

# Cross-lingual Parsing with Polyglot Training and Multi-treebank Learning: A Faroese Case Study

James Barry and Joachim Wagner and Jennifer Foster

ADAPT Centre

School of Computing, Dublin City University, Ireland

firstname.lastname@adaptcentre.ie

## Abstract

Cross-lingual dependency parsing involves transferring syntactic knowledge from one language to another. It is a crucial component for inducing dependency parsers in low-resource scenarios where no training data for a language exists. Using Faroese as the target language, we compare two approaches using annotation projection: first, projecting from multiple *monolingual* source models; second, projecting from a single *polyglot* model which is trained on the combination of all source languages. Furthermore, we reproduce multi-source projection (Tyers et al., 2018), in which dependency trees of multiple sources are combined. Finally, we apply multi-treebank modelling to the projected treebanks, in addition to or alternatively to polyglot modelling on the source side. We find that polyglot training on the source languages produces an overall trend of better results on the target language but the single best result for the target language is obtained by projecting from monolingual source parsing models and then training multi-treebank POS tagging and parsing models on the target side.

## 1 Introduction

Cross-lingual transfer methods, i. e. methods that transfer knowledge from one or more source languages to a target language, have led to substantial improvements for low-resource dependency parsing (Rosa and Mareček, 2018; Agić et al., 2016; Guo et al., 2015; Lynn et al., 2014; McDonald et al., 2011; Hwa et al., 2005) and part-of-speech (POS) tagging (Plank and Agić, 2018). In low-resource scenarios, there may be not enough data for data-driven models to learn how to parse. In cases where no annotated data is available, knowledge is often transferred from annotated data in other languages and when there is only a small amount of annotated data, additional knowl-

edge can be induced from external corpora such as by learning distributed word representations (Mikolov et al., 2013; Al-Rfou' et al., 2013) and more recent contextualized variants (Peters et al., 2018; Devlin et al., 2019).

This work focuses on dependency parsing for low-resource languages by means of annotation projection (Yarowsky et al., 2001) and synthetic treebank creation (Tiedemann and Agić, 2016). We build on recent work by Tyers et al. (2018) who show that in the absence of annotated training data for the target language, a lexicalized treebank can be created by translating a target language corpus into a number of related source languages and parsing the translations using models trained on the source language treebanks.<sup>1</sup> These annotations are then projected to the target language using separate word alignments for each source language, combined into a single graph for each sentence and decoded (Sagae and Lavie, 2006), resulting in a treebank for the target language, Faroese in the case of Tyers et al.'s and our experiments.

Inspired by recent literature involving multilingual learning (Mulcaire et al., 2019; Smith et al., 2018; Vilares et al., 2016), we investigate whether additional improvements can be made by:

1. using a single polyglot<sup>2</sup> parsing model which is trained on the combination of all source languages to create synthetic source treebanks (which are subsequently projected to the target language)

<sup>1</sup>In this paper, *source language* and *target language* always refer to the projection, not the direction of translation.

<sup>2</sup>We adopt the same terminology used in Mulcaire et al. (2019), who use the term *cross-lingual transfer* to describe methods involving the use of one or more source languages to process a target language. They reserve the term *polyglot learning* for training a single model on multiple languages and where parameters are shared between languages. For the polyglot learning technique applied to multiple treebanks of a single language, we use the term *multi-treebank learning*.

2. training a multi-treebank model on the individually projected treebanks and the treebank produced with multi-source projections.

The former differs from the approach of [Tyers et al. \(2018\)](#), who use multiple discrete, monolingual models to parse the translated sentences, whereas in this work we use a single model trained on multiple source treebanks. The latter differs from training on the target treebank produced by multi-source projection in that the information of the individual projections is still available and training data is not reduced to cases where all source languages provide a projection.

In other words, we aim to investigate whether the current state-of-the-art approach for Faroese, which relies on cross-lingual transfer, can be improved upon by adopting an approach based on source-side polyglot learning and/or target-side multi-treebank learning. We hypothesize that a polyglot model can exploit similarities in morphology and syntax across the included source languages, which will result in a better model to provide annotations for projection. On the target side, we expect that combining different sources of information will result in a more robust target model.

We evaluated our various models on the Faroese test set and experienced considerable gains for three of the four source languages (Danish, Norwegian Bokmål and Swedish) by adopting a polyglot model. However, for Norwegian Nynorsk, a stronger monolingual model was able to outperform the polyglot approach. When we extended multi-treebank learning to the target side, we experienced additional gains for all cases. Our best result of 71.5 – an absolute improvement of 7.2 points over the result reported by [Tyers et al. \(2018\)](#) – was achieved with multi-treebank target learning over the monolingual projections.

