

Cross-lingual Dependency Transfer : What Matters?

Assessing the Impact of Pre- and Post-processing

Ophélie Lacroix, Guillaume Wisniewski and François Yvon
LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, F-91405 Orsay
{ophelie.lacroix, guillaume.wisniewski, francois.yvon}@limsi.fr

Abstract

In this paper, we propose to analyze the pre- and post-processing steps applied in the context of cross-lingual dependency transfer. To this aim, we employ a simple transfer strategy that operates on partially annotated projected data. We show that a good data selection strategy is a key point in successfully transferring dependencies and that better data selection techniques need to be developed in order to achieve the performance of fully supervised methods.

1 Introduction

Supervised learning techniques nowadays lie at the core of most Natural Language Processing (NLP) tools. Their use is however hindered by the scarcity of annotated data, which are only available for a restricted number of tasks, genres, domain, and languages. The supervision information that exists for well-resourced languages can however be transferred to under-resourced languages through the use of cross-lingual techniques. In this work, we focus on the transfer of syntactic dependency annotations.

Two main transfer strategies have been proposed in the literature: direct transfer model and annotation transfer. The first approach is mainly based on delexicalized parsing (Zeman and Resnik, 2008; McDonald et al., 2011) which assumes of common morpho-syntactic representation (e.g. PoS tags) between the source and target languages. It has been improved with the use of self-training, data selection, relexicalization and multi-source transfer (Naseem et al., 2010; Cohen et al., 2011; Sjøgaard, 2011; Täckström et al., 2013).

The second approach (transfer of annotations), relies on parallel corpora to project, through alignment links, the dependencies automatically predicted from a resource-rich language to a resource-poor language. This approach pioneered by Hwa et al. (2005) requires various heuristic transformation rules to cope with the non-isomorphism between the source and target structures as well as with the noise in source annotations and in alignments. It has since enjoyed a great popularity and been improved by many works (see the overview in Section 2).

In spite of the simplicity of the annotation transfer principle, all these methods have several (hidden) parameters, such as the symmetrization heuristic or filtering thresholds, that make any direct comparison of their performance very hard. That is why, in this work, we aim at analyzing the impact of external factors used as pre- and post-processing steps and their significance in the whole transfer process. To this end, we propose to use the simple transfer strategy exploiting partially annotated data introduced in Lacroix et al. (2016) to systematically compare various design decisions.

The transfer strategy used in our experiments is explained in Section 2. We then propose to explore and analyze different external factors: projected data filtering (Section 3.1), enhancement of the parsing strategy (Section 3.2) and multi-source transfer (Section 3.3). Finally, we compare the efficiency of dependency transfer and supervised parsing (Section 3.4) and analyze the performance achieved for the different kind of labels (Section 3.5).

2 Transfer strategy

In this Section, we describe the two main steps of the transfer process considered in our experiments: the projection of the dependencies through word alignments from a source language to a target language and the training method of Lacroix et al. (2016) for transition-based parser that can learn a parser from partial dependency trees. We also present the dataset used in our experiments and evaluate the proposed approach.

2.1 Dependency Projection

Most works on dependency transfer use a similar setting. They consider sentence-aligned bitexts for which an automatically parsed text in a resource-rich language is associated with its translation in a target language. Parallel sentences are then aligned in both directions and these alignments are merged with a symmetrization heuristics. To further guarantee their quality, alignments are generally filtered using various hand-crafted rules, for instance, to remove alignment links that associate words with different PoS tags (Rasooli and Collins, 2015) or sentences in which the number of alignment links is too low. Finally dependencies are projected through alignment links.

Dependencies for which both the head and the dependent are each aligned to exactly one word (i.e. 1 : 1 alignment) can be readily transferred to the target language. Difficulties with the projection arise with many-to-many links and un-aligned tokens. Some heuristics were proposed by Hwa et al. (2005), and reused by Tiedemann (2014), to deal with these multiple alignments. Nonetheless, to avoid this problem several works have proposed ad hoc rules to complete the trees. For instance, Spreyer and Kuhn (2009) propose to attach unaligned tokens to a fake root in order to ignore the actions associated with these dependencies during learning; Li et al. (2014) choose to add all dependency possibilities of the unaligned tokens to preserve ambiguity during learning; Ma and Xia (2014) consider, the potentially noisy, dependencies predicted by a delexicalized parser. Applying these heuristics allows, at the expense of adding potentially fake tokens or noisy dependencies, to label automatically a corpus of full parsing trees in the target language on which a stan-

dard learning method can be used.

