

How to Parse Low-Resource Languages: Cross-Lingual Parsing, Target Language Annotation, or Both?

Ailsa Meechan-Maddon

Uppsala University

Department of Linguistics and Philology

aime5651@student.uu.se

Joakim Nivre

Uppsala University

Department of Linguistics and Philology

joakim.nivre@lingfil.uu.se

Abstract

To develop a parser for a language with no syntactically annotated data, we either have to develop a (small) treebank for the target language or rely on cross-lingual learning or projection, or possibly use some combination of these methods. In this paper, we compare the usefulness of cross-lingual model transfer and target language annotation for three different languages, with varying support from closely related high-resource languages. The results show that annotating even a very small amount of data in the target language is superior to any cross-lingual setup and that accuracy can be further improved by adding training data from related languages in a multilingual model.

1 Introduction

Despite significant advances in natural language processing over several decades, even basic technologies like part-of-speech tagging and syntactic parsing are still available only for a tiny fraction of the languages of the world. This observation has led to an increasing interest in techniques for supporting low-resource languages, typically by making use of data from high-resource languages together with methods for cross-lingual learning or transfer. These techniques include annotation projection (Hwa et al., 2002), model transfer (Zeman and Resnik, 2008; McDonald et al., 2011), treebank translation (Tiedemann et al., 2014), and multilingual parsing models (Duong et al., 2015a; Ammar et al., 2016). Despite the undeniable progress in this line of research, the question always looms large whether it is not more effective to simply annotate a small amount of training data in the target language of interest. Daniel Zeman, one of the inventors of delexicalized transfer parsing, maintains that you can get over 50% accuracy for many languages with just 100 annotated sentences, citing as evidence the results of Ramasamy (2014) for some Indian languages. Further support comes from the study of Garcia et al. (2018), who compares cross-lingual parsing to target language annotation in the context of building a treebank for Galician.

In this paper, we approach this question by comparing three ways of training dependency parsers for low-resource languages: *monolingual* models trained on small amounts of *target* language data; *cross-lingual* models trained only on data from related *support* languages; and *multilingual* models trained on both support and target language data. We perform experiments on three target languages with varying support from related high-resource languages: Faroese (supported by Danish, Norwegian, and Swedish), Upper Sorbian (supported by Czech, Polish, and Slovak), and North Saami (supported by Estonian, Finnish, and Hungarian). Our results show that monolingual models consistently outperform cross-lingual models even with very limited amounts of training data. In addition, there is always a multilingual model that outperforms the best monolingual model. Taken together, these results suggest that the most effective strategy for low-resource parser development may well be to annotate as much data as you can afford in the target language and then add training data from related languages if available.

2 Methodology

To be able to compare monolingual, cross-lingual and multilingual models, we adopt the multilingual parsing approach pioneered by Ammar et al. (2016) and deployed on a large scale by Smith et al. (2018a)

Language	Treebank	Train	Dev	Test
Faroese	OFT	4.9k	2.5k	2.5k
Danish	DDT	80k		
Norwegian	Nynorsk	245k		
Swedish	Talbanken	67k		
Upper Sorbian	UFAL	5.8k	2.7k	2.7k
Czech	PDT	300k		
Polish	LFG	105k		
	SZ	63k		
Slovak	SNK	81k		
North Saami	Giella	14.3k	2.5k	10k
Estonian	EDT	288k		
Finnish	FTB	128k		
	TDT	163k		
Hungarian	Szeged	20k		

Table 1: Data sets (UD v2.3) with number of tokens.

in the 2018 CoNLL shared task on universal dependency parsing (Zeman et al., 2018). This approach differs from early work on model transfer, which relied on delexicalized models with part-of-speech tags as pivot features (Zeman and Resnik, 2008; McDonald et al., 2011). Although these models initially gave encouraging results, especially for closely related languages, the results were mostly based on experiments with gold part-of-speech tags, severely overestimating the accuracy achievable under more realistic conditions (Tiedemann, 2015). We instead use lexicalized models, which do not presuppose part-of-speech tagging or any other preprocessing except tokenization for the target language, and instead rely on word, character and language embeddings. Besides being more realistic in a low-resource setting, this is justified by the reduced importance of part-of-speech tagging for neural dependency parsers (Dozat et al., 2017; Ma et al., 2018; Smith et al., 2018b).