## 2 Background

[Tyers et al. \(2018\)](#) describe a method for creating synthetic treebanks for Faroese based on previous work which uses machine translation and word alignments to transfer trees from source language(s) to the target language. Sentences from Faroese are translated into the four source languages Danish, Swedish, Norwegian Nynorsk and Norwegian Bokmål. The translated sentences are then tokenized, POS tagged and parsed using the relevant source language model trained on the source language treebank. The resulting trees

are projected back to the Faroese sentences using word alignments. The four trees for each sentence are combined into a graph with edge scores one to four (the number of trees that support them), from which a single tree per sentence is produced using the Chu-Liu-Edmonds algorithm ([Chu and Liu, 1965](#); [Edmonds, 1967](#)). The resulting trees make up a synthetic treebank for Faroese which is then used to train a Faroese parsing model. The parser output is evaluated using the gold-standard Faroese test treebank developed by [Tyers et al. \(2018\)](#). The approach is compared to a delexicalized baseline, which it outperforms by a large margin. It is also shown that, for Faroese, a combination of the four source languages (multi-source projection) is superior to individual language projection.

The idea of annotation projection using word-alignments originates from ([Yarowsky et al., 2001](#)) who used word alignments to transfer information such as POS tags from source to target languages. This method was later used in dependency parsing by [Hwa et al. \(2005\)](#), who project dependencies to a target language and use a set of heuristics to form dependency trees in the target language. A parser is then trained on the projected treebank and evaluated against gold-standard treebanks. [Zeman and Resnik \(2008\)](#) introduced the idea of delexicalized dependency parsing whereby a parser is trained using only POS information and is then applied to a target language.

[McDonald et al. \(2011\)](#) perform delexicalized dependency parsing using direct transfer and show that this approach outperforms unsupervised approaches for grammar induction. Importantly, this approach can be extended to the multi-source case by training on multiple source languages and predicting a target language. In an additional experiment, they combine annotation projection and multi-source transfer.

[Tiedemann and Agić \(2016\)](#) present a thorough comparison of pre-neural cross-lingual parsing. Various forms of projected annotation methods are compared to delexicalized baselines, and the use of machine translation instead of parallel corpora to produce synthetic treebanks in the target language is explored. In contrast to [Tyers et al. \(2018\)](#), they translate a target sentence and project the source parse tree back to the target during test time instead of using this approach to obtain training data for the target language.

Agić et al. (2016) leverage massively multi-lingual parallel corpora such as translations of the Bible and web-scraped data from the Watchtower Online Library website<sup>3</sup> for low-resource POS tagging and dependency parsing using annotation projection. They project weight matrices (as opposed to decoded dependency arcs) from multiple source languages and average the matrices weighted by word alignment confidences. They then decode the weight matrices into dependency trees on the target side, which are then used to train a parser. This approach utilizes dense information from multiple source languages, which helps reduce noise from source side predictions but to the best of our knowledge, the source-side parsing models learn information between source languages independently and the cross-lingual interaction only occurs when projecting the edge scores into multi-source weight matrices.

The idea of projecting dense information in the form of a weighted graph has been further extended by Schlichtkrull and Søgaard (2017) who bypass the need to train the target parser on decoded trees and develop a parser which can be trained directly on weighted graphs.

Plank and Agić (2018) use annotation projection for POS tagging. They find that choosing high quality training instances results in superior accuracy than randomly sampling a larger training set. To this end, they rank the target sentences by the percentage of words covered by word alignments across all source languages and choose the top  $k$  covered sentences for training.

Meechan-Maddon and Nivre (2019) carry out an evaluation on cross-lingual parsing for three low-resource languages which are supported by related languages. They include three experiments: first, training a *monolingual* model on a small number of sentences in the target language; second, training a *cross-lingual* model on related source languages which is then applied to the target data and lastly, training a *multilingual* model which includes target data as well as data from the related support languages. They found that training a monolingual model on the target data was always superior to training a cross-lingual model. Interestingly, they found that the best results were achieved by training a model on the various support languages as well as the target data, i. e. their multilingual model. While we do not combine

<sup>3</sup><https://wol.jw.org/>

the synthetic target treebanks with the source treebanks in our experiments, the results of Meechan-Maddon and Nivre (2019) motivate us to carry out this experiment in the future.

### 3 Method

We outline the process used for creating a synthetic treebank for cross-lingual dependency parsing. We use the following resources: raw Faroese sentences taken from Wikipedia, a machine translation system to translate these sentences into all source languages (Danish, Swedish, Norwegian Nynorsk and Norwegian Bokmål), a word-aligner to provide word alignments between the words in the target and source sentences, treebanks for the four source languages on which to train parsing models, POS tagging and parsing tools, and, lastly a target language test set. We use the same raw corpus, alignments and tokenized and segmented versions of the source translations<sup>4</sup> as Tyers et al. (2018) who release all of their data.<sup>5</sup> In this way, the experimental pipeline is the same as theirs but we predict POS tags and dependency annotations using our own models.

**Target Language Corpus** We use the target corpus built by Tyers et al. (2018) which comprises 28,862 sentences which were extracted from Faroese Wikipedia dumps<sup>6</sup> using the WikiExtractor script<sup>7</sup> and further pre-processed to remove any non-Faroese texts and other forms of unsuitable sentences.

**Machine Translation** As noted by Tyers et al. (2018), popular repositories for developing machine translation systems such as OPUS (Tiedemann, 2016) contain an inadequate amount of sentences to train a data-driven machine translation system for Faroese. For instance, there are fewer than 7,000 sentence pairs between Faroese and Danish, Faroese and English, Faroese and Norwegian and Faroese and Swedish. Consequently, to create parallel source sentences, Tyers et al. (2018) use a rule-based machine translation system available in Apertium<sup>8</sup> to translate from Faroese to

<sup>4</sup>The original authors tokenize and segment the source translations with UDPipe.

