Parsing in the absence of related languages: Evaluating low-resource dependency parsers on Tagalog

Angelina Aquino and **Franz de Leon**
Digital Signal Processing Laboratory, Electrical and Electronics Engineering Institute
University of the Philippines, Diliman, Quezon City, Philippines
{angelina.aquino, franz.de.leon}@eee.upd.edu.ph

Abstract

Cross-lingual and multilingual methods have been widely suggested as options for dependency parsing of low-resource languages; however, these typically require the use of annotated data in related high-resource languages. In this paper, we evaluate the performance of these methods versus monolingual parsing of Tagalog, an Austronesian language which shares little typological similarity with any existing high-resource languages. We show that a monolingual model developed on minimal target language data consistently outperforms all cross-lingual and multilingual models when no closely-related sources exist for a low-resource language.

1 Introduction

Dependency parsing is a fundamental component of many natural language understanding (Roth and Lapata, 2016; Zhang et al., 2018) and machine translation systems (Ding and Palmer, 2005; Chen et al., 2017). State-of-the-art parsers which annotate syntactic dependencies from raw text have achieved high accuracy for languages with large datasets but continue to yield poor results for low-resource languages which have little to no annotated data (Zeman et al., 2018).

Various methods have been proposed to solve the problem of dependency parsing in a low-resource setting, including cross-lingual transfer (Zeman and Resnik, 2008; McDonald et al., 2011), multilingual modeling (Duong et al., 2015; Ammar et al., 2016), and annotation projection (Hwa et al., 2002; Agić et al., 2016). These methods have been shown to be effective on target languages when datasets (such as treebanks and parallel corpora) are readily available for closely-related source languages; however, would the same hold true in the absence of related language data?

Such is the problem for Tagalog, an Austronesian language of the Philippines with over 25 million speakers worldwide (Eberhard et al., 2020). Despite its widespread use in both spoken and digital domains, it remains largely under-resourced, lacking basic language processing resources such as syntactic treebanks and parsers. Moreover, while dependency treebanks are available for Indonesian, another Austronesian language, the extensive phylogenetic distance between the two languages (Greenhill and Gray, 2009; Reid, 2018) suggests that Indonesian may have too many typological differences to serve as an effective source language for Tagalog.

In this paper, we investigate the performance on Tagalog of three strategies for low-resource dependency parsing: monolingual modeling (using only minimal target language data), cross-lingual modeling (using only data from similar source languages), and multilingual modeling (using data from both target and non-target languages). We present a new Tagalog dependency treebank on which to train and test these approaches together with available treebanks from the Universal Dependencies (UD) project (Zeman et al., 2020), and compare our results to those of previous studies.

2 Related work

To our knowledge, only two dependency treebanks for Tagalog have been created prior to this work. The first is the Tagalog Dependency Treebank (Manguilimotan and Matsumoto, 2011), which includes 2,500

Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020), pages 8–15

PUBLICATION	TREEBANK	NO. OF TOKENS	ANNOTATION	POS	FEATS	LEMM	UAS	LAS
Manguilimotan & Matsumoto (2011)	TDT (Test)	6,557	own native	88.96	—	—	75.90	—
Dehouck & Denis(2019)	TRG	292	UD	—	—	—	70.89	50.38
Kondratyuk & Straka (2019)	TRG	292	UD	61.64	35.27	75.00	64.73	39.38
This work	Ugnayan	1,011	UD	80.54	—	85.47	63.47	55.37

Table 1: Overview of best reported results for dependency parsing of Tagalog treebanks.

sentences annotated with part-of-speech (POS) tags and dependency heads for each word. However, the treebank does not contain labels for dependency relations, nor any other levels of annotation.

The second is the TRG treebank (Samson, 2018) released as part of UD since version 2.2. The treebank contains 55 sentences taken from grammar examples in the Tagalog Reference Grammar (Schachter and Otanes, 1972). Upon inspection, we found that most of these were simple declarative sentences which used the basic predicate-initial word order of Tagalog and contained only one or two arguments. Moreover, the treebank did not contain any examples of other sentence types such as compound sentences, interrogatives, and imperatives, nor of common grammatical components such as adjectival modifiers and plural forms. Table 1 provides a summary of parsing results previously reported for these treebanks.

