Cross-lingual Transfer of Semantic Role Labeling Models

Mikhail Kozhevnikov and **Ivan Titov**
Saarland University, Postfach 15 11 50
66041 Saarbrücken, Germany
{mkozhevn|titov}@mmci.uni-saarland.de

Abstract

Semantic Role Labeling (SRL) has become one of the standard tasks of natural language processing and proven useful as a source of information for a number of other applications. We address the problem of transferring an SRL model from one language to another using a shared feature representation. This approach is then evaluated on three language pairs, demonstrating competitive performance as compared to a state-of-the-art unsupervised SRL system and a cross-lingual annotation projection baseline. We also consider the contribution of different aspects of the feature representation to the performance of the model and discuss practical applicability of this method.

1 Background and Motivation

Semantic role labeling has proven useful in many natural language processing tasks, such as question answering (Shen and Lapata, 2007; Kaisser and Webber, 2007), textual entailment (Sammons et al., 2009), machine translation (Wu and Fung, 2009; Liu and Gildea, 2010; Gao and Vogel, 2011) and dialogue systems (Basili et al., 2009; van der Plas et al., 2009).

Multiple models have been designed to automatically predict semantic roles, and a considerable amount of data has been annotated to train these models, if only for a few more popular languages. As the annotation is costly, one would like to leverage existing resources to minimize the human effort required to construct a model for a new language.

A number of approaches to the construction of semantic role labeling models for new languages have been proposed. On one end of the scale is unsupervised SRL, such as Grenager and Manning (2006), which requires some expert knowledge, but no labeled data. It clusters together arguments that should bear the same semantic role, but does not assign a particular role to each cluster. On the other end is annotating a new dataset from scratch. There are also intermediate options, which often make use of similarities between languages. This way, if an accurate model exists for one language, it should help simplify the construction of a model for another, related language.

The approaches in this third group often use parallel data to bridge the gap between languages. Cross-lingual annotation projection systems (Padó and Lapata, 2009), for example, propagate information directly via word alignment links. However, they are very sensitive to the quality of parallel data, as well as the accuracy of a source-language model on it.

An alternative approach, known as cross-lingual model transfer, or cross-lingual model adaptation, consists of modifying a source-language model to make it directly applicable to a new language. This usually involves constructing a shared feature representation across the two languages. McDonald et al. (2011) successfully apply this idea to the transfer of dependency parsers, using part-of-speech tags as the shared representation of words. A later extension of Täckström et al. (2012) enriches this representation with cross-lingual word clusters, considerably improving the performance.

In the case of SRL, a shared representation that is purely syntactic is likely to be insufficient, since structures with different semantics may be realized by the same syntactic construct, for example "in August" vs "in Britain". However with the help of recently introduced cross-lingual word represen-

tations, such as the cross-lingual clustering mentioned above or cross-lingual distributed word representations of Klementiev et al. (2012), we may be able to transfer models of shallow semantics in a similar fashion.

In this work we construct a shared feature representation for a pair of languages, employing cross-lingual representations of syntactic and lexical information, train a semantic role labeling model on one language and apply it to the other one. This approach yields an SRL model for a new language at a very low cost, effectively requiring only a source language model and parallel data.

We evaluate on five (directed) language pairs – EN-ZH, ZH-EN, EN-CZ, CZ-EN and EN-FR, where EN, FR, CZ and ZH denote English, French, Czech and Chinese, respectively. The transferred model is compared against two baselines: an unsupervised SRL system and a model trained on the output of a cross-lingual annotation projection system.

In the next section we will describe our setup, then in section 3 present the shared feature representation we use, discuss the evaluation data and other technical aspects in section 4, present the results and conclude with an overview of related work.

2 Setup

The purpose of the study is not to develop a yet another semantic role labeling system – any existing SRL system can (after some modification) be used in this setup – but to assess the practical applicability of cross-lingual model transfer to this problem, compare it against the alternatives and identify its strong/weak points depending on a particular setup.

2.1 Semantic Role Labeling Model

We consider the dependency-based version of semantic role labeling as described in Hajič et al. (2009) and transfer an SRL model from one language to another. We only consider verbal predicates and ignore the predicate disambiguation stage. We also assume that the predicate identification information is available – in most languages it can be obtained using a relatively simple heuristic based on part-of-speech tags.

The model performs argument identification and classification (Johansson and Nugues, 2008) separately in a pipeline – first each candidate is classified as being or not being a head of an argument phrase with respect to the predicate in question and then each of the arguments is assigned a role from a given inventory. The model is factorized over arguments – the decisions regarding the classification of different arguments are made independently of each other.