In this work, we consider another approach and choose to ignore unattached words as well as many-to-many alignments: we focus on the projection through 1 : 1 alignments that are, intuitively, the most reliable. More precisely, to prevent annotations of noisy dependencies to be transferred, we decide to remove all multiple alignment links, and among the remaining links, we remove the ones that associate words with different PoS (following the soft rules proposed by Rasooli and Collins (2015)). After projection, we filter the sentences in which less than 80% of the words receive a dependency as they often result from bad quality alignments. At the end, we obtain an automatically annotated corpus for the target language that contains partial but accurate annotations.¹ We will describe, in the following Section, how a parser can be trained on such data.

In spite of its simplicity, this way to transfer dependency has several (hidden) parameters, such as the symmetrization heuristic or the filtering threshold, that can have a large impact on the quality of the transferred parser. We will evaluate, in Section 3 through 3.1 the impact of these design decisions.

2.2 Partial Transition-based Learning

We consider a transition-based dependency parser based on the arc-eager algorithm (Nivre, 2003): this parser builds a dependency tree incrementally by performing a sequence of *actions*. At each step of the parsing process, a classifier scores each possible action and the highest scoring one is applied.

Training relies on the dynamic oracle of Goldberg and Nivre (2012): for each sentence, a parse tree is built incrementally; at each step, if the predicted action creates an erroneous dependency (or, equivalently, prevents the creation of a gold dependency), a weight vector is updated, according to the perceptron rule. The set of all ‘correct’ actions is built considering the (potentially wrong) predicted tree and the gold action is defined as the correct action with the highest model score.

It is crucial to note that the training algorithm is an

¹We also remove from target sentences containing non-projective dependencies, as sentences containing non-projective dependencies often results in low-quality projected dependency structures (Mareček, 2011) and cannot be taken into account in the standard ARCEAGER training method.

error-correction learning procedure that solely relies on its capacity to detect when an action choice will result in an error: when no error is detected, the construction of the parse tree continues according to the model prediction. Consequently, this training procedure can also be used, as such, to train a dependency parser from partially annotated data: when no supervision information is available (no correct dependencies are known), all actions are considered as correct; in this case, the predicted action is necessarily equal to the correct action, the weight vector is not updated, and the training process goes on.

2.3 Dataset

All our experiments are carried out on six languages² of the Universal Dependency Treebank Project v2.0 (UDT) (McDonald et al., 2013): German (de), English (en), Spanish (es), French (fr), Italian (it) and Swedish (sv). We consider as parallel corpora a subset of the Europarl corpus that have exactly the same English sentences, collecting 1, 231, 216 parallel sentences for the 6 language pairs.

For the evaluation, the original splits (train/test/dev) of the UDT corpora are kept for training the source and evaluate the target.

2.4 Experiments

The parallel sentences are aligned in both directions with Giza++ (Och and Ney, 2003). These alignments are then merged with the intersection and grow-diag heuristics. For each language pair, the source dataset (Europarl) is PoS-tagged and parsed using the transition-based version of the MateParser (Bohnet and Nivre, 2012) with a beam of 40, which was trained on the UDT corpus. These predicted annotations are then partially projected on the target language data using the projection strategy described in Section 2.1.

To train a parser on partially projected target data, we used our own implementation of the arc-eager dependency parser, using the features described in Zhang and Nivre (2011). The greedy version of the parser is used in all but one experiments of the Section 3 while a beam-search (with a beam-size of 8 for learning & parsing) is used to achieve

²These are the languages present in both Europarl and UDT.

the best performances of the proposed method (Section 2.5).³

2.5 Performance

We present the results of our transfer strategy in the table 1. The results are first presented for cross-lingual transfer from English and second, applying a voting method for transferring from multiple source languages.⁴ These scores are obtained using the most appropriate external factors: filtering of the projected sentences for which less than 80% words are attached, beam-search strategy for parsing the source and target data. The effects of this parameters on the transfer results are clarified respectively in sections 3.1 and 3.2.

This method achieves results that are competitive with recent state-of-the-art methods such as (Ma and Xia, 2014; Rasooli and Collins, 2015), at a much cheaper computational cost,⁵ which allows us to make all the experiments required to compare the various design decisions. The results of Table 1 also show that using the grow-diag heuristic to symmetrize the alignments rather than the intersection heuristic hurts performance for all languages.

		intersection		grow-diag		sup.
source		(en)	(multi)	(en)	(multi)	
target	de	73.7	76.8	71.2	75.6	84.4
	es	76.8	79.3	75.4	79.0	85.5
	fr	77.9	80.9	76.7	81.0	85.8
	it	77.8	80.1	76.2	80.1	86.9
	sv	82.1	83.3	80.1	83.1	87.8

Table 1: Results of our transfer method. ‘sup’ present the fully supervised scores.