<sup>5</sup><https://github.com/ftyers/cross-lingual-parsing>

<sup>6</sup><https://dumps.wikimedia.org/>

<sup>7</sup><https://github.com/attardi/wikiextractor>

<sup>8</sup><https://github.com/apertium>

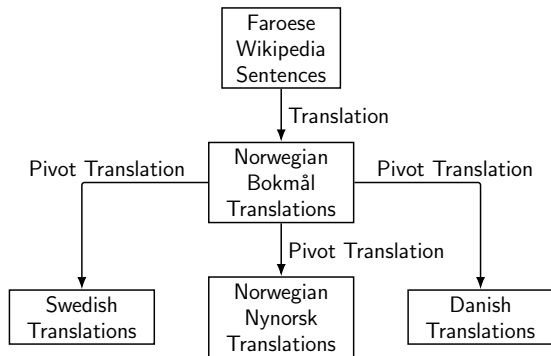


Figure 1: Overview of the machine translation process. The Faroese sentences are first translated into Norwegian Bokmål and then from Norwegian Bokmål into the other source languages (pivot translation).

Norwegian Bokmål. There also exists translation systems from Norwegian Bokmål to Norwegian Nynorsk, Swedish and Danish in Apertium. As a result, the authors use *pivot translation* from Norwegian Bokmål into the other source languages. The process is illustrated in Fig. 1. For a more thorough description of the machine translation process and for resource creation in general, see the work of Tyers et al. (2018).

**Word Alignments** We use word alignments between the Faroese text and the source translations generated by Tyers et al. (2018) using fast\_align (Dyer et al., 2013), a word alignment tool based on IBM Model 2.<sup>9</sup>

**Source Treebanks** We use the Universal Dependencies v2.2 treebanks (Nivre et al., 2018) to train our source parsing models. This is the version used for the 2018 CoNLL shared task on Parsing Universal Dependencies (Zeman et al., 2018).

**Source Tagging and Parsing Models** In order for our parsers to work well with predicted POS tags, we follow the same steps as used in the 2018 CoNLL shared task for creating training and development treebanks with automatically predicted POS tags (henceforth referred to as silver POS). Since we are required to parse translated text which only has lexical features available, we

<sup>9</sup>Note that previous related work (Agić et al., 2016) report better results using IBM Model 1 with a more diverse language setup. They claim that IBM Model 2 introduces a bias towards more closely related languages. As we are working with related languages and translations and alignments are largely word-for-word, we expect that this will have less of an impact on our experiments although IBM Model 1 should also be tried in future work.

disregard lemmas, language-specific POS (XPOS) and morphological features and only use the word form and universal POS (UPOS) tag as input features to our parsers. We develop our POS tagging and parsing models using the AllenNLP library (Gardner et al., 2018).

We use jackknife resampling to predict the UPOS tags for the training treebanks. We split the training treebank into ten parts, train models on nine parts and predict UPOS for the excluded part. The process is repeated until all ten parts are predicted and they are then combined to recreate the treebank with silver POS tags. Only token features are used to predict the UPOS tag.<sup>10</sup> Finally, we train a model per source language on the full training data to check performance on the respective development set and to POS tag the source language translations before parsing.

We train two variants of parsing models. The first is a monolingual biaffine dependency parser (Dozat and Manning, 2017) trained on the individual source treebanks. The second is a polyglot model trained on all source treebanks using the multilingual parser of Schuster et al. (2019), which is the same graph-based biaffine dependency parser, extended to enable parsing with multiple treebanks. We additionally include a treebank embedding (Ammar et al., 2016; Stymne et al., 2018) to the input of the polyglot parser to help the parser differentiate between the source languages. We optimize the model for average development set LAS across the included languages. The process is illustrated in Fig. 2.

To ensure that our parser is realistic, we add a pre-trained monolingual word embedding to each monolingual parser, giving a considerable improvement in accuracy on the development sets of the source languages. We use the precomputed Word2Vec embeddings<sup>11</sup> released as part of the 2017 CoNLL shared task on UD parsing (Zeman et al., 2017) which were trained on CommonCrawl and Wikipedia.

In order to use pre-trained word embeddings for the polyglot setting, we need to consider that a polyglot model uses a shared vocabulary across all input languages. In our experiments, we simply

<sup>10</sup>We observe slightly lower POS tagging scores on fully annotated test sets than UDpipe, which uses gold lemmas, XPOS and morphological features to predict the UPOS label and therefore cannot be applied to the translated text without also building predictors for these features.

<sup>11</sup><https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1989>

use the union of the word embeddings and average the word vector for words that occur in more than one language. Future work should explore cross-lingual word embeddings with limited amount of parallel data or use aligned contextual embeddings as in (Schuster et al., 2019).