3 Language data

Tagalog treebank. In order to properly assess the performance of a dependency parser on a target language, we need to have a treebank available in that language which more extensively captures its grammatical complexities and contains universally comparable annotations. Since neither of the previous Tagalog treebanks fulfill both requirements, we developed Ugnayan, a new Tagalog dependency treebank manually annotated in the UD framework. The treebank currently consists of 94 sentences (1011 tokens) taken from educational texts (Almario and Tan, 2016). These sentences include examples of various syntactic phenomena such as compound and complex sentences, clausal modifiers, question forms, and sentence inversion.

Source treebanks. To train the cross-lingual and multilingual models, we also needed to identify which UD languages with available training data are most similar to Tagalog. For this, we used a WALS-reliant distance measure, which compares the typological similarity of a source language S and a target language T based on their features as described in the World Atlas for Language Structures (Dryer and Haspelmath, 2013). We use the distance measure defined by Agić (2017) as the Hamming distance d_h between the WALS feature vectors v_S and v_T for the source and target languages, normalized with respect to the number of features $f_{S,T}$ which are non-empty for both S and T. The resulting WALS measure d_W is given as:

$$\mathrm{d}_W(S,T) = \frac{d_h(v_S,v_T)}{f_{S,T}}$$

Using this measure, we found that the five closest source languages for Tagalog included Indonesian (the only other Austronesian UD language), Vietnamese (the only Austro-Asiatic UD language), and three Indo-European languages: Ukrainian, Romanian, and Catalan (see Table 2). Interestingly, we also found that Tagalog was *not* among the five most similar sources for any of the languages above. Unsurprisingly, Ukrainian was much closer to its Slavic neighbors with distances well below 0.2, while Romanian and Catalan were all within distances of 0.3 of other Romance languages. But even among the Asian sources, Indonesian and Vietnamese were much closer to each other, to Mandarin, and even

T	*Tagalog*	d_W	*Indonesian*	d_W	*Ukrainian*	d_W	*Vietnamese*	d_W	*Romanian*	d_W	*Catalan*	d_W
S	Indonesian	0.446	Vietnamese	0.275	Slovenian	0.029	Indonesian	0.275	Italian	0.179	Italian	0.183
	Ukrainian	0.455	Arabic	0.360	Russian	0.054	Mandarin	0.331	Portuguese	0.182	Romanian	0.211
	Vietnamese	0.469	Ukrainian	0.393	Polish	0.087	Ukrainian	0.358	Catalan	0.211	Spanish	0.262
	Romanian	0.471	Mandarin	0.401	Serbo-Croat	0.150	Slovenian	0.385	Bulgarian	0.246	Portuguese	0.267
	Catalan	0.472	Polish	0.416	Estonian	0.175	Portuguese	0.392	Greek	0.256	French	0.297

Table 2: Top 5 most similar source languages S for target languages T, with corresponding WALS distances d_W. Lower d_W indicates higher typological similarity. Distances are symmetric; i.e. $d_W(S,T) = d_W(T,S)$.

Language	Treebank	Train	Dev	Test
Tagalog	Ugnayan			1.0k
Tagalog	TRG			0.3k
Indonesian	GSD	97.5k	12.6k	
Ukrainian	IU	92.4k	12.6k	
Vietnamese	VTB	20.3k	11.5k	
Romanian	Nonstandard	410.4k	18.6k	
Romanian	RRT	185.1k	17.1k	
Catalan	AnCora	416.7k	56.3k	

Table 3: UD v2.6 treebanks used with sizes in tokens.

to other Indo-European languages than they were to Tagalog. This supports the findings by Georgi et al. (2010) that phylogenetic relatedness does not guarantee typological similarity.

For our cross-lingual modeling, we selected all UD v2.6 treebanks with available train and dev sets in the source languages identified above. We also decided to train a model on the Tagalog TRG treebank as a point of comparison. We report the sizes of these data sets in Table 3.