3 Analysis

3.1 The importance of filtering

To assess the usefulness of filtering on transfer performance, we conduct experiments on several lan-

³The beam search parser used in our experiments is described in (Aufrant and Wisniewski, 2016)

⁴For a target language, projection is applied from each of the other 5 languages. For each token of a sentence the most frequent head among the projected heads is saved (from the most frequent one to the less) if it does not impede the projectivity of the resulted tree (which then can be partially annotated).

⁵See (Lacroix et al., 2016) for more details.

guage pairs, considering two symmetrization heuristics: the intersection and the grow-diag heuristics. After the projection step, several greedy parsers are learned for increasing sizes of projected datasets. The sentences are included in the learning set in order of decreasing percentage of attached tokens.

The results for French, German and Swedish are presented respectively in Figure 1. Similar curves are obtained for Italian and Spanish. These results show that adding partially annotated sentences improves parsing performance as long as these sentences have enough attached tokens. For instance, in French (focusing on the scores obtained with the intersection heuristics), a parser trained only on fully labelled sentences achieves a UAS of 75.6%; when sentences in which more than 80% of dependencies are known are added, parsing performance is improved to 76.9%, but adding more sentences hurts performance. Indeed, sentences with a small number of attached tokens correspond to sentence pairs with few alignment links that are often not perfect translation of each other and may have very different grammatical structures.

One can notice that, while the number of sentences needed to reach the top scores varies greatly from a language to another, the average percentage of attached token per sentence remains in a short interval (from 74.9 % (de) to 84.9 % (sv)). Adding more sparse data seems to bring more noise than relevant syntactic information. Controlling the quality of the projected data, over the quantity, is therefore a key point in the success of the transfer process. This observation justifies our decision to consider only sentences with more than 80% of attached tokens.

Finally, the scores obtained with the use of the heuristic of symetrization intersection are mostly higher than those obtained with the grow-diag heuristic. It is worth noting that the number of sentences fully annotated with the grow-diag heuristic is far less important than with the intersection. For instance in French, the training filtered data contains 21,381 sentences when the intersection heuristic is used, but only 6,534 for the grow-diag heuristic. Indeed, the number of projected dependencies is lower because multiple alignments impede the projection of dependencies. This restrains the projection of potentially wrong dependencies from ambiguous alignments but also the diversity of the syn-