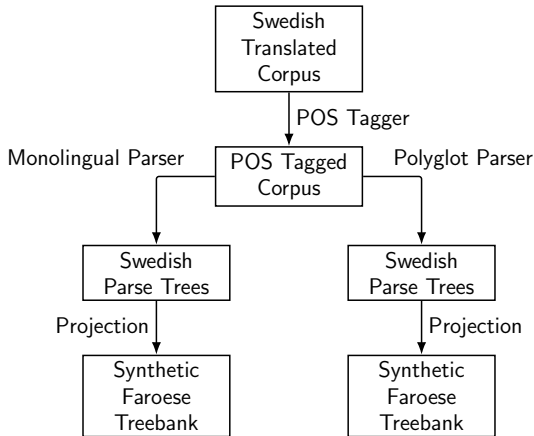


Figure 2: Overview of the *monolingual* and *polyglot* parse experiments using Swedish translations as an example. This process is repeated for all source languages.

**Synthetic Source Treebanks** Source translations are tokenized with UDPipe (Straka and Straková, 2017) by Tyers et al. (2018). For each source language, the POS model trained on the full training data (see previous section) is used to tag the tokenized translations. Once the text is tagged, we predict dependency arcs and labels with the parsing models of the previous section, and use annotation projection (described below) to provide syntactic annotations for the target sentences.

**Annotation Projection** Once the synthetic source treebanks are compiled, i.e. the translations are parsed, the annotations are then projected from the source translations to the target language using the word alignments and Tyers et al.’s projection tool, resulting in a Faroese treebank. In some cases, not all tokens are aligned and Tyers et al. (2018) work around this by falling back to a 1:1 mapping between the target index and the source index. There are also cases where there is a mismatch in length between the source and target sentences and some dependency structures cannot be projected to the target language. Tyers et al.’s projection setup removes unsuitable projected trees containing e.g. more than one root token, a

token that is its own head or a token with a head outside the range of the sentence.

**Multi-source Projection** For multi-source projection, the four source-language dependency trees for a Faroese sentence are projected into a single graph, scoring edges according to the number of trees that contain them (Sagae and Lavie, 2006; Nivre et al., 2007). The dependency structure is first built by voting over the directed edges. Afterward, dependency labels and POS tags are decided using the same voting procedure. The process is illustrated in Fig. 3.

**Target Tagging and Parsing Models** At this stage we have Faroese treebanks to train our POS tagging and parsing models. The Faroese treebanks come in two variants: the result of projection from source trees produced by either 1) a monolingual, or 2) the polyglot model. For each case, we train our POS tagging and parsing models directly on these synthetic treebanks and do not make use of word embeddings as we do not have them for Faroese.

**Multi-treebank Target Parsing** Since we have several synthetic Faroese treebanks, we have the option of training on a single treebank or using a multi-treebank approach where we train on all target treebanks in the same way as we did for inducing the polyglot source model. The process is illustrated in Fig. 4. When training a multi-treebank target model, for each target treebank, we add a treebank embedding denoting the source model used to project annotations to the target treebank. At predict time, we must include one of these treebank embeddings as input to the model. As we do not have real Faroese data in our target training treebanks, we must choose the treebank embedding of one of the synthetic target treebanks. Stymne et al. (2018) introduce the term “proxy treebank” to refer to cases where the test treebank is not in the training set and a treebank embedding from the training set must be used instead.

## 4 Experiments

In this section, we describe our experiments, which include a replication of the main findings of Tyers et al. (2018), using AllenNLP (Gardner et al., 2018) for POS tagging and parsing instead of UDPipe (Straka and Straková, 2017).<sup>12</sup>

<sup>12</sup>All of the code and scripts to reproduce the experiments can be found at <https://github.com/Jbar-ry/>

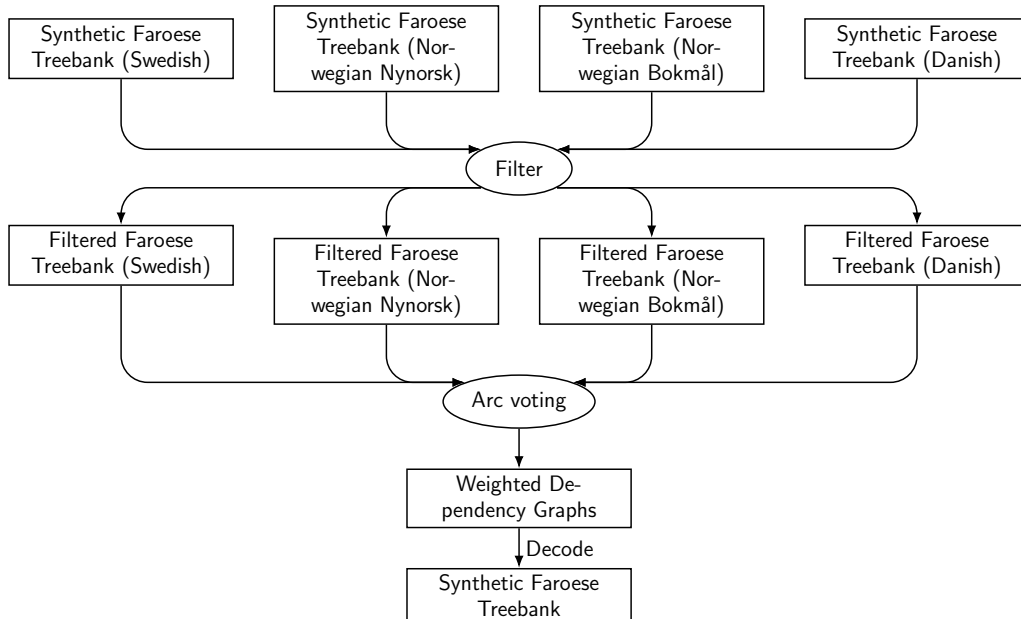


Figure 3: Multi-source projection. The source language is listed in brackets.