2.1 Languages and Treebanks

From Universal Dependencies v2.3 (Nivre et al., 2016; Nivre et al., 2018), we select three language clusters with one low-resource language and three related support languages with larger treebanks: a Scandinavian cluster with Faroese supported by Danish, Norwegian (Nynorsk) and Swedish; a West Slavic cluster with Upper Sorbian supported by Czech, Polish and Slovak; and a Uralic cluster with North Saami supported by Estonian, Finnish and Hungarian. It is worth noting that the support languages are much more closely related to the target language in the Scandinavian and West Slavic clusters than in the Uralic cluster. Table 1 lists the treebanks used for each language and the number of tokens in each data set.

2.2 Parser

We use UUParser v2.3 (de Lhoneux et al., 2017a; Smith et al., 2018a), which is an adaptation of the transition-based parser of Kiperwasser and Goldberg (2016) specifically for multilingual models. The original parsing architecture relies on a BiLSTM to learn representations of tokens in context and a multi-layer perceptron to predict transitions and arc labels based on a few BiLSTM vectors. The multilingually motivated extensions in UUParser include an extended transition system for handling non-projective structures (de Lhoneux et al., 2017b) and a richer representation of input tokens. More specifically, each input token w_i in language l is represented by:

$$\mathbf{x} = e(w) \circ \text{BiLSTM}(ch_{1:m}) \circ e(t)$$

Here \mathbf{x} is the concatenation of a word embedding $e(w)$, a character-based vector $\text{BiLSTM}(ch_{1:m})$ obtained by running a BiLSTM over the characters $ch_{1:m}$ of w , and a treebank embedding $e(t)$ representing the

Scandinavian				West Slavic				Uralic						
-Target		+Target		-Target		+Target		-Target		+Target				
UAS	LAS	UAS	LAS	UAS	LAS	UAS	LAS	UAS	LAS	UAS	LAS			
Faroese		78.6	71.1	Upper Sorbian		66.0	58.4	N Saami		66.0	58.6			
Dan	45.9	30.6	81.5	74.2	Cze	33.1	23.5	71.5	64.2	Est	22.4	8.5	68.8	60.1
Nor	47.4	35.7	84.0	76.2	Pol	44.9	34.3	71.3	64.2	Fin	22.5	7.5	67.3	59.5
Swe	45.9	24.9	81.1	73.9	Slo	41.2	29.5	68.6	61.9	Hun	19.4	4.9	65.6	57.5
Dan+Nor	48.5	34.5	83.7	76.7	Cze+Pol	51.1	41.4	72.2	63.9	Est+Fin	24.7	9.4	64.5	55.3
Dan+Swe	55.9	35.9	82.7	75.5	Cze+Slo	44.1	32.6	72.5	65.3	Est+Hun	23.8	8.9	67.0	58.4
Nor+Swe	56.4	39.6	83.9	77.0	Pol+Slo	47.5	38.8	72.2	64.9	Fin+Hun	20.8	8.0	65.9	57.7
All	57.7	44.4	82.8	75.3	All	52.4	43.3	69.6	62.8	All	27.1	11.6	65.4	56.4

Table 2: Test set accuracy for target languages (UAS, LAS). -Target = cross-lingual models trained without target language data. +Target = models trained on target language data; monolingual (first row) and multilingual.

treebank t that the input comes from. The treebank embedding is used to distinguish data from different languages as well as different treebanks from the same language (Stymne et al., 2018; Smith et al., 2018a). We drop the treebank embedding when training models on a single treebank and otherwise train all models with default settings and no pre-trained embeddings. For cross-lingual and multilingual models, word, character and language embeddings are thus learned jointly for all languages.

2.3 Experimental Setup

Within each cluster, we train cross-lingual models on data from every combination of one, two or three support languages (7 models), multilingual models on the same data sets plus target language data (7 models), and a monolingual model only on target language data, for a total of 15 models. For the support languages, we only use the dedicated training sets from Universal Dependencies v2.3 (Nivre et al., 2018). We do not standardize training set sizes, since the parser has been shown to be robust to size differences when training multi-treebank models (Stymne et al., 2018), but we limit the size of the Czech training set (which is about four times bigger than any other) to 300k tokens. For the target languages, we need a training set for the mono- and multilingual models, a development set to tune hyper-parameters, and a test set for the final evaluation. For Faroese and Upper Sorbian, there is only about 10k tokens of data, which we subdivide into 50% training, 25% development, and 25% test. For North Saami, there is more data, so we leave the dedicated test set of 10k tokens intact and extract a development set of 2.5k tokens from the training set, leaving 14.3k words for training.