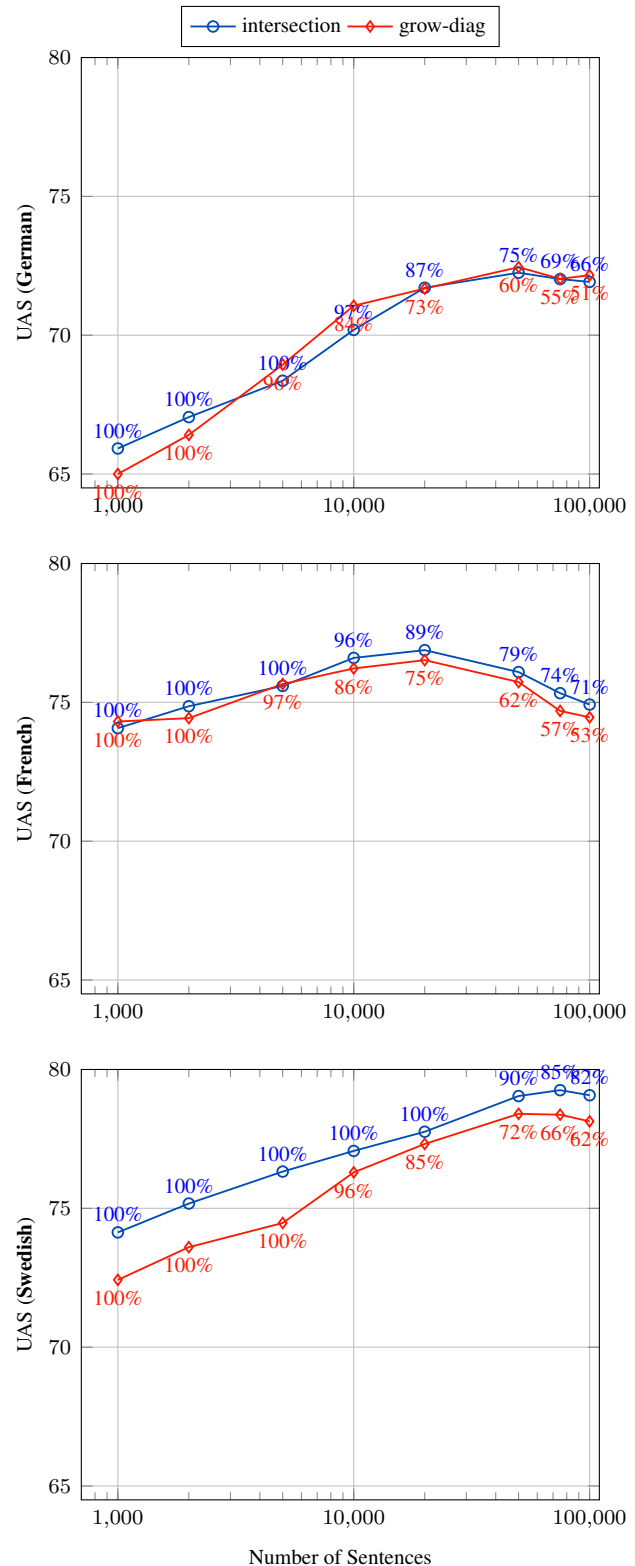


Figure 1: Variation of the transfer scores (UAS) depending on the number of sentences selected for learning. For each node the average percentage of attached token per sentence is specified. Greedy target parsing. Evaluation on gold PoS-tagged data.

tactic information transferred on the target language, and then the parsing performances. In the rest of the paper we will only consider the intersection heuristic.

This experiment shows that an appropriate selection of the alignment strategy and thus of the projected data used for learning could benefit the transfer scores of the methods that exploit (raw or even completed) partial data.

3.2 Pumping the parsing

It is well known that different techniques can boost parsing performance. For instance, clusters (Koo et al., 2008) may be used to reduce lexical sparseness, which is particularly appropriate in the case of dependency parser transfer since parallel data are generally not from the same domain as the corpus used to train and evaluate the parser. Another approach for boosting parsing performance is the use of a beam-search strategy that reduces the number of search errors (Zhang and Nivre, 2012). In this section, we aim at assessing, first, how parsing performance of the source language impacts the quality of the transferred parser, and second, how using more ‘advanced’ parsing techniques may boost parsing in the target language.

Using a similar transfer process as in the previous section, we conduct experiments in which the source and target parsers will be progressively enriched: we consider, in a first experiment, a greedy parser to predict the dependencies of the source and target sentences; the source greedy parsers are then replaced by a beam-search parser and features describing Brown clusters learned from the Europarl data are added. Finally, we also consider a beam-search parser for parsing the target language, using our own implementation of the transition based parser presented in Section 2.2 with a beam size of 8, and enrich it with Brown clusters.

The transfer scores are presented in Table 2. First, these results show that the alignments are not good enough to reflect improvements in (source) parsing quality on the target data: the use of beam-search on parsing the source language allows an average improvement of 0.2 UAS point on the target languages, while the source (English) performance is improved by 2.3. The use of clusters does not improve the

Parsing strategy				Transfer target				
src		tgt		de	es	fr	it	sv
beam	clusters	beam	clusters					
-	-	-	-	71.3	75.6	76.3	76.2	79.9
✓	-	-	-	71.3	75.8	76.8	76.0	80.4
✓	✓	-	-	71.4	75.8	76.0	76.2	80.3
✓	✓	✓	-	73.5	76.7	77.7	78.3	82.2
✓	✓	✓	✓	73.8	76.7	77.7	78.0	81.9

Table 2: Transfer results according to the parsing strategy used on the source and target data. Evaluation, in UAS, on Gold PoS-tagged data.

average score, nor the source parsing performance.⁶ However, the beam-search strategy is surely useful for parsing the projected target languages: scores are, on average, 1.34 higher. The use of clusters is not interesting: only German performance is improved (+0.3) while Italian and Swedish are both negatively impacted (−0.3 for both).

We have observed that the use of clusters is mostly useless in any case and, globally, that boosting the source parsing performance have very little effect on transfer final scores. However, not surprisingly, the use of beam-search for parsing the target data is highly effective for boosting the transfer scores.

3.3 Multisource impact

We have seen in Section 3.1 that filtering the projected data is a key point to achieve good transfer scores, as adding too much data for learning reduces the parsing performance. However, for a similar percentage of attached token, the number of sentences kept for learning varies a lot depending on the target language. For instance, when transferring from English, the number of sentences having more than 80% dependencies reaches 52,554 for Swedish but only 15,191 for German; the parsing scores differ greatly as well: their UAS is, respectively, 81.9 and 73.8. Spanish, Italian and French achieve relatively

⁶The inefficiency of clusters on English may be due to the fact that clusters are learned from Europarl data which are out-of-domain for the UDT data.