4 Evaluation of parsing models

Methodology. To train parsing models on the treebanks above, we used UDPipe (Straka and Straková, 2017), a pipeline for processing of CoNLL-U treebanks which has served as the baseline system in several CoNLL UD Shared Tasks (Zeman et al., 2017; Zeman et al., 2018). We trained cross-lingual models for each of the identified source treebanks using their specified train and dev partitions, as well as a model using all test data in TRG, and tested these models on all test data in Ugnayan. We performed ten-fold cross-validation to evaluate monolingual models trained on Ugnayan, with a train/dev/test partition of roughly 80/10/10 for each iteration. We also used cross-validation on the multilingual models, which were trained using each of the ten Ugnayan train/dev partitions combined with the individual source treebanks, and tested on the ten Ugnayan test partitions. We used the default settings on UDPipe 3.1 for all training and testing instances.

We then investigated the performance on Ugnayan of the two approaches previously applied to TRG as described in Section 2 by evaluating the pre-trained Indonesian model of Stanza (previously StanfordNLP), a neural pipeline developed by Qi et al. (2020) which reportedly outperforms all submissions to the CoNLL 2018 UD Shared Task for the low-resource categories on all metrics. We selected this as an approximation of the neural parser used by Dehouck & Denis (2019), which was based on the parser of Dozat et al. (2017) currently integrated into Stanza. We also evaluated an updated version of UDify, the multilingual parser by Kondratyuk & Straka (2019).

Results. Table 4 reports the performance of each model tested on the Ugnayan treebank. We recorded

			Tokenization			*Tagging*		*Parsing*	
	Parser	Model	Token	Word	Sent	UPOS	Lemm	UAS	LAS
monolingual	UDPipe	tl-ugnayan	**99.27**	**95.67**	**95.41**	**80.54**	**85.47**	**63.47**	**55.37**
		tl-trg	98.08	86.00	64.04	41.58	65.72	29.23	13.11
cross-lingual	UDPipe	id-gsd	97.40	85.22	90.32	27.45	65.64	18.81	**9.69**
		uk-iu	**97.56**	**85.43**	63.41	12.80	65.55	15.48	8.31
		vi-vtb	74.31	63.81	90.62	22.83	49.44	7.24	3.67
		ro-nonstandard	92.65	81.00	89.80	26.15	39.04	16.56	5.64
		ro-rrt	96.95	84.81	91.98	26.07	48.01	**20.03**	8.15
		ca-ancora	97.40	85.22	94.68	23.70	50.96	14.49	4.89
	Stanza	id-gsd	97.40	85.22	**95.14**	**28.60**	**66.60**	14.88	5.76
multilingual	UDPipe	tl-ugnayan + id-gsd	98.67	94.17	**98.57**	78.16	83.46	48.20	39.73
		tl-ugnayan + uk-iu	99.07	**95.49**	90.46	78.57	**85.32**	**58.54**	48.31
		tl-ugnayan + vi-vtb	98.49	95.13	93.95	**79.28**	84.79	58.05	**48.65**
		tl-ugnayan + ro-nonstd.	98.30	94.69	94.79	71.00	71.18	45.93	34.26
		tl-ugnayan + ro-rrt	**98.71**	95.04	96.42	77.61	80.51	46.70	37.12
		tl-ugnayan + ca-ancora	97.69	94.32	95.23	75.86	80.57	43.01	32.68
	UDify	universal	—*	—*	—*	59.62*	70.92*	51.96*	32.09*

Table 4: F_1 scores on parsing tasks for each parser and model tested on the Ugnayan treebank. Scores for all models trained with Ugnayan data were averaged over 10-fold cross-validation. *(*UDify uses gold tokenization.)* **Bold**: highest scores per method. Gray: highest scores across all models.

the F_1 scores automatically generated by each of the three parsers on the following tasks: token, word, and sentence tokenization, universal part-of-speech tagging, lemmatization, unlabeled attachment, and labeled attachment. Since the Ugnayan treebank currently does not contain features or language-specific part-of-speech tags, metrics involving those annotations were excluded from the report. For the monolingual and multilingual models trained on Ugnayan data, we present the average scores across ten iterations for cross-validation.