With respect to the use of syntactic annotation we consider two options: using an existing dependency parser for the target language and obtaining one by means of cross-lingual transfer (see section 4.2).

Following McDonald et al. (2011), we assume that a part-of-speech tagger is available for the target language.

2.2 SRL in the Low-resource Setting

Several approaches have been proposed to obtain an SRL model for a new language with little or no manual annotation. Unsupervised SRL models (Lang and Lapata, 2010) cluster the arguments of predicates in a given corpus according to their semantic roles. The performance of such models can be impressive, especially for those languages where semantic roles correlate strongly with syntactic relation of the argument to its predicate. However, assigning meaningful role labels to the resulting clusters requires additional effort and the model's parameters generally need some adjustment for every language.

If the necessary resources are already available for a closely related language, they can be utilized to facilitate the construction of a model for the target language. This can be achieved either by means of cross-lingual annotation projection (Yarowsky et al., 2001) or by cross-lingual model transfer (Zeman and Resnik, 2008).

This last approach is the one we are considering in this work, and the other two options are treated as baselines. The unsupervised model will be further referred to as UNSUP and the projection baseline as PROJ.

2.3 Evaluation Measures

We use the F_1 measure as a metric for the argument identification stage and accuracy as an aggregate measure of argument classification performance. When comparing to the unsupervised SRL system the clustering evaluation measures are used instead. These are purity and collocation

$$PU = \frac{1}{N} \sum_i \max_j |G_j \cap C_i|$$

$$CO = \frac{1}{N} \sum_j \max_i |G_j \cap C_i|,$$

where C_i is the set of arguments in the i-th induced cluster, G_j is the set of arguments in the jth gold cluster and N is the total number of arguments. We report the harmonic mean of the two (Lang and Lapata, 2011) and denote it F_1^c to avoid confusing it with the supervised metric.

3 Model Transfer

The idea of this work is to abstract the model away from the particular source language and apply it to a new one. This setup requires that we use the same feature representation for both languages, for example part-of-speech tags and dependency relation labels should be from the same inventory.

Some features are not applicable to certain languages because the corresponding phenomena are absent in them. For example, consider a strongly inflected language and an analytic one. While the latter can usually convey the information encoded in the word form in the former one (number, gender, etc.), finding a shared feature representation for such information is non-trivial. In this study we will confine ourselves to those features that are applicable to all languages in question, namely: part-of-speech tags, syntactic dependency structures and representations of the word's identity.

3.1 Lexical Information

We train a model on one language and apply it to a different one. In order for this to work, the words of the two languages have to be mapped into a common feature space. It is also desirable that closely related words from both languages have similar representations in this space.

Word mapping. The first option is simply to use the source language words as the shared representation. Here every source language word would have itself as its representation and every target word would map into a source word that corresponds to it. In other words, we supply the model with a gloss of the target sentence.

The mapping (bilingual dictionary) we use is derived from a word-aligned parallel corpus, by identifying, for each word in the target language, the word in the source language it is most often aligned to.

Cross-lingual clusters. There is no guarantee that each of the words in the evaluation data is present in our dictionary, nor that the corresponding source-language word is present in the training data, so the model would benefit from the ability to generalize over closely related words. This can, for example, be achieved by using cross-lingual word clusters induced in Täckström et al. (2012). We incorporate these clusters as features into our model.

3.2 Syntactic Information

Part-of-speech Tags. We map part-of-speech tags into the universal tagset following Petrov et al. (2012). This may have a negative effect on the performance of a monolingual model, since most part-of-speech tagsets are more fine-grained than the universal POS tags considered here. For example Penn Treebank inventory contains 36 tags and the universal POS tagset – only 12. Since the finer-grained POS tags often reflect more language-specific phenomena, however, they would only be useful for very closely related languages in the cross-lingual setting.

The universal part-of-speech tags used in evaluation are derived from gold-standard annotation for all languages except French, where predicted ones had to be used instead.

Dependency Structure. Another important aspect of syntactic information is the dependency structure. Most dependency relation inventories are language-specific, and finding a shared representation for them is a challenging problem. One could map dependency relations into a simplified form that would be shared between languages, as it is done for part-of-speech tags in Petrov et al. (2012). The extent to which this would be useful, however, depends on the similarity of syntactic-semantic interfaces of the languages in question.