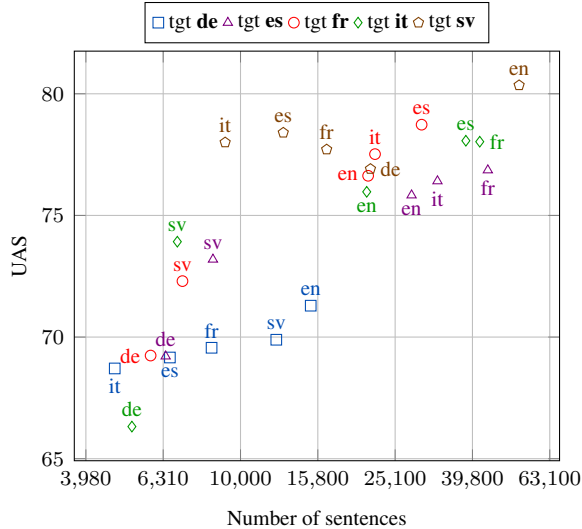


Figure 2: Results of the multi-lingual transfer depending on the number of sentences saved for learning. The label of each node indicates the source language and its color the target language.

close scores for the same order of number of sentences (around 20/30K).

These observations show that there is a correlation between the number of dependencies transferred and the parsing performance: the more filtered sentences there are the better the scores are. A natural way to increase the number of sentences with a high number of dependencies is to transfer dependencies from different languages: good projections result from good word alignments, that depend on source and target languages at stake.

We conduct multi-lingual experiments, similar to the experiments from English, in which each language (among German, Spanish, French, Italian and Swedish) is considered as the source language for the other ones. We consider as a parallel corpora a subset of the Europarl corpus (as detailed in Section 2.3). The sentences are filtered as previously after the dependencies have been transferred across alignment links. All reported results are achieved by a greedy parser considering gold PoS-tagged data.

The results of this multi-lingual experiment are presented in Figure 2, with, for every source language, the number of sentences that survived filtering.⁷ These results show that, for each target language, the best score is achieved for the source lan-

⁷It is important to note that the number of sentences is proportional to the number of dependencies.

guage with the largest training set (which is: English for German and Swedish, French for Spanish and Italian, and Spanish for French). With the exception of Swedish, for a given target language, the UAS is proportional to the number of sentences, regardless of the source parsing performance.⁸ As expected, languages from the same family, such as the ones derived from Latin (Spanish, French and Italian), are beneficial for each other. This observation has already been reported several times (e.g. by McDonald et al. (2013)), but Figure 2 suggests that the increase in performance may mainly result from a good alignment between the source and the target languages. Overall, these results stress the fact that the alignment quality has a large impact on the transfer performance and should not be neglected.

As seen in the previous section, the source parsing performance does not appear to be the parameter having the greatest impact on transfer. The quality of the alignment deriving from the choice of the source language is quite crucial.

3.4 Transfer vs Supervised Parsing

Cross-lingual transfer strategies are mainly used with the aim of developing swiftly NLP tools for resource-poor languages without the need of annotated corpora that are expensive to build from scratch. Results presented in the previous sections show that parsers trained on transferred data are still outperformed by supervised parsers. It is however difficult to evaluate how prejudicial this loss is. That is why, to assess the usefulness of transfer methods, we propose to determine the amount of gold annotated sentences needed to achieve performance similar to the performance of transferred parsers.

In a first series of experiments, we compare the performance of a cross-lingual dependency parser to supervised parsers learned from increasing amount of data (starting with 50 sentences). All the experiments are performed on 10 runs to mitigate the impact of selecting labelled data randomly, using the greedy version of the parser. Scores are evaluated on gold PoS-tagged data. The results, presented in Figure 3,⁹ show that the amount of supervised

⁸The source performances obtained with the MateParser (predicted PoS-tags) are: 92.4 (en), 80.4 (de), 83.1 (es), 83.8 (fr), 84.2 (it) and 85.7 (sv).

⁹Results for Spanish and Italian are quite similar to French.

data needed to achieve transfer scores are respectively around 250, 200 and 400 sentences for German, French and Swedish. This observation strongly questions the interest of cross-lingual transfer: only a very limited amount of annotated data is required to outperform a parser trained on transferred annotations.

Wisniewski et al. (2014) show that the performance difference between a supervised and transferred PoS taggers partially results from divergences in annotation conventions and from evaluating the tagger on out-of-domain data: as in the setting described in Section 2, the taggers are trained on Europarl and evaluated on UDT. To assess the impact on parsing performance of these two elements, we propose, in a second series of experiments, to enrich the data labelled automatically by transferring annotations with an increasing amount of in-domain labelled data to learn new parsers.

Results of these experiments are presented in Figure 3. They show that a small amount of supervised data ($\simeq 300$ sentences) provides useful information to projected data for parsing target languages: the performance of a parser trained on the combination of transferred and labelled data outperforms both a parser trained on the labelled data only and a parser trained only on the transferred data. However, above a specific threshold, projected data become useless, and worse, adding them to labelled data hurts parsing performance. This could mean that the projected data do not just lack syntactic diversity but also contain a substantial amount of projection errors, even if the alignments have been filtered with very conservative rules.

