapter Cross-Language Frame Semantics Transfer in Bilingual Corpora, p. 332–345. Springer Berlin / Heidelberg.

BURCHARDT A., ERK K., FRANK A., KOWALSKI A., PADO S. & PINKAL M. (2009). *Multilingual FrameNets in Computational Lexicography : Methods and Applications*, chapter FrameNet for the semantic analysis of German : Annotation, representation and automation, p. 209–244. De Gruyter Mouton, Berlin.

CARPUAT M. & WU D. (2007). Improving Statistical Machine Translation using Word Sense Disambiguation. In *Proceedings of the Joint EMNLP-CoNLL Conference*, p. 61–72, Prague, Czech Republic.

DAS D. & PETROV S. (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*, p. 600–609, Portland, Oregon, USA : Association for Computational Linguistics.

DOWTY D. (1991). Thematic proto-roles and argument selection. *Language*, **67.3**, 547–619.

FILLMORE C. J., JOHNSON R. & PETRUCK M. (2003). Background to FrameNet. *International journal of lexicography*, **16.3**, 235–250.

HAJIČ J., CIARAMITA M., JOHANSSON R., KAWAHARA D., MARTÍ M. A., MÀRQUEZ L., MEYERS A., NIVRE J., PADÓ S., ŠTEPÁNEK J., STRAŇÁK P., SURDEANU M., XUE N. & ZHANG Y. (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)*.

HENDERSON J., MERLO P., MUSILLO G. & TITOV I. (2008). A latent variable model of synchronous parsing for syntactic and semantic dependencies. In *Proceedings of CONLL 2008*, p. 178–182.

HWA R., RESNIK P., A.WEINBERG, CABEZAS C. & KOLAK O. (2005). Bootstrapping parsers via syntactic projection accross parallel texts. *Natural language engineering*, **11**, 311–325.

KIPPER K. (2005). *VerbNet : A broad-coverage, comprehensive verb lexicon*. PhD thesis, University of Pennsylvania.

KOEHN P. (2005). Europarl : A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit X*, p. 79–86, Phuket, Thailand.

KOZHEVNIKOV M. & TITOV I. (2013). Crosslingual transfer of semantic role models. In *In Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria : Association for Computational Linguistics.

LEVIN B. (1993). *English Verb Classes and Alternations : A preliminary investigation*. Rapport interne, University of Chicago Press.

LOPER E., YI S.-T. & PALMER M. (2007). Combining lexical resources : Mapping between PropBank and VerbNet. In *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*, p. 118–129, Tilburg, The Netherlands.

MARCUS M., SANTORINI B. & MARCINKIEWICZ M. (1993). Building a large annotated corpus of English : the Penn Treebank. *Comp. Ling.*, **19**, 313–330.

MEYERS A. (2007). *Annotation guidelines for NomBank - noun argument structure for PropBank*. Rapport interne, New York University.

NAVIGLI R. (2009). Word Sense Disambiguation : a Survey. *ACM Computing Surveys*, **41**(2), 1–69.

NG H. T., WANG B. & CHAN Y. S. (2003). Exploiting Parallel Texts for Word Sense Disambiguation : An Empirical Study. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, p. 455–462, Sapporo, Japan.

OCH F. J. & NEY H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, **29**, 19–51.

PADÓ S. (2007). *Cross-lingual Annotation Projection Models for Role-Semantic Information*. PhD thesis, Saarland University.

PALMER M., GILDEA D. & KINGSBURY P. (2005). The Proposition Bank : An annotated corpus of semantic roles. *Computational Linguistics*, **31**, 71–105.

RESNIK P. & YAROWSKY D. (2000). Distinguishing Systems and Distinguishing Senses : New Evaluation Methods for Word Sense Disambiguation. *Natural Language Engineering*, **5**(3), 113–133.

SAMARDŽIĆ T., VAN DER PLAS L., KASHAEVA G. & MERLO P. (2010). The scope and the sources of variation in verbal predicates in English and French. In *Proceedings of the 9th International Workshop on Treebanks and Linguistic Theories*.

SCHMID H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, p. 44–49, Manchester, UK. http ://www.ims.uni-stuttgart.de/~schmid/.

TÄCKSTRÖM O., DAS D., PETROV S., MCDONALD R. & NIVRE J. (2013). Token and type constraints for cross-lingual part-of-speech tagging. In *Transactions of the ACL* : Association for Computational Linguistics.

TITOV I., HENDERSON J., MERLO P. & MUSILLO G. (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.

VAN DER PLAS L., MERLO P. & HENDERSON J. (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*.

VAN DER PLAS L., SAMARDŽIĆ T. & MERLO P. (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.

YAROWSKY D. & NGAI G. (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, p. 1–8, Stroudsburg, PA, USA : Association for Computational Linguistics.