The development sets for target languages are used for model selection as follows:

- For support languages with two treebanks of roughly equal size, we run preliminary experiments with cross-lingual models to decide whether to use both or only one. The resulting selection can be seen in Table 1.
- To improve compatibility of character-based representations across (support and target) languages, we try mapping characters that exist only in a target language to characters that exist in one or more support language. This is helpful only for Faroese, where we map {ÍÚýúđí} to {IUyudi}.
- For cross-lingual models, the parser does not learn a language embedding for the target language, so we select a support language to use as proxy during parsing based on labelled attachment score (LAS) on the target language development set.
- All models are trained for 30 epochs, and the best epoch is selected according to LAS on the target language development set.

Finally, the development sets are also used in learning curve experiments (see Section 3).

3 Results and Discussion

Table 2 reports results on the test sets for all cross-lingual, multilingual and monolingual models on our three target languages. The first thing to note is that the monolingual models, trained only on about 5k tokens for Faroese and Upper Sorbian and 14k tokens for North Saami, consistently outperform all cross-lingual models by a wide margin. The difference is especially large for the Uralic cluster, where the target language is in a different branch of the language family from all support languages, and where the best cross-lingual model does not even reach 10% for LAS (25% for unlabelled attachment score (UAS)). But even for the Scandinavian and West Slavic clusters, where languages are more closely related and the best cross-lingual models get LAS over 40% and UAS over 50%, the monolingual model gives a higher LAS score by at least 15% absolute (12% absolute for UAS). This indicates that annotating a relatively small amount of training data in the target language is generally superior to using cross-lingual model transfer.

The second main trend is that, despite the poor results for cross-lingual models, the best multilingual models consistently outperform the monolingual models. For the Scandinavian and West Slavic clusters, *all* multilingual models outperform the monolingual model and the best model improves by as much as 5.9/6.9 LAS and 5.3/6.5 UAS. But even for the Uralic cluster, where data from the related support languages seem completely useless in the cross-lingual scenario, using the same data in a multilingual model improves on the monolingual model in 3 out of 7 cases for UAS (2 out of 7 for LAS). The relative improvement for the best multilingual model is smaller than in the other two clusters, but it should be kept in mind that the target language training set is almost three times bigger for North Saami than for Faroese and Upper Sorbian. These results suggest that, even if target language annotation is more effective than cross-lingual transfer, adding data from related support languages can nevertheless lead to further improvements.

To understand why multilingual models work so much better than cross-lingual models, it is important to note that the former learn word, character and language embeddings for the target language and that these embeddings are learned together for all languages. The cross-lingual models have no target language specific representations and have to rely on a proxy language embedding and the existence of cognates for matching word and character representations. This works especially poorly for the Uralic cluster, where the distance from the target to the support languages is much larger.

So how much target language data do we need to outperform a cross-lingual model? To answer this question, we run learning curve experiments for the monolingual and best multilingual models, using the development sets for evaluation, and gradually increasing the amount of target language training data from 0 to 50, 100, 500, 1k, 3k, 5k and 10k tokens (Figure 1). For the Scandinavian and Uralic clusters, we only need 1k tokens for the monolingual model to surpass the cross-lingual model with respect to LAS. For the West Slavic cluster, results are slightly erratic for the smallest training sets, but 3k tokens definitely suffice to reach the accuracy of the best cross-lingual model.¹ In all three cases, this is less than 200 sentences,² so the results seem to support Daniel Zeman’s claim that something like 100 sentences can be sufficient to train a decent parser, although in our study it is only Faroese that reaches a (labeled) accuracy of 50% with only 100 sentences.