In this work we discard the dependency relation labels where the inventories do not match and only consider the unlabeled syntactic dependency graph. Some discrepancies, such as variations in attachment order, may be present even there, but this does not appear to be the case with the datasets we use for evaluation. If a target language is poor in resources, one can obtain a dependency parser for the target language by means of cross-lingual model transfer (Zeman and Resnik, 2008). We

take this into account and evaluate both using the original dependency structures and the ones obtained by means of cross-lingual model transfer.

3.3 The Model

The model we use is based on that of Björkelund et al. (2009). It is comprised of a set of linear classifiers trained using Liblinear (Fan et al., 2008). The feature model was modified to accommodate the cross-lingual cluster features and the reranker component was not used.

We do not model the interaction between different argument roles in the same predicate. While this has been found useful, in the cross-lingual setup one has to be careful with the assumptions made. For example, modeling the sequence of roles using a Markov chain (Thompson et al., 2003) may not work well in the present setting, especially between distant languages, as the order or arguments is not necessarily preserved. Most constraints that prove useful for SRL (Chang et al., 2007) also require customization when applied to a new language, and some rely on language-specific resources, such as a valency lexicon. Taking into account the interaction between different arguments of a predicate is likely to improve the performance of the transferred model, but this is outside the scope of this work.

3.4 Feature Selection

Compatibility of feature representations is necessary but not sufficient for successful model transfer. We have to make sure that the features we use are predictive of similar outcomes in the two languages as well.

Depending on the pair of languages in question, different aspects of the feature representation will retain or lose their predictive power. We can be reasonably certain that the identity of an argument word is predictive of its semantic role in any language, but it might or might not be true of, for example, the word directly preceding the argument word. It is therefore important to prevent the model from capturing overly specific aspects of the source language, which we do by confining the model to first-order features. We also avoid feature selection, which, performed on the source language, is unlikely to help the model to better generalize to the target one. The experiments confirm that feature selection and the use of second-order features degrade the performance of the transferred model.

`POS`	part-of-speech tags
`Synt`	unlabeled dependency graph
`Cls`	cross-lingual word clusters
`Gloss`	glossed word forms
`Deprel`	dependency relations

Table 1: Feature groups.

3.5 Feature Groups

For each word, we use its part-of-speech tag, cross-lingual cluster id, word identity (glossed, when evaluating on the target language) and its dependency relation to its parent. Features associated with an argument word include the attributes of the predicate word, the argument word, its parent, siblings and children, and the words directly preceding and following it. Also included are the sequences of part-of-speech tags and dependency relations on the path between the predicate and the argument.

Since we are also interested in the impact of different aspects of the feature representation, we divide the features into groups as summarized in table 1 and evaluate their respective contributions to the performance of the model. If a feature group is enabled – the model has access to the corresponding source of information. For example, if only `POS` group is enabled, the model relies on the part-of-speech tags of the argument, the predicate and the words to the right and left of the argument word. If `Synt` is enabled too, it also uses the POS tags of the argument's parent, children and siblings.

Word order information constitutes an implicit group that is always available. It includes the `Position` feature, which indicates whether the argument is located to the left or to the right of the predicate, and allows the model to look up the attributes of the words directly preceding and following the argument word. The model we compare against the baselines uses all applicable feature groups (`Deprel` is only used in EN-CZ and CZ-EN experiments with original syntax).

4 Evaluation

4.1 Datasets and Preprocessing

Evaluation of the cross-lingual model transfer requires a rather specific kind of dataset. Namely, the data in both languages has to be annotated

with the same set of semantic roles following the same (or compatible) guidelines, which is seldom the case. We have identified three language pairs for which such resources are available: English-Chinese, English-Czech and English-French.

The evaluation datasets for English and Chinese are those from the CoNLL Shared Task 2009 (Hajič et al., 2009) (henceforth CoNLL-ST). Their annotation in the CoNLL-ST is not identical, but the guidelines for "core" semantic roles are similar (Kingsbury et al., 2004), so we evaluate only on core roles here. The data for the second language pair is drawn from the Prague Czech-English Dependency Treebank 2.0 (Hajič et al., 2012), which we converted to a format similar to that of CoNLL-ST[1]. The original annotation uses the tectogrammatical representation (Hajič, 2002) and an inventory of semantic roles (or *functors*), most of which are interpretable across various predicates. Also note that the syntactic annotation of English and Czech in PCEDT 2.0 is quite similar (to the extent permitted by the difference in the structure of the two languages) and we can use the dependency relations in our experiments.