3.5 Label scores and frequencies

The previous experiments suggest that cross-lingual parsers suffer from alignment errors (or from their absence) even if the alignments are filtered (restricted to 1:1 and PoS-coherent links). To reveal systematic syntactic errors, we propose to examine the transfer scores depending on the (gold) syntactic label of the dependency. Our hypothesis is that the UAS of a given target label depends both on the capacity of the source parser to predict this label and on the ability of the transfer method to project the syntactic information. For the sake of clarity, we only report results for English to French transfer.

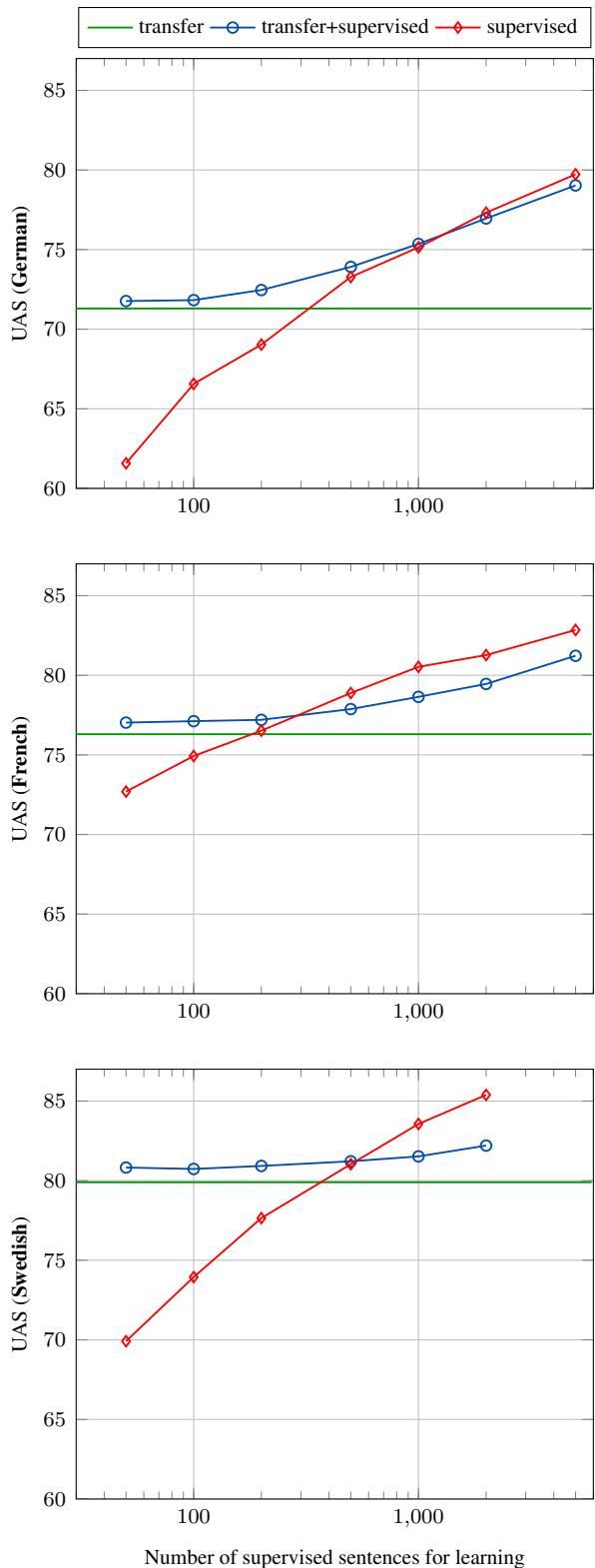


Figure 3: Evaluation of a parser for German learned from transferred and supervised data according to the number of included supervised data, and compared to fully supervised parsing. Greedy parsing. Evaluation on Gold PoS-tagged data.

Table 3 shows¹⁰ the frequencies of the labels predicted by a supervised parser on the English and French EUROPARL corpora as well as the frequencies of these labels on the French projected (from English) and filtered data. It appears that the frequencies of the labels tend to look like the source frequencies, introducing a systematic bias in the data used to train a cross-lingual parser. In particular, some dependencies, such as the root, are over-projected but no label is entirely skipped. The syntactic information are quite proportionally transferred. Table 3 also reported the supervised UAS of these labels for English and French, as well as the UAS of the transferred parser for French. We observe that the prediction of each label suffers from transfer: each score is generally lower than the (source or target) supervised score. One can also notice that over-projection do not benefit label scores. Overall, it may suggest that projection errors are also quite proportionally transferred among the various labels.