4 Related Work

Work on cross-lingual learning for parsing and related tasks has focused on three main approaches: annotation projection (Hwa et al., 2002; Hwa et al., 2005; Tiedemann, 2014), model transfer (Zeman and Resnik, 2008; McDonald et al., 2011), and (to a lesser extent) treebank translation (Tiedemann et al., 2014). Annotation projection and treebank translation presupposes parallel data, so we will focus on model transfer, which is closest to our work. Model transfer was pioneered for closely related languages by Zeman and Resnik (2008), using delexicalized models and relying on a common part-of-speech tagset for the source and target language. The idea was refined and generalized to multi-source transfer by

¹If we consider UAS instead of LAS, the patterns are very similar, with the curves crossing at around 1k tokens for the Scandinavian and Uralic clusters and just under 3k tokens for the West Slavic cluster, so we omit these figures to save space.

²Upper Sorbian has significantly longer sentences than the other two target languages.

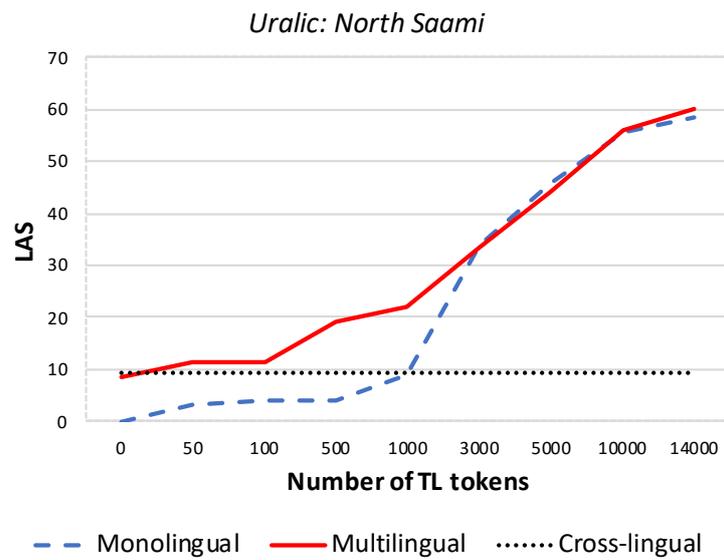
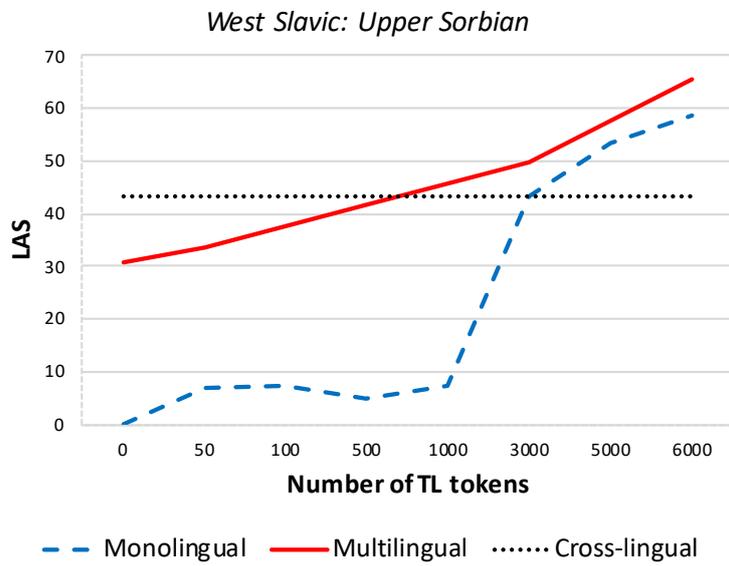
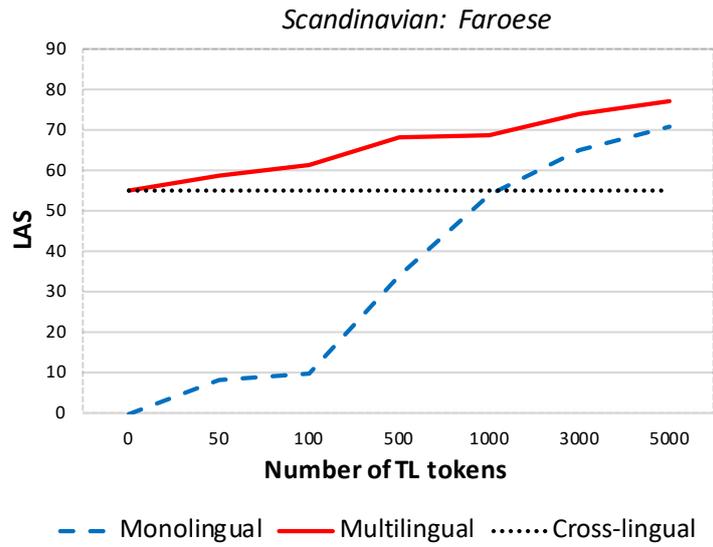


Figure 1: Learning curves (LAS) for the monolingual model (blue dashed line) and the best multilingual model (red solid line), compared to the best cross-lingual model (black dotted line).