For English-French, the English CoNLL-ST dataset was used as a source and the model was evaluated on the manually annotated dataset from van der Plas et al. (2011). The latter contains one thousand sentences from the French part of the Europarl (Koehn, 2005) corpus, annotated with semantic roles following an adapted version of PropBank (Palmer et al., 2005) guidelines. The authors perform annotation projection from English to French, using a joint model of syntax and semantics and employing heuristics for filtering. We use a model trained on the output of this projection system as one of the baselines. The evaluation dataset is relatively small in this case, so we perform the transfer only one-way, from English to French.

The part-of-speech tags in all datasets were replaced with the universal POS tags of Petrov et al. (2012). For Czech, we have augmented the mappings to account for the tags that were not present in the datasets from which the original mappings were derived. Namely, tag "t" is mapped to "VERB" and "Y" – to "PRON".

We use parallel data to construct a bilingual dictionary used in word mapping, as well as in the projection baseline. For English-Czech and English-French, the data is drawn from Europarl (Koehn, 2005), for English-Chinese – from MultiUN (Eisele and Chen, 2010). The word alignments were obtained using GIZA++ (Och and Ney, 2003) and the intersection heuristic.

[1] see http://www.ml4nlp.de/code-and-data/treex2conll

4.2 Syntactic Transfer

In the low-resource setting, we cannot always rely on the availability of an accurate dependency parser for the target language. If one is not available, the natural solution would be to use cross-lingual model transfer to obtain it.

Unfortunately, the models presented in the previous work, such as Zeman and Resnik (2008), McDonald et al. (2011) and Täckström et al. (2012), were not made available, so we reproduced the direct transfer algorithm of McDonald et al. (2011), using Malt parser (Nivre, 2008) and the same set of features. We did not reimplement the projected transfer algorithm, however, and used the default training procedure instead of perceptron-based learning. The dependency structure thus obtained is, of course, only a rough approximation – even a much more sophisticated algorithm may not perform well when transferring syntax between such languages as Czech and English, given the inherent difference in their structure. The scores are shown in table 2.

We will henceforth refer to the syntactic annotations that were provided with the datasets as *original*, as opposed to the annotations obtained by means of syntactic transfer.

4.3 Baselines

Unsupervised Baseline: We are using a version of the unsupervised semantic role induction system of Titov and Klementiev (2012a) adapted to

Setup	UAS, %
EN-ZH	35
ZH-EN	42
EN-CZ	36
CZ-EN	39
EN-FR	67

Table 2: Syntactic transfer accuracy, unlabeled attachment score (percent). Note that in case of French we evaluate against the output of a supervised system, since manual annotation is not available for this dataset. This score does not reflect the true performance of syntactic transfer.

the shared feature representation considered in order to make the scores comparable with those of the transfer model and, more importantly, to enable evaluation on transferred syntax. Note that the original system, tailored to a more expressive language-specific syntactic representation and equipped with heuristics to identify active/passive voice and other phenomena, achieves higher scores than those we report here.

Projection Baseline: The projection baseline we use for English-Czech and English-Chinese is a straightforward one: we label the source side of a parallel corpus using the source-language model, then identify those verbs on the target side that are aligned to a predicate, mark them as predicates and propagate the argument roles in the same fashion. A model is then trained on the resulting training data and applied to the test set.

For English-French we instead use the output of a fully featured projection model of van der Plas et al. (2011), published in the CLASSiC project.

5 Results

In order to ensure that the results are consistent, the test sets, except for the French one, were partitioned into five equal parts (of 5 to 10 thousand sentences each, depending on the dataset) and the evaluation performed separately on each one. All evaluation figures for English, Czech or Chinese below are the average values over the five subsets. In case of French, the evaluation dataset is too small to split it further, so instead we ran the evaluation five times on a randomly selected 80% sample of the evaluation data and averaged over those. In both cases the results are consistent over the subsets, the standard deviation does not exceed 0.5% for the transfer system and projection baseline and 1% for the unsupervised system.

5.1 Argument Identification

We summarize the results in table 3. Argument identification is known to rely heavily on syntactic information, so it is unsurprising that it proves inaccurate when transferred syntax is used. Our simple projection baseline suffers from the same problem. Even with original syntactic information available, the performance of argument identification is moderate. Note that the model of (van der Plas et al., 2011), though relying on more expressive syntax, only outperforms the transferred system by 3% (F_1) on this task.