#### 4.1 Details

The hyper-parameters of our POS tagging and parsing models are given in Table 1. For POS tagging, we adopt a standard architecture with a word and character-level Bi-LSTM (Plank et al., 2016; Graves and Schmidhuber, 2005) to learn context-sensitive representations of our words. These representations are passed to a multilayer perceptron (MLP) classifier followed by a softmax function to choose a tag with the highest probability. For both the POS tagging and parsing models, we use a word embedding dimension of size 100 and a character embedding dimension of size 64. POS tag embeddings of dimension 50 are included in the parser. We train our Faroese models for fifty epochs. We do not split the synthetic Faroese treebanks into training/development portions though we suspect doing so will help the models to not overfit on the training data. For all experiments we report labelled attachment scores produced by the official CoNLL 2018 evaluation script.<sup>13</sup>

#### 4.2 Results

The development results of our monolingual and polyglot models on the source language treebanks are shown in Table 2. The results for the polyglot

model are better for three out of four source languages, whereas for no.nynorsk, the monolingual model marginally outperforms the polyglot one. These results suggest that the polyglot model will contribute better syntactic annotations for Faroese treebanks.

The statistics of the filtered Faroese treebanks obtained via projection with our source parsing models are given in Table 3. The treebank sizes are fairly similar regardless of whether source annotations are provided by a monolingual or a polyglot model which is expected because the word alignments are the major factor in determining whether a projection is successful. There is a proportionally lower number of sentences for multi-source projection. This is because this method only uses the intersection of sentences which are present across all synthetic treebanks after filtering. The treebank originating from Norwegian Bokmål has the highest number of valid sentences, suggesting that it could be a good candidate for projection to Faroese. It also has the highest source language parsing accuracy (Table 2).

The results of training on our various synthetic Faroese treebanks and predicting the Faroese test set are shown in the first result column of Table 4 (SINGLE). In terms of monolingual vs. polyglot, we find that projecting from a polyglot model helps with four out of the five possible treebanks (with three of them being statistically

multilingual-parsing

<sup>13</sup>[https://github.com/ufal/conll2018/blob/master/evaluation\\_script/conll18\\_ud\\_eval.py](https://github.com/ufal/conll2018/blob/master/evaluation_script/conll18_ud_eval.py)

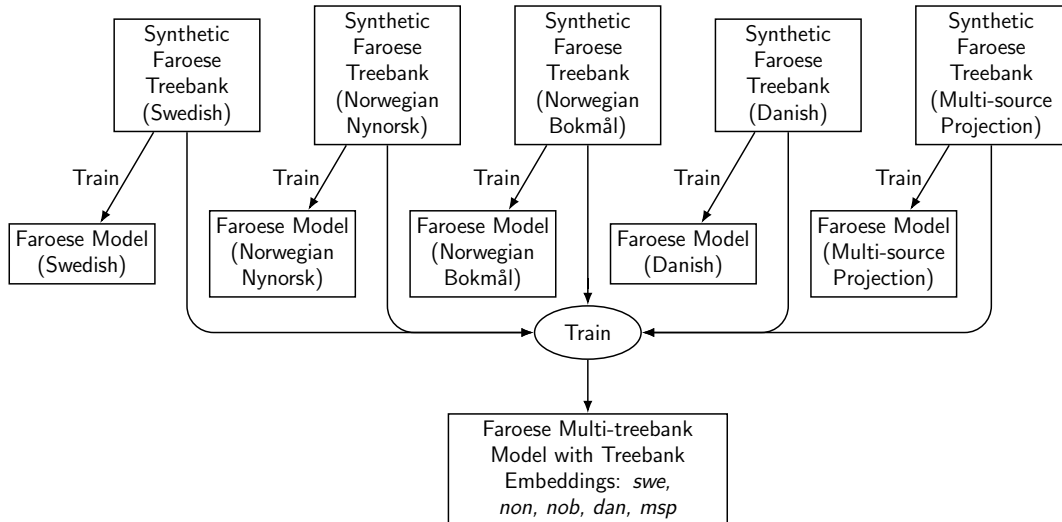


Figure 4: Single versus multi-treebank training. The source language is listed in brackets.

POS Tagger Architecture	
Parameter	Value
Char-BiLSTM layers	2
BiLSTM layers	2
BiLSTM size	400
Dropout LSTMs	0.33
Dropout MLP	0.33
Dropout embeddings	0.33
Nonlinear act. (MLP)	ELU
Parser Architecture	
Parameter	Value
Char-BiLSTM layers	2
BiLSTM layers	3
BiLSTM size	400
Arc MLP size	500
Label MLP size	100
Dropout LSTMs	0.33
Dropout MLP	0.33
Dropout embeddings	0.33
Nonlinear act. (MLP)	ELU
Embeddings	
Parameter	Size
Word embedding (both)	100
Char embedding (both)	64
POS embedding (parser)	50
Treebank embedding (both)	12
Training	
Parameter	Value
Optimizer	Adam
Learning rate	0.001
Adam epsilon	1e-08
beta1 (both)	0.9
beta2 (parser)	0.9
beta2 (tagger)	0.999

Table 1: Chosen hyperparameters for our POS tagging and parsing models. *both* means the feature is common to both the POS tagger and parser.