We find that the monolingual Ugnayan models, each trained on less than 90 sentences (or approximately 900 tokens), outperform all other models on the tagging and parsing tasks, and are surpassed only by the Tagalog-Indonesian mixed model on sentence tokenization. These results support the hypothesis by Zeman that "You can actually train a parser and get over 50% accuracy for many languages with just about 100 sentences," (Nivre et al., 2017) which has previously been shown for Indian languages (Ramasamy, 2014), Galician (Garcia et al., 2018), and Faroese (Meechan-Maddon and Nivre, 2019). Garrette and Baldridge (2013) have achieved similar POS tagging performance for Kinyarwanda and Malagasy using similarly limited annotation and graph-based label propagation onto larger amounts of raw text; here we show that supervised modeling using *only* limited annotation can yield good results.

We also observe that the cross-lingual models in particular score much lower than the Ugnayan models or any of the multilingual models on the UPOS, UAS, and LAS metrics. Interestingly, the multilingual models, which use Ugnayan training data together with each of the cross-lingual source treebanks, yield consistently *lower* accuracy than the monolingual models alone. This runs contrary to the findings of Meechan-Maddon & Nivre (2019) who observed that adding related language data to train a multilingual model further improves parsing accuracy. These results suggest that the typological distance between Tagalog and any of its closest UD languages may be too great for the latter to be useful as cross-lingual or multilingual source languages out of the box, and that upweighting of the Ugnayan data may be necessary to account for the size difference between the source and target training corpora.

As for the pre-trained parsers, the Stanza Indonesian model slightly outperforms its UDPipe equivalent on sentence tokenization, POS tagging, and lemmatization, while underperforming on UAS and LAS. Because of the large discrepancy between these results and the 70.89% UAS & 50.38% LAS previously reported for parsing TRG using an Indonesian-only model (Dehouck and Denis, 2019), we further investigated the performance of Indonesians models on both Ugnayan and TRG when gold tokenization and gold tags are made available. We found that UAS and LAS higher than 50% were achievable only with gold tags for both treebanks, and that these results could not be matched when parsing from raw text.

On the other hand, the UDify universal model outperforms all cross-lingual models and even the monolingual TRG model on all tagging and parsing tasks. This is quite remarkable, considering that no annotated Tagalog data was used to train the UDify model, although the availability of gold tokenization may have yielded a performance improvement compared to the other models which parse from raw text. These support the results of Kondratyuk & Straka (2019) which show that UDify's BERT pretraining and multilingual learning produce reasonably high scores even in a zero-resource setting.

5 Extended analysis

Performance on cross-domain data. So far, the experiments we have described above involved the use of Tagalog training and test data from the same corpus (Ugnayan) and domain (educational text). However, as Plank and Agić (2018) have observed, in-domain training naturally results in better performance than the cross-domain scenario for the same amount of data. To test the cross-domain performance of the Ugnayan model, we annotated an additional 7 sentences (265 tokens) of Tagalog news text, and evaluated each of the single-language UDPipe models above on this new dataset. We found that the tagging and parsing results of the Ugnayan model on the news dataset were significantly lower than the in-domain results (see Table 5a). But comparatively, the Ugnayan model still far surpassed any of the other single-language models: Tagalog-TRG achieved the closest scores for each task, followed ID-GSD for UPOS, RO-Nonstandard for UAS, and RO-RRT for LAS respectively. Aside from the dissimilarity of content between domains, the decrease in performance may be attributed to the length of the news sentences—each at least thrice as long as the average sentence in the Ugnayan treebank.

	a. News text (raw)			*b. Ugnayan (raw)*			*c. Ugnayan + POS tags*		*d. News text + POS tags*	
	tl-ugnayan	tl-trg	next-best	es-pud	en-pud	es+en+id	tl-ugnayan	next-best	tl-ugnayan	next-best
UPOS	**64.74**	35.62	28.46 (id-gsd)	25.13	28.98	28.08	—	—	—	—
UAS	**34.22**	18.00	15.83 (ro-nstd)	16.09	13.87	13.05	**74.12**	63.54 (en-pud)	**61.51**	59.25 (en-pud)
LAS	**25.86**	8.22	4.32 (ro-rrt)	5.65	4.81	4.79	**66.89**	49.77 (en-pud)	**52.08**	47.55 (en-pud)

Table 5: F_1 scores for extended analysis experiments. **Bold**: highest scores for each test set.