Setup	Syntax	TRANS	PROJ
EN-ZH	trans	34.5	13.9
ZH-EN	trans	32.6	15.6
EN-CZ	trans	46.3	12.4
CZ-EN	trans	42.3	22.2
EN-FR	trans	61.6	43.5
EN-ZH	orig	51.7	19.6
ZH-EN	orig	53.2	29.7
EN-CZ	orig	63.9	59.3
CZ-EN	orig	67.3	60.9
EN-FR	orig	71.0	51.3

Table 3: Argument identification, transferred model vs. projection baseline, F_1.

Most unsupervised SRL approaches assume that the argument identification is performed by some external means, for example heuristically (Lang and Lapata, 2011). Such heuristics or unsupervised approaches to argument identification (Abend et al., 2009) can also be used in the present setup.

5.2 Argument Classification

In the following tables, TRANS column contains the results for the transferred system, UNSUP – for the unsupervised baseline and PROJ – for projection baseline. We highlight in bold the higher score where the difference exceeds twice the maximum of the standard deviation estimates of the two results.

Table 4 presents the unsupervised evaluation results. Note that the unsupervised model performs as well as the transferred one or better where the

Setup	Syntax	TRANS	UNSUP
EN-ZH	trans	**83.3**	73.9
ZH-EN	trans	**79.2**	67.6
EN-CZ	trans	66.4	66.1
CZ-EN	trans	68.2	68.7
EN-FR	trans	**74.6**	65.1
EN-ZH	orig	84.5	**89.7**
ZH-EN	orig	79.2	**83.0**
EN-CZ	orig	74.1	74.0
CZ-EN	orig	74.6	**76.7**
EN-FR	orig	73.3	72.3

Table 4: Argument classification, transferred model vs. unsupervised baseline in terms of the clustering metric F_1^c (see section 2.3).

Setup	Syntax	TRANS	PROJ
EN-ZH	trans	**70.1**	69.2
ZH-EN	trans	**65.6**	61.3
EN-CZ	trans	**50.1**	46.3
CZ-EN	trans	53.3	**54.7**
EN-FR	trans	65.1	**66.1**
EN-ZH	orig	**71.7**	69.7
ZH-EN	orig	**66.1**	64.4
EN-CZ	orig	**59.0**	53.2
CZ-EN	orig	61.0	60.8
EN-FR	orig	63.0	**68.0**

Table 5: Argument classification, transferred model vs. projection baseline, accuracy.

original syntactic dependencies are available. In the more realistic scenario with transferred syntax, however, the transferred model proves more accurate.

In table 5 we compare the transferred system with the projection baseline. It is easy to see that the scores vary strongly depending on the language pair, due to both the difference in the annotation scheme used and the degree of relatedness between the languages. The drop in performance when transferring the model to another language is large in every case, though, see table 6.

Setup	Target	Source
EN-ZH	71.7	87.1
ZH-EN	66.1	86.2
EN-CZ	59.0	80.1
CZ-EN	61.0	75.4
EN-FR	63.0	82.5

Table 6: Model accuracy on the source and target language using original syntax. The source language scores for English vary between language pairs because of the difference in syntactic annotation and role subset used.

We also include the individual F_1 scores for the top-10 most frequent labels for EN-CZ transfer with original syntax in table 7. The model provides meaningful predictions here, despite low overall accuracy.

Most of the labels[2] are self-explanatory: Patient (PAT), Actor (ACT), Time (TWHEN), Effect (EFF), Location (LOC), Manner (MANN), Addressee (ADDR), Extent (EXT). CPHR marks the nominal part of a complex predicate, as in "to have [a plan]$_{CPHR}$", and DIR3 indicates destination.

[2]http://ufal.mff.cuni.cz/~toman/pcedt/en/functors.html

Label	Freq.	F_1	Re.	Pr.
PAT	14707	69.4	70.0	68.7
ACT	14303	81.1	81.7	80.4
TWHEN	3631	70.6	65.1	77.0
EFF	2601	45.4	67.2	34.3
LOC	1990	41.8	35.3	51.3
MANN	1208	54.0	63.8	46.9
ADDR	1045	30.2	34.4	26.8
CPHR	791	20.4	13.1	45.0
EXT	708	42.2	40.5	44.1
DIR3	695	20.1	17.3	23.9

Table 7: EN-CZ transfer (with original syntax), F_1, recall and precision for the top-10 most frequent roles.

5.3 Additional Experiments

We now evaluate the contribution of different aspects of the feature representation to the performance of the model. Table 8 contains the results for English-French.