McDonald et al. (2011) and gained further momentum with the advent of cross-linguistically consistent syntactic annotation, which facilitated evaluation (McDonald et al., 2013). Other studies concerned methods for selecting optimal source languages (Søgaard and Wulff, 2012; Rosa and Zabokrtsky, 2015). However, most of the early studies of model transfer relied on evaluation with gold part-of-speech tags on the target side, which was later shown to give over-optimistic results (Tiedemann, 2015).

A study of special relevance to our own work is that of Garcia et al. (2018), who specifically study the amount of target language training data needed to outperform cross-lingual model transfer in the context of building a UD treebank for Galician. Drawing on data from 7 other Romance language varieties (Brazilian Portuguese, Catalan, European Portuguese, French, Italian, Romanian and Spanish), they show that a single-source transfer parser achieves LAS corresponding to about 3,000 tokens of target language training data and UAS corresponding to about 7,000 tokens. However, they also show that careful combination and adaptation of source language data from multiple languages can increase these numbers to 16,000 (LAS) and 20,000 (UAS). One difference between their study and ours is that they make use of part-of-speech tags as pivot features, which may explain why especially the adapted multi-source transfer parsers seem to perform better than in our study. In addition, the similarity between Galician and some of the Romance languages is probably greater than in most of our support-target language pairs. Another difference is that Garcia et al. (2018) find that cross-lingual parsers are more competitive with respect to UAS than LAS, whereas we find that about the same number of target language training tokens is needed to reach cross-lingual performance with respect to both metrics. It is possible but by no means obvious that this difference is also related to the presence or absence of part-of-speech tags.

The increasing use of neural networks and distributed representations in syntactic parsing has led to more flexible models for cross-lingual and multilingual learning embeddings that go beyond delexicalized models and their reliance on part-of-speech tags (Duong et al., 2015a; Duong et al., 2015b; Guo et al., 2015a; Guo et al., 2015b). Especially important for our own work is the multilingual model of Ammar et al. (2016) with its use of language embeddings, which were later generalized to treebank embeddings that allow seamless integration of multiple languages as well as heterogeneous treebanks for a single language (de Lhoneux et al., 2017a; Stymne et al., 2018; Smith et al., 2018a). A more recent line of research involves the use of synthetic treebanks (Wang and Eisner, 2016; Wang and Eisner, 2018), an approach recently applied to parser development for one of our target languages, Faroese (Tyers et al., 2018).

Finally, it is worth noting that the superiority of annotating target language data over using cross-lingual methods has also been demonstrated for the related part-of-speech tagging problem, in the context of historical text processing, by Schultz and Kuhn (2016) and Schultz and Ketchik (2019).

5 Conclusion

We have compared cross-lingual, multilingual and monolingual parser training for three low-resource languages, supported to different degrees by related languages with more resources. Our main conclusion is that training a monolingual model on target language data gives better performance than any cross-lingual model as soon as we have at least 200 annotated target language sentences. Moreover, adding data from related languages to train a multilingual model can improve performance further by up to 7 LAS points. In conclusion, to develop a parser for a low-resource language, annotate as much data as you can afford and add data from related languages if available.

Acknowledgements

We would like to thank our friends and colleagues, Ali Basirat, Miryam de Lhoneux, Harald Hammarström, Christian Hardmeier, Artur Kulmizev, Paola Merlo and Gongbo Tang, for their invaluable comments on this paper. We also wish to extend our thanks to the anonymous reviewers who provided great insight and suggestions. Joakim Nivre’s contributions to this work were supported by grant 2016-01817 of the Swedish Research Council.