On historical contact and lexical similarity. In addition to cognates within the Austronesian language family, the Tagalog language is known to have incorporated many loanwords from both Spanish and English as a result of colonial occupation and, in the case of the latter, continued use within the country. To check whether either of these would be viable source languages, we trained single-language UDPipe models using the Spanish and English PUD treebanks. The results were roughly at par with the other source languages tested above (see Table 5b). We also trained a multilingual model using the combination of PUD treebanks for Spanish, English, and Indonesian (to account for Malay cognates), but found no significant improvement. More complex parsing models such as those proposed for code-switching (Partanen et al., 2018; Bhat et al., 2018) may be necessary to effectively utilize these treebanks for Tagalog parsing.

Parsing from gold-tagged data. This paper has largely focused on UD parsing from raw text. In a low-resource context, however, if good POS tagging performance can be achieved for the target language independent of treebank data, delexicalized parsing (which uses only POS tags as input) has been widely thought of as a suitable parsing strategy. In relation to this, we compare the performance of the single-language models when parsing with gold tags available for all test tokens. In contrast, the Ugnayan model achieves the best performance on both in-corpus and news data (see Tables 5c and 5d), outperforming all other single-language models in both cases by a comfortable margin; the next-best source model was English-PUD for both tasks and test sets. These provide partial support for the findings of Falenska and Çetinoğlu (2017), who have demonstrated that lexicalized parsing with limited target data generally outperforms delexicalized parsing with large amounts of source data when no good sources for the target language exist.

6 Conclusion

We have evaluated the performance of monolingual, cross-lingual, and multilingual parsing models on Ugnayan, a new Universal Dependencies treebank for the Tagalog language, given the task of dependency parsing from raw text. We have also identified potential source treebanks for the cross-lingual and multilingual models by measuring the typological similarity between Tagalog and existing high-resource UD languages. We find that a monolingual model trained on roughly 900 tokens of annotated target language data yields better performance than cross-lingual or multilingual models trained on 20,000 or more tokens of annotated data in other high-resource languages if these source languages exhibit low similarity to the target language. We also find that when no annotated training data is available for a target language, a model pre-trained on high-quality multilingual embeddings can give reasonable performance over cross-lingual models trained on individual source languages. We conclude that, when developing a parser for a low-resource language in the absence of any annotations for closely related languages, even a minimal amount of target language annotation greatly improves parsing performance over alternative methods.

We currently plan to expand the Ugnayan treebank in both size and scope, with additional annotations for morphological features and language-specific relation subtypes. Further investigation is warranted on the effects of domain coverage, lexical similarity, and word order differences (Ahmad et al., 2019) on parsing performance, as well as the application of other methods such as data augmentation (Vania et al., 2019) and annotation transfer using parallel corpora (Ma and Xia, 2014) in parser modeling.

Acknowledgements

This work is supported by the U.P. Teaching Assistantship Program. We thank Adrian Vidal, Herlan Benitez, and four anonymous reviewers for their incisive comments and valuable suggestions.

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.

Željko Agić. 2017. Cross-lingual parser selection for low-resource languages. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 1–10, Gothenburg, Sweden, May. Association for Computational Linguistics.

Wasi Ahmad, Zhisong Zhang, Xuezhe Ma, Eduard Hovy, Kai-Wei Chang, and Nanyun Peng. 2019. On difficulties of cross-lingual transfer with order differences: A case study on 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 2440–2452, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Ani Rosa Almario and Yvette Tan. 2016. *Leveled Readers in Filipino – Isang Kakaibang Araw – Alamin Natin ang mga Anyong-Tubig sa Pilipinas!* Department of Education, Philippines.

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.

Irshad Bhat, Riyaz A. Bhat, Manish Shrivastava, and Dipti Sharma. 2018. Universal dependency parsing for Hindi-English code-switching. 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 987–998, New Orleans, Louisiana, June. Association for Computational Linguistics.