Features	Orig	Trans
`POS`	47.5	47.5
`POS, Synt`	53.0	53.1
`POS, Cls`	53.7	53.7
`POS, Gloss`	63.7	63.7
`POS, Synt, Cls`	55.9	56.4
`POS, Synt, Gloss`	65.2	66.3
`POS, Cls, Gloss`	61.5	61.5
`POS, Synt, Cls, Gloss`	63.0	65.1

Table 8: EN-FR model transfer accuracy with different feature subsets, using original and transferred syntactic information.

The fact that the model performs slightly better with transferred syntax may be explained by two factors. Firstly, as we already mentioned, the original syntactic annotation is also produced automatically. Secondly, in the model transfer setup it is more important how closely the syntactic-semantic interface on the target side resembles that on the source side than how well it matches the "true" structure of the target language, and in this respect a transferred dependency parser may have an advantage over one trained on target-language data.

The high impact of the `Gloss` features here

may be partly attributed to the fact that the mapping is derived from the same corpus as the evaluation data – Europarl (Koehn, 2005) – and partly by the similarity between English and French in terms of word order, usage of articles and prepositions. The moderate contribution of the cross-lingual cluster features are likely due to the insufficient granularity of the clustering for this task.

For more distant language pairs, the contributions of individual feature groups are less interpretable, so we only highlight a few observations. First of all, both EN-CZ and CZ-EN benefit noticeably from the use of the original syntactic annotation, including dependency relations, but not from the transferred syntax, most likely due to the low syntactic transfer performance. Both perform better when lexical information is available, although the improvement is not as significant as in the case of French – only up to 5%.

The situation with Chinese is somewhat complicated in that adding lexical information here fails to yield an improvement in terms of the metric considered. This is likely due to the fact that we consider only the core roles, which can usually be predicted with high accuracy based on syntactic information alone.

6 Related Work

Development of robust statistical models for core NLP tasks is a challenging problem, and adaptation of existing models to new languages presents a viable alternative to exhaustive annotation for each language. Although the models thus obtained are generally imperfect, they can be further refined for a particular language and domain using techniques such as active learning (Settles, 2010; Chen et al., 2011).

Cross-lingual annotation projection (Yarowsky et al., 2001) approaches have been applied extensively to a variety of tasks, including POS tagging (Xi and Hwa, 2005; Das and Petrov, 2011), morphology segmentation (Snyder and Barzilay, 2008), verb classification (Merlo et al., 2002), mention detection (Zitouni and Florian, 2008), LFG parsing (Wróblewska and Frank, 2009), information extraction (Kim et al., 2010), SRL (Padó and Lapata, 2009; van der Plas et al., 2011; Annesi and Basili, 2010; Tonelli and Pianta, 2008), dependency parsing (Naseem et al., 2012; Ganchev et al., 2009; Smith and Eisner, 2009; Hwa et al., 2005) or temporal relation prediction (Spreyer and Frank, 2008). Interestingly, it has also been used to propagate morphosyntactic information between old and modern versions of the same language (Meyer, 2011).

Cross-lingual model transfer methods (McDonald et al., 2011; Zeman and Resnik, 2008; Durrett et al., 2012; Søgaard, 2011; Lopez et al., 2008) have also been receiving much attention recently. The basic idea behind model transfer is similar to that of cross-lingual annotation projection, as we can see from the way parallel data is used in, for example, McDonald et al. (2011).

A crucial component of direct transfer approaches is the unified feature representation. There are at least two such representations of lexical information (Klementiev et al., 2012; Täckström et al., 2012), but both work on word level. This makes it hard to account for phenomena that are expressed differently in the languages considered, for example the syntactic function of a certain word may be indicated by a preposition, inflection or word order, depending on the language. Accurate representation of such information would require an extra level of abstraction (Hajič, 2002).

A side-effect of using adaptation methods is that we are forced to use the same annotation scheme for the task in question (SRL, in our case), which in turn simplifies the development of cross-lingual tools for downstream tasks. Such representations are also likely to be useful in machine translation.

Unsupervised semantic role labeling methods (Lang and Lapata, 2010; Lang and Lapata, 2011; Titov and Klementiev, 2012a; Lorenzo and Cerisara, 2012) also constitute an alternative to cross-lingual model transfer.

For an overview of of semi-supervised approaches we refer the reader to Titov and Klementiev (2012b).

7 Conclusion

We have considered the cross-lingual model transfer approach as applied to the task of semantic role labeling and observed that for closely related languages it performs comparably to annotation projection approaches. It allows one to quickly construct an SRL model for a new language without manual annotation or language-specific heuristics, provided an accurate model is available for one of the related languages along with a certain amount of parallel data for the two languages. While an-

notation projection approaches require sentence- and word-aligned parallel data and crucially depend on the accuracy of the syntactic parsing and SRL on the source side of the parallel corpus, cross-lingual model transfer can be performed using only a bilingual dictionary.