Results of Table 3 also suggest that the capacity of the source parser to predict a given kind of dependency has not much impact on the performance of the target parser. For instance, the ADPOBJ and DET are both very well predicted by the English parser, but the prediction of the cross-lingual French parser are far better for DET than for ADPOBJ

Similar scores and frequencies are observed for different pair of source-target languages. In addition, it is worth noting that we observe comparable behaviour when scores are computed depending on the PoS of the tokens. Frequencies are relatively well preserved and the loss in UAS is shared among the PoS tags.

4 Conclusion

We have proposed to apply a simple method that learns transferred dependency parsers from partially projected data with the aim of analyzing the various parameters that impact the parsing performance. Our observations are valid for many methods of dependency transfer that operates on annotations (partially) projected via alignments links.

We have shown that the selection of the align-

¹⁰We are only considering the most frequent labels. The frequency of all ignored labels is less than 1%.

label	frequencies (%)			UAS (%)		
	parsed		tr.	sup.		tr.
	fr	en	fr	en	fr	fr
adpmod	14.5	10.9	9.9	83.5	78.8	70.0
adpobj	13.2	10.7	9.8	95.4	92.2	85.1
det	13.3	10.4	12.9	96.9	96.6	95.3
nsubj	6.2	7.1	8.8	90.4	88.3	81.8
amod	6.0	5.9	6.7	94.1	92.1	88.2
root	3.3	3.6	8.4	89.8	78.0	71.3
dobj	4.9	4.2	4.8	92.4	91.6	86.2
advmod	3.4	4.3	4.2	83.5	74.2	71.5
conj	3.3	3.3	2.8	69.7	57.3	49.5
cc	2.7	3.2	2.8	75.2	62.9	54.3

Table 3: English-to-French transfer scores per dependency label and frequencies of these labels in the Europarl corpus, parsed and transferred (tr.). Supervised (sup.) scores for the source and target languages are also reported.

ments (thus the dependencies) and the filtering of the projected data are crucial. The quantity of projected data used for learning is not relevant if quality is not controlled. However, the quantity of training data is correlated to the parsing performance, as quantity is rather a consequence of the quality of the alignments. Finally, the quality of alignment greatly depends on the relation shared between the source and target languages. It appear that all these choices are far more important than the quality of the source parsing.

Moreover, we have seen that performance of transfer techniques still lag behind those of fully supervised learning. Our experiments suggest that many attachment errors are produced during the dependency projection and that these errors are spread over all kind of syntactic phenomena. They surely derived from alignment errors and variation in the annotation scheme between languages. The recent development of more coherent annotation schemes and corpora (universal dependencies (Nivre et al., 2015)) tends to alleviate these problems but there is still work to be done concerning the quality of the alignments. The main difficulty is to preserve enough sentences for learning while preventing the projection of erroneous dependencies.

Acknowledgments

This work has been partly funded by a DGA-RAPID project under grant agreement N.º1429060465 (Papyrus). We thank the reviewers for their accurate comments and suggestions.