TREEBANK	MONOLINGUAL	POLYGLOT
da_ddt	81.10	<b>82.75</b>
sv_talbanken	80.61	<b>83.85</b>
no_nynorsk	<b>88.54</b>	88.29
no_bokmaal	89.29	<b>90.29</b>
average	84.88	<b>86.30</b>

Table 2: Source model LAS scores on the development treebanks using silver POS tags.

SOURCE	MONOLINGUAL	POLYGLOT
Danish	13,950	13,944
Swedish	10,894	10,874
Norwegian Nynorsk	13,177	13,194
Norwegian Bokmål	17,345	17,378
Multi-source	6,716	6,833

Table 3: The number of valid sentences in the Faroese synthetic treebank for each source language after annotation projection and sentence filtering.

formed by the monolingual model using Norwegian Nynorsk for projection though the difference is not statistically significant. On the source side, the monolingual Norwegian Nynorsk model also performed slightly better than the polyglot model (Table 2). This observation supports the intuition that the quality of the projected annotations can be improved by contributing better source annotations, i. e. improving the source model(s) is one way to improve performance of the target model. This is supported by the fact that the source lan-

significant).<sup>14</sup> The polyglot model was outper-

<sup>14</sup>Statistical significance is tested with `udapi-python`

<https://github.com/udapi/udapi-python>. LAS differences are reported as significant if  $p < 0.05$ .

SOURCE LANGUAGE	SOURCE MODEL	TARGET MODEL	
		SINGLE	MULTI
Danish	Monolingual	61.24	63.40
	Polyglot	65.29 <sup>†</sup>	<b>65.53<sup>†</sup></b>
Swedish	Monolingual	65.93	66.15
	Polyglot	68.60 <sup>†</sup>	<b>69.69<sup>†</sup></b>
Norwegian Nynorsk	Monolingual	70.27	<b>71.51</b>
	Polyglot	69.80	71.13
Norwegian Bokmål	Monolingual	67.46	67.94
	Polyglot	70.51 <sup>†</sup>	<b>70.58<sup>†</sup></b>
Multi-source	Monolingual	68.00	69.80
	Polyglot	68.55	<b>70.07</b>
Average	Monolingual	66.58	67.76
	Polyglot	68.55	<b>69.40</b>

Table 4: LAS on the target Faroese test treebank. *Single* refers to using a single synthetic Faroese treebank to train a Faroese model, *Multi* uses both a multi-treebank POS tagger and a multi-treebank parser with all synthetic Faroese treebanks. The multi-treebank model is tested with each of the five training treebanks (four projected from individual source languages and one using multi-source projection) as proxy treebank. Statistically significant differences between the monolingual and polyglot setting are indicated by <sup>†</sup> for each result pair, excluding averages.

SOURCE LANGUAGE	MONOLINGUAL	POLYGLOT
Danish	61.13	<b>64.43</b>
Swedish	63.19	<b>67.46</b>
Norwegian Nynorsk	68.72	<b>69.28</b>
Norwegian Bokmål	66.13	<b>68.77</b>
Multi-source	68.00	<b>68.55</b>
Average	65.43	<b>67.70</b>

Table 5: LAS scores between target models trained on the subset of sentences eligible for multi-source projection (with annotations from the stated source).

guage with the highest LAS (Norwegian Bokmål) is also the best choice for projection (in this single target model setting).

The multi-source approach was not that effective in our case and some individual better sources were able to surpass this combination approach. One could argue that this may be due to the lower amount of training data when using the multi-source treebank. We test this hypothesis by only including those sentences which contributed to multi-source projection in the single-source synthetic treebanks. The results are given in Table 5. Comparing the results in Tables 4 and 5, we see that LAS scores tend to be slightly lower than on the version which included all target sen-

WORK	RESULT
Rosa and Mareček (2018)	49.4
Tyers et al. (2018)	64.4
Our implementation of Tyers et al. (2018)	68.0
Our Best Model	<b>71.5</b>

Table 6: Comparison to previous work. LAS on Faroese test set. Note that the first results uses predicted segmentation and tokenization whereas the rest used gold.

tences, indicating that we did lose some information by filtering out a large number of sentences. However, Norwegian Nynorsk still outperforms the multi-source model for the monolingual setting and both Norwegian models perform better than the multi-source model in the polyglot setting, suggesting that size alone does not explain the under-performance of the multi-source model. It is also worth noting that polyglot training is superior to all monolingual models which hints that for no\_nynorsk (the previously better performing model), the monolingual model was not able to achieve its full potential with the reduced data while the polyglot model was able to provide richer annotations.