Unsupervised SRL approaches have their advantages, in particular when no annotated data is available for any of the related languages and there is a syntactic parser available for the target one, but the annotation they produce is not always sufficient. In applications such as Information Retrieval it is preferable to have precise labels, rather than just clusters of arguments, for example.

Also note that when applying cross-lingual model transfer in practice, one can improve upon the performance of the simplistic model we use for evaluation, for example by picking the features manually, taking into account the properties of the target language. Domain adaptation techniques can also be employed to adjust the model to the target language.

Acknowledgments

The authors would like to thank Alexandre Klementiev and Ryan McDonald for useful suggestions and Täckström et al. (2012) for sharing the cross-lingual word representations. This research is supported by the MMCI Cluster of Excellence.

References

Omri Abend, Roi Reichart, and Ari Rappoport. 2009. Unsupervised argument identification for semantic role labeling. 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*, ACL '09, pages 28–36, Stroudsburg, PA, USA. Association for Computational Linguistics.

Paolo Annesi and Roberto Basili. 2010. Cross-lingual alignment of FrameNet annotations through hidden Markov models. In *Proceedings of the 11th international conference on Computational Linguistics and Intelligent Text Processing*, CICLing'10, pages 12–25, Berlin, Heidelberg. Springer-Verlag.

Roberto Basili, Diego De Cao, Danilo Croce, Bonaventura Coppola, and Alessandro Moschitti. 2009. Cross-language frame semantics transfer in bilingual corpora. In Alexander F. Gelbukh, editor, *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 332–345.

Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 43–48, Boulder, Colorado, June. Association for Computational Linguistics.

Ming-Wei Chang, Lev Ratinov, and Dan Roth. 2007. Guiding semi-supervision with constraint-driven learning. In *ACL*.

Chenhua Chen, Alexis Palmer, and Caroline Sporleder. 2011. Enhancing active learning for semantic role labeling via compressed dependency trees. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 183–191, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

Dipanjan Das and Slav Petrov. 2011. Unsupervised part-of-speech tagging with bilingual graph-based projections. *Proceedings of the Association for Computational Linguistics*.

Greg Durrett, Adam Pauls, and Dan Klein. 2012. Syntactic transfer using a bilingual lexicon. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1–11, Jeju Island, Korea, July. Association for Computational Linguistics.

Andreas Eisele and Yu Chen. 2010. MultiUN: A multilingual corpus from United Nation documents. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA).

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

Kuzman Ganchev, Jennifer Gillenwater, and Ben Taskar. 2009. Dependency grammar induction via bitext projection constraints. In *Proceedings of the 47th Annual Meeting of the ACL*, pages 369–377, Stroudsburg, PA, USA. Association for Computational Linguistics.

Qin Gao and Stephan Vogel. 2011. Corpus expansion for statistical machine translation with semantic role label substitution rules. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 294–298, Portland, Oregon, USA.

Trond Grenager and Christopher D. Manning. 2006. Unsupervised discovery of a statistical verb lexicon. In *Proceedings of EMNLP*.

Jan Hajič. 2002. Tectogrammatical representation: Towards a minimal transfer in machine translation. In Robert Frank, editor, *Proceedings of the 6th International Workshop on Tree Adjoining Grammars*

and Related Frameworks (TAG+6), pages 216—226, Venezia. Universita di Venezia.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi 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): Shared Task*, pages 1–18, Boulder, Colorado.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English dependency treebank 2.0. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel text. *Natural Language Engineering*, 11(3):311–325.

Richard Johansson and Pierre Nugues. 2008. Dependency-based semantic role labeling of PropBank. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 69–78, Honolulu, Hawaii.

Michael Kaisser and Bonnie Webber. 2007. Question answering based on semantic roles. In *ACL Workshop on Deep Linguistic Processing*.

Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, and Gary Geunbae Lee. 2010. A cross-lingual annotation projection approach for relation detection. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 564–571, Stroudsburg, PA, USA. Association for Computational Linguistics.

Paul Kingsbury, Nianwen Xue, and Martha Palmer. 2004. Propbanking in parallel. In *In Proceedings of the Workshop on the Amazing Utility of Parallel and Comparable Corpora, in conjunction with LREC'04*.

Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Bombay, India.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT.

Joel Lang and Mirella 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.