References

- Lauriane Aufrant and Guillaume Wisniewski. 2016. PanParser: a Modular Implementation for Efficient Transition-Based Dependency Parsing. Technical report, LIMSI-CNRS, March.
- Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Jeju Island, Korea, July. Association for Computational Linguistics.
- Shay B. Cohen, Dipanjan Das, and Noah A. Smith. 2011. Unsupervised Structure Prediction with Non-Parallel Multilingual Guidance. In *Proceedings of EMNLP 2011, the Conference on Empirical Methods in Natural Language Processing*, pages 50–61, Edinburgh, Scotland, UK., July.
- Yoav Goldberg and Joakim Nivre. 2012. A Dynamic Oracle for Arc-Eager Dependency Parsing. In *Proceedings of COLING 2012, the International Conference on Computational Linguistics*, pages 959–976, Bombay, India.
- Rebecca Hwa, Philip Resnik, A. Weinberg, C. Cabezas, and O. Kolak. 2005. Bootstrapping Parsers via Syntactic Projection across Parallel Texts. *Natural language engineering*, 11:311–325.
- Terry Koo, Xavier Carreras Pérez, and Michael Collins. 2008. Simple Semi-supervised Dependency Parsing. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 595–603.
- Ophélie Lacroix, Lauriane Aufrant, Guillaume Wisniewski, and François Yvon. 2016. Frustratingly easy cross-lingual transfer for transition-based dependency parsing. In *The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2016, San Diego, California, USA.
- Zhenghua Li, Min Zhang, and Wenliang Chen. 2014. Soft Cross-lingual Syntax Projection for Dependency Parsing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 783–793, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Xuezhe Ma and Fei Xia. 2014. Unsupervised dependency parsing with transferring distribution via parallel guidance and entropy regularization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1337–1348, Baltimore, Maryland, June.
- David Mareček. 2011. Combining Diverse Word-Alignment Symmetrizations Improves Dependency Tree Projection. In *Computational Linguistics and Intelligent Text Processing*, pages 144–154. Springer.
- Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source Transfer of Delexicalized Dependency Parsers. In *Proceedings of EMNLP 2011, the Conference on Empirical Methods in Natural Language Processing*, pages 62–72.
- Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of ACL 2013, the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria, August.
- Tahira Naseem, Harr Chen, Regina Barzilay, and Mark Johnson. 2010. Using Universal Linguistic Knowledge to Guide Grammar Induction. In *Proceedings of EMNLP 2010, the Conference on Empirical Methods in Natural Language Processing*, pages 1234–1244, Stroudsburg, PA, USA.
- Joakim Nivre, Željko Agić, Maria Jesus Aranzabe, Masayuki Asahara, Aitziber Atutxa, Miguel Ballesteros, John Bauer, Kepa Bengoetxea, Riyaz Ahmad Bhat, Cristina Bosco, Sam Bowman, Giuseppe G. A. Celano, Miriam Connor, Marie-Catherine de Marneffe, Arantza Diaz de Ilarraza, Kaja Dobrovoljc, Timothy Dozat, Tomaž Erjavec, Richárd Farkas, Jennifer Foster, Daniel Galbraith, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Berta Gonzales, Bruno Guillaume, Jan Hajič, Dag Haug, Radu Ion, Elena Irimia, Anders Johannsen, Hiroshi Kanayama, Jenna Kanerva, Simon Krek, Veronika Laippala, Alessandro Lenci, Nikola Ljubešić, Teresa Lynn, Christopher Manning, Ctlna Mrnduc, David Mareček, Héctor Martínez Alonso, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Anna Missilä, Verginica Mititelu, Yusuke Miyao, Simonetta Montemagni, Shunsuke Mori, Hanna Nurmi, Petya Osenova, Lilja Øvrelid, Elena Pascual, Marco Passarotti, Cemel-Augusto Perez, Slav Petrov, Jussi Piitulainen, Barbara Plank, Martin Popel, Prokopis Prokopidis, Sampo Pyysalo, Loganathan Ramasamy,

- Rudolf Rosa, Shadi Saleh, Sebastian Schuster, Wolfgang Seeker, Mojgan Seraji, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Kiril Simov, Aaron Smith, Jan Štěpánek, Alane Suhr, Zsolt Szántó, Takaaki Tanaka, Reut Tsarfaty, Sumire Uematsu, Larraitz Uribe, Viktor Varga, Veronika Vincze, Zdeněk Žabokrtský, Daniel Zeman, and Hanzhi Zhu. 2015. Universal dependencies 1.2. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague.
- Joakim Nivre. 2003. An Efficient Algorithm for Projective Dependency Parsing. In *Proceedings of IWPT 2003, the 8th International Workshop on Parsing Technologies*, Nancy, France.
- Franz Joseph Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29:19–51.
- Mohammad Sadegh Rasooli and Michael Collins. 2015. Density-driven cross-lingual transfer of dependency parsers. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 328–338, Lisbon, Portugal, September. Association for Computational Linguistics.
- Anders Søgaard. 2011. Data point selection for cross-language adaptation of dependency parsers. In *Proceedings of ACL 2011, the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 682–686, Portland, Oregon, USA, June.
- Kathrin Spreyer and Jonas Kuhn. 2009. Data-Driven Dependency Parsing of New Languages Using Incomplete and Noisy Training Data. In *Proceedings of CoNLL 2009, the Thirteenth Conference on Computational Natural Language Learning*, pages 12–20, Boulder, Colorado, June.
- Oscar Täckström, Ryan McDonald, and Joakim Nivre. 2013. Target Language Adaptation of Discriminative Transfer Parsers. In *Proceedings of ACL 2013, the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1061–1071, Atlanta, Georgia.
- Jörg Tiedemann. 2014. Rediscovering Annotation Projection for Cross-Lingual Parser Induction. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1854–1864, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.
- Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham, and François Yvon. 2014. Cross-lingual part-of-speech tagging through ambiguous learning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1779–1785, Doha, Qatar, October.
- 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.
- Yue Zhang and Joakim Nivre. 2011. Transition-based Dependency Parsing with Rich Non-local Features. In *Proceedings of ACL 2011, the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 188–193, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Yue Zhang and Joakim Nivre. 2012. Analyzing the Effect of Global Learning and Beam-Search on Transition-Based Dependency Parsing. In *Proceedings of COLING 2012, the 24th International Conference on Computational Linguistics: Technical Papers*, pages 1391–1400, Mumbai, India, December.