Another reason why the multi-source model does not work as well in our experiments as it does in those of Tyers et al. (2018) might be that we use pre-trained embeddings whereas Tyers et al. (2018) do not. In this way, our monolingual models are stronger and likely do not benefit as much from voting.

The second result column (MULTI) of Table 4 shows the effect of training a multi-treebank POS tagger and parser on the Faroese treebanks created by each of the four source languages as well as the treebank which is produced by multi-source projection. This experiment is orthogonal to the experiment using a polyglot model on the source side and so we also test a combination of polyglot source side parsing and multi-treebank target side parsing. We see improvements over the single treebank setting for all cases.<sup>15</sup>

<sup>15</sup>The multi-treebank tagger closely resembles the dependency parser, where we add a treebank embedding and optimize for average accuracy across the included treebanks. To the best of our knowledge, this is the first reported use of a multi-treebank POS tagger using a treebank embedding (Stymne et al., 2018). We also tested the effect of training only the dependency parser using multiple treebanks



Table 6 places our systems in the context of previous results on the same Faroese test set. The highest scoring system in the 2018 CoNLL shared task was that of Rosa and Mareček (2018) who achieved a LAS score of 49.4 on the Faroese test set. Note that they use predicted tokenization and segmentation whereas our experiments and Tyers et al.’s use gold tokenization and segmentation, which provides a small artificial boost. Tyers et al. (2018) report an LAS of 64.43 with a monolingual multi-source approach. Our implementation which uses a different parser (AllenNLP versus UDPipe) and pre-trained word embeddings achieves an LAS of 68. Our highest score of 71.51 is achieved through the combination of projecting from strong monolingual source models and then training multi-treebank POS tagging and parsing models on the outputs.

## 5 Conclusion

We have presented parsing results on Faroese, a low-resource language, using annotation projection from multiple monolingual sources versus a single polyglot model. We also extended the idea of multi-treebank learning to the target treebanks.

The results of our experiments show that the use of a polyglot source model helps in four out of five cases using single treebank target models. The two source languages that have lowest LAS when using monolingual parsers, namely Danish and Swedish, see significant improvements when switching to a polyglot model. Our best performing single target model is trained on Faroese trees projected from Norwegian Bokmål trees produced by a polyglot model. However, the strongest language with monolingual modelling, Norwegian Nynorsk, does not benefit from switching to a polyglot model. When we filtered the target treebank to the subset of sentences selected by multi-source projection, the polyglot model is superior to all five monolingual models, even outperforming the Norwegian Nynorsk model. One explanation of the improvements seen with polyglot modelling is that it introduces a new interaction point for cross-lingual features via the feature extractor of the polyglot parser. With monolingual source models, cross-lingual features only interact indirectly in the graph-decoding stage of multi-source projection.

---

but found that it always helps to also perform multi-treebank training for the POS tagger.

We also applied the multi-treebank approach to the target-side POS tagger and parser and see improvements for all settings. The overall best result is with the setting that uses monolingual source models to create the source trees that are projected to Faroese and combined in a multi-treebank model. The proxy treebank for the multi-treebank model is the treebank that also gave best results with single treebank target models, projected from Norwegian Nynorsk.

We presented a simple solution to deal with using multiple pre-trained embeddings in a model with a shared vocabulary. It was a rather naïve solution and we want to explore the use of available cross-lingual word embedding tools. Additionally, the use of contextual embeddings such as ELMo (Peters et al., 2018) or multilingual BERT (Devlin et al., 2019) would likely provide better representations, with the effect of contributing better annotations for the target language. Indeed, recent work has already shown promising work in this area (Schuster et al., 2019; Kondratyuk, 2019).

In the multi-source projection experiments, our criteria for filtering is based on whether the sentence was present across all target treebanks and more sophisticated approaches could be used to select better training instances as in Plank and Agić (2018).

More generally, we would like to investigate how our findings might change when the number of source languages or treebanks is changed and how the observations carry over to other languages than Faroese. It would also be interesting to use multiple sources of arc weights in a dense graph as in (Agić et al., 2016) but with models induced from training on multiple source languages together. To work with language pairs with more deviating word orders and/or translations that are not word-for-word, the choice of word alignment algorithm and the projection algorithm may have to be revised.

## Acknowledgments

This research is supported by Science Foundation Ireland through the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. We thank Tyers et al. (2018) for releasing their data and the anonymous reviewers for their helpful feedback.