Huadong Chen, Shujian Huang, David Chiang, and Jiajun Chen. 2017. Improved neural machine translation with a syntax-aware encoder and decoder. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1936–1945, Vancouver, Canada, July. Association for Computational Linguistics.

Mathieu Dehouck and Pascal Denis. 2019. Phylogenic multi-lingual 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 192–203, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Yuan Ding and Martha Palmer. 2005. Machine translation using probabilistic synchronous dependency insertion grammars. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 541–548, Ann Arbor, Michigan, June. Association for Computational Linguistics.

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, Vancouver, Canada, August. Association for Computational Linguistics.

Matthew S. Dryer and Martin Haspelmath, editors. 2013. *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. 2015. 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 and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 845–850, Beijing, China, July. Association for Computational Linguistics.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig, editors. 2020. *Ethnologue: Languages of the World*. SIL International, Dallas, Texas, 23 edition.

Agnieszka Falenska and Özlem Çetinoğlu. 2017. Lexicalized vs. delexicalized parsing in low-resource scenarios. In *Proceedings of the 15th International Conference on Parsing Technologies*, pages 18–24, Pisa, Italy, September. Association for Computational Linguistics.

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(1):91–122.

Dan Garrette and Jason Baldridge. 2013. Learning a part-of-speech tagger from two hours of annotation. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 138–147, Atlanta, Georgia, June. Association for Computational Linguistics.

Ryan Georgi, Fei Xia, and William Lewis. 2010. Comparing language similarity across genetic and typologically-based groupings. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 385–393, Beijing, China, August. Coling 2010 Organizing Committee.

Simon J. Greenhill and Russell D. Gray. 2009. Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In Alexander Adelaar and Andrew Pawley, editors, *Austronesian Historical Linguistics and Culture History: A festschrift for Robert Blust*, chapter 22, pages 375–398. Pacific Linguistics, Canberra, Australia.

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*, pages 392–399, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing universal dependencies universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China, November. 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. Association for Computational Linguistics.

Erlyn Manguilimotan and Yuji Matsumoto. 2011. Dependency-based analysis for Tagalog sentences. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, pages 343–352, Singapore, December. Institute of Digital Enhancement of Cognitive Processing, Waseda 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., July. 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 the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 112–120, Paris, France, August. Association for Computational Linguistics.

Joakim Nivre, Daniel Zeman, Filip Ginter, and Francis M. Tyers. 2017. Tutorial on Universal Dependencies: Adding a new language to UD. Presented at the 15th Conference of the European Chapter of the Association for Computational Linguistics, April.

Niko Partanen, Kyungtae Lim, Michael Rießler, and Thierry Poibeau. 2018. Dependency parsing of code-switching data with cross-lingual feature representations. In *Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages*, pages 1–17, Helsinki, Finland, January. 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, October-November. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In Association for Computational Linguistics (ACL) System Demonstrations.

Logathan Ramasamy. 2014. *Parsing under-resourced languages: Cross-lingual transfer strategies for Indian languages*. Ph.D. thesis, Charles University, Prague, Czech Republic.

Lawrence Reid. 2018. Modeling the linguistic situation in the Philippines. *Senri Ethnological Studies*, 98:91–105.

Michael Roth and Mirella Lapata. 2016. Neural semantic role labeling with dependency path embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202, Berlin, Germany, August. Association for Computational Linguistics.

Stephanie Dawn Samson. 2018. *A treebank prototype of Tagalog*. Undergraduate thesis, University of Tübingen, Germany.

Paul Schachter and Fe T. Otanes. 1972. *Tagalog Reference Grammar*. University of California Press, Berkeley and Los Angeles, California.

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, August. Association for Computational Linguistics.

Clara Vania, Yova Kementchedjhieva, Anders Søgaard, and Adam Lopez. 2019. A systematic comparison of methods for low-resource dependency parsing on genuinely low-resource languages. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1105–1116, Hong Kong, China, November. 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*.

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 Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, 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 Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, 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 Mendonça, Tatiana Lando, Rattima Nitisaroj, 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, August. 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, October. Association for Computational Linguistics.

Daniel Zeman, Joakim Nivre, et al. 2020. Universal Dependencies 2.6. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, Brussels, Belgium, October-November. Association for Computational Linguistics.