References

- Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah Smith. 2016. Many languages, one parser. *Transactions of the Association for Computational Linguistics*, 4:431–444.
- Miryam de Lhoneux, Yan Shao, Ali Basirat, Eliyahu Kiperwasser, Sara Stymne, Yoav Goldberg, and Joakim Nivre. 2017a. From raw text to Universal Dependencies – Look, no tags! In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 207–217.
- Miryam de Lhoneux, Sara Stymne, and Joakim Nivre. 2017b. Arc-hybrid non-projective dependency parsing with a static-dynamic oracle. In *Proceedings of the 15th International Conference on Parsing Technologies*, pages 99–104.
- Timothy Dozat, Peng Qi, and Christopher D. Manning. 2017. Stanford’s graph-based neural dependency parser at the conll 2017 shared task. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 20–30.
- Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. 2015a. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 845–850.
- Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. 2015b. A neural network model for low-resource universal dependency parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 339–348.
- Marcos Garcia, Carlos Gómez-Rodríguez, and Miguel A. Alonso. 2018. New treebank or repurposed? on the feasibility of cross-lingual parsing of romance languages with universal dependencies. *Natural Language Engineering*, 24:91–122.
- Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015a. Cross-lingual dependency parsing based on distributed representations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1234–1244.
- Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015b. A representation learning framework for multi-source transfer parsing. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2734–2740.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. 2002. Evaluating translational correspondence using annotation projection. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 392–399.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural Language Engineering*, 11(3):311–325.
- Eliyahu Kiperwasser and Yoav Goldberg. 2016. Simple and accurate dependency parsing using bidirectional lstm feature representations. *Transactions of the Association for Computational Linguistics*, 4:313–327.
- Xuezhe Ma, Zecong Hu, Jingzhou Liu, Nanyun Peng, Graham Neubig, and Eduard Hovy. 2018. Stack-pointer networks for dependency parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1403–1414.
- 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)*, 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 the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Dan Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Rogier Blokland, Victoria Bobicev, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Drostanova, Puneet Dwivedi, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomáš Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Radu Ion, Elena Irimia, Olájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Kamil Kopacewicz, Natalia Kotsyba, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horriacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayo Olùòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roşca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle Teller, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uribe, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Seyi Williams, Mats Wirén, Tsegay Woldemariam, Tak-sum Wong, Chunxiao Yan, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, Manying Zhang, and Hanzhi Zhu. 2018. Universal dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Logathan Ramasamy. 2014. *Parsing Under-Resourced languages: Cross-Lingual Transfer Strategies for Indian Languages*. Ph.D. thesis, Charles University in Prague.

Rudolf Rosa and Zdenek Zabokrtsky. 2015. KLcpo3 – a language similarity measure for delexicalized parser transfer. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 243–249.

Aaron Smith, Bernd Bohnet, Miryam de Lhoneux, Joakim Nivre, Yan Shao, and Sara Stymne. 2018a. 82 treebanks, 34 models: Universal dependency parsing with multi-treebank models. In *Proceedings of the 2018 CoNLL Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.

Aaron Smith, Miryam de Lhoneux, Sara Stymne, and Joakim Nivre. 2018b. An investigation of the interactions between pre-trained word embeddings, character models and pos tags in dependency parsing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

- Anders Søgaard and Julie Wulff. 2012. An empirical study of non-lexical extensions to delexicalized transfer. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 1181–1190.
- Sara Stymne, Miryam Lhoneux, Aaron Smith, and Joakim Nivre. 2018. Parser training with heterogeneous treebanks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 619–625.
- Jörg Tiedemann, Željko Agić, and Joakim Nivre. 2014. Treebank translation for cross-lingual parser induction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning (CoNLL)*, pages 130–140.
- Jörg Tiedemann. 2014. Rediscovering annotation projection for cross-lingual parser induction. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1854–1864.
- Jörg Tiedemann. 2015. Cross-lingual dependency parsing with Universal Dependencies and predicted PoS labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling)*, pages 340–349.
- Francis Tyers, Mariya Sheyanova, Aleksandra Martynova, Pavel Stepachev, and Konstantin Vinogorodskiy. 2018. Multi-source synthetic treebank creation for improved cross-lingual dependency parsing. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 144–150.
- Dingquan Wang and Jason Eisner. 2016. The galactic dependencies treebanks: Getting more data by synthesizing new languages. *Transactions of the Association for Computational Linguistics*, 4:491–505.
- Dingquan Wang and Jason Eisner. 2018. Synthetic data made to order: The case of parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1325–1337.
- 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.
- Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthtyersast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*.