## References

- Željko Agić, Anders Johannsen, Barbara Plank, Héctor Martínez Alonso, Natalie Schluter, and Anders Søgaard. 2016. [Multilingual projection for parsing truly low-resource languages](#). *Transactions of the Association for Computational Linguistics*, 4:301–312.
- Rami Al-Rfou’, Bryan Perozzi, and Steven Skiena. 2013. [Polyglot: Distributed word representations for multilingual NLP](#). In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192, Sofia, Bulgaria. Association for Computational Linguistics.
- Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. 2016. [Many languages, one parser](#). *Transactions of the Association for Computational Linguistics*, 4:431–444.
- Yoeng-Jin Chu and Tseng-Hong Liu. 1965. On the shortest arborescence of a directed graph. *Scientia Sinica*, 14:1396–1400.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Timothy Dozat and Christopher D. Manning. 2017. [Deep biaffine attention for neural dependency parsing](#). In *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)*.
- Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. [A simple, fast, and effective reparameterization of IBM model 2](#). In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.
- Jack Edmonds. 1967. Optimum branchings. *Journal of Research of the national Bureau of Standards B*, 71(4):233–240.
- Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. [AllenNLP: A deep semantic natural language processing platform](#). In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 1–6, Melbourne, Australia. Association for Computational Linguistics.
- Alex Graves and Jürgen Schmidhuber. 2005. Frame-wise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610.
- Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015. [Cross-lingual dependency parsing based on distributed representations](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1234–1244, Beijing, China. Association for Computational Linguistics.
- 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.
- Daniel Kondratyuk. 2019. [75 languages, 1 model: Parsing universal dependencies universally](#). *CoRR*, abs/1904.02099.
- Teresa Lynn, Jennifer Foster, Mark Dras, and Lamia Tounsi. 2014. [Cross-lingual transfer parsing for low-resourced languages: An Irish case study](#). In *Proceedings of the First Celtic Language Technology Workshop*, pages 41–49, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.
- Ryan McDonald, Slav Petrov, and Keith Hall. 2011. [Multi-source transfer of delexicalized dependency parsers](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Ailsa Meechan-Maddon and Joakim Nivre. 2019. How to parse low-resource languages: Cross-lingual parsing, target language annotation, or both? In *Proceedings of DepLing*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. [Polyglot contextual representations improve crosslingual transfer](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3912–3918, Minneapolis, Minnesota. Association for Computational Linguistics.
- Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, 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, John Bauer, Sandra Bellato, Kepa Bengoetxea, 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, Giuseppe G. A. Celano, 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 Droganova, 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, 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, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Tolga Kayadelen, Václava Kettnerová, Jesse Kirchner, Natalia Kotsyba, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê H'ông, 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 Manuring, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisepp, Pinkey Nainwani, Juan Ignacio Navarro Horňáček, Anna Nedoluzhko, Gunta Nešpore-Bėrzkalne, Lng Nguy'ên Thì, Huy'ên Nguy'ên Thì Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayo Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Siyao Peng, Cene-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 Roca, Olga Rudina, Shoval Sadde, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamel 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, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle Tellier, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Veronika Vincze, Lars Wallin, 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 2.2](#). LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007. [The CoNLL 2007 shared task on dependency parsing](#). In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*, pages 915–932, Prague, Czech Republic. Association for Computational Linguistics.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Barbara Plank and Željko Agić. 2018. [Distant supervision from disparate sources for low-resource part-of-speech tagging](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 614–620, Brussels, Belgium. Association for Computational Linguistics.
- Barbara Plank, Anders Søgaard, and Yoav Goldberg. 2016. [Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 412–418, Berlin, Germany. Association for Computational Linguistics.
- Rudolf Rosa and David Mareček. 2018. [CUNI x-ling: Parsing under-resourced languages in CoNLL 2018 UD shared task](#). In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text*

- to *Universal Dependencies*, pages 187–196, Brussels, Belgium. Association for Computational Linguistics.
- Kenji Sagae and Alon Lavie. 2006. [Parser combination by reparsing](#). In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 129–132, New York City, USA. Association for Computational Linguistics.
- Michael Schlichtkrull and Anders Søgaard. 2017. [Cross-lingual dependency parsing with late decoding for truly low-resource languages](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 220–229, Valencia, Spain. Association for Computational Linguistics.
- Tal Schuster, Ori Ram, Regina Barzilay, and Amir Globerson. 2019. [Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1599–1613, Minneapolis, Minnesota. Association for Computational Linguistics.
- Aaron Smith, Bernd Bohnet, Miryam de Lhoneux, Joakim Nivre, Yan Shao, and Sara Stymne. 2018. [82 treebanks, 34 models: Universal dependency parsing with multi-treebank models](#). In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 113–123, Brussels, Belgium. Association for Computational Linguistics.
- Milan Straka and Jana Straková. 2017. [Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe](#). In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.
- Sara Stymne, Miryam de 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. Association for Computational Linguistics.
- Jörg Tiedemann. 2016. Opus-parallel corpora for everyone. *Baltic Journal of Modern Computing*, page 384.
- Jrg Tiedemann and Željko Agić. 2016. Synthetic treebanking for cross-lingual dependency parsing. *Journal of Artificial Intelligence Research*, 5.
- 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, Brussels, Belgium. Association for Computational Linguistics.
- David Vilares, Carlos Gómez-Rodríguez, and Miguel A. Alonso. 2016. [One model, two languages: training bilingual parsers with harmonized treebanks](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 425–431, Berlin, Germany. Association for Computational Linguistics.
- David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on Human language technology research*, pages 1–8. Association for Computational Linguistics.
- Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, 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*, pages 1–21, Brussels, Belgium. Association for Computational Linguistics.
- Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinkova, Jan Hajic jr., Jaroslava Hlavacova, Václava Kettnerová, Zdenka Uresova, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droганova, Héctor Martínez Alonso, Çağr Çöltekin, Umut Sulubacak, Hans Uszkor-eit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonca, Tatiana Lando, Rattima Nitisaraj, and Josie Li. 2017. [Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies](#). In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.
- 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*.