Joel Lang and Mirella Lapata. 2011. Unsupervised semantic role induction via split-merge clustering. In *Proc. of Annual Meeting of the Association for Computational Linguistics (ACL)*.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijing, China.

Adam Lopez, Daniel Zeman, Michael Nossal, Philip Resnik, and Rebecca Hwa. 2008. Cross-language parser adaptation between related languages. In *IJCNLP-08 Workshop on NLP for Less Privileged Languages*, pages 35–42, Hyderabad, India, January.

Alejandra Lorenzo and Christophe Cerisara. 2012. Unsupervised frame based semantic role induction: application to French and English. In *Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, pages 30–35, Jeju, Republic of Korea, July. Association for Computational Linguistics.

Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 62–72, Stroudsburg, PA, USA. Association for Computational Linguistics.

Paola Merlo, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. 2002. A multi-lingual paradigm for automatic verb classification. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pages 207–214, Philadelphia, PA.

Roland Meyer. 2011. New wine in old wineskins?–Tagging old Russian via annotation projection from modern translations. *Russian Linguistics*, 35(2):267(15).

Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 629–637, Jeju Island, Korea, July. Association for Computational Linguistics.

Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. *Comput. Linguist.*, 34(4):513–553, December.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1).

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection for semantic roles. *Journal of Artificial Intelligence Research*, 36:307–340.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31:71–105.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A universal part-of-speech tagset. In *Proceedings of LREC*, May.

Mark Sammons, Vinod Vydiswaran, Tim Vieira, Nikhil Johri, Ming wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundu, Yuancheng Tu, Kevin Small, Joshua Rule, Quang Do, and Dan Roth. 2009. Relation alignment for textual entailment recognition. In *Text Analysis Conference (TAC)*.

Burr Settles. 2010. Active learning literature survey. *Computer Sciences Technical Report*, 1648.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *EMNLP*.

David A Smith and Jason Eisner. 2009. Parser adaptation and projection with quasi-synchronous grammar features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 822–831. Association for Computational Linguistics.

Benjamin Snyder and Regina Barzilay. 2008. Cross-lingual propagation for morphological analysis. In *Proceedings of the 23rd national conference on Artificial intelligence*.

Anders Søgaard. 2011. Data point selection for cross-language adaptation of dependency parsers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 2 of *HLT '11*, pages 682–686, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kathrin Spreyer and Anette Frank. 2008. Projection-based acquisition of a temporal labeller. *Proceedings of IJCNLP 2008*.

Oscar Täckström, Ryan McDonald, and Jakob Uszkoreit. 2012. Cross-lingual word clusters for direct transfer of linguistic structure. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*, pages 477–487, Montréal, Canada.

Cynthia A. Thompson, Roger Levy, and Christopher D. Manning. 2003. A generative model for semantic role labeling. In *Proceedings of the 14th European Conference on Machine Learning*, ECML 2003, pages 397–408, Dubrovnik, Croatia.

Ivan Titov and Alexandre Klementiev. 2012a. A Bayesian approach to unsupervised semantic role induction. In *Proc. of European Chapter of the Association for Computational Linguistics (EACL)*.

Ivan Titov and Alexandre Klementiev. 2012b. Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Bombay, India, December.

Sara Tonelli and Emanuele Pianta. 2008. Frame information transfer from English to Italian. In *Proceedings of LREC 2008*.

Lonneke van der Plas, James Henderson, and Paola Merlo. 2009. Domain adaptation with artificial data for semantic parsing of speech. In *Proc. 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 125–128, Boulder, Colorado.

Lonneke van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, HLT '11, pages 299–304, Stroudsburg, PA, USA. Association for Computational Linguistics.

Alina Wróblewska and Anette Frank. 2009. Cross-lingual projection of LFG F-structures: Building an F-structure bank for Polish. In *Eighth International Workshop on Treebanks and Linguistic Theories*, page 209.

Dekai Wu and Pascale Fung. 2009. Can semantic role labeling improve SMT? In *Proceedings of 13th Annual Conference of the European Association for Machine Translation (EAMT 2009)*, Barcelona.

Chenhai Xi and Rebecca Hwa. 2005. A backoff model for bootstrapping resources for non-English languages. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 851–858, Stroudsburg, PA, USA.

David Yarowsky, Grace Ngai, and Ricahrd Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of Human Language Technology Conference*.

Daniel Zeman and Philip Resnik. 2008. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*, pages 35–42, Hyderabad, India, January. Asian Federation of Natural Language Processing.

Imed Zitouni and Radu Florian. 2008. Mention detection crossing the language barrier. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.