Cross-lingual Constituency Parsing with Linguistic Typology Knowledge

Anonymous ACL-IJCNLP submission

Abstract

Cross-lingual Transfer learning (CLT) has successfully been applied to the dependency parsing task. This is the first work that evaluates a CLT based approach to the Constituency parsing task. Furthermore, we utilized the linguistic typology knowledge in WALS database to improve the cross-lingual transferring ability of our proposed parser.

1 Introduction

Constituency parsing is a classic NLP task which aims to construct a phrase-structure parse-tree to represent the syntax of a given sentence. Numerous approaches to the constituency-parsing task have been proposed in the past (Charniak, 2000; Collins, 2003; Petrov et al., 2006) including sophisticated neural-network based approaches to it (Kuncoro et al., 2016; Takase et al., 2018; Mrini et al., 2019; Yang and Deng, 2020). The state-of-the-art neural approaches to the constituency-parsing task are mono-lingual supervised approaches which require large amount of labelled data to be trained on, thus limiting their utility to only handful of high-resource languages. To address this issue of data-sparsity, researchers have proposed numerous unsupervised approaches to constituency-parsing (Kann et al., 2019; Zhao and Titov, 2021; Kim et al., 2020a; Wu et al., 2020). However these approaches significantly under-perform the monolingual supervised approaches.

In this work, we evaluate the performance of the cross-lingual variant of the popular Discriminative Recurrent Neural Network Grammar (RNNG) (Dyer et al., 2016) constituency parser. Cross-lingual Transfer-learning (CLT) typically involves training a model on the high-resource source-languages and applying it on a low-resource target-language. The CLT based approaches utilise various multilingual word-embeddings such as MUSE (Conneau et al., 2017), mBERT (Wu and Dredze, 2019) etc. for text-representation to ensure the cross-lingual transferring from the source to the target language. CLT has successfully been applied to numerous NLP-tasks including Dependency Parsing (Daniel et al., 2017; Zeman et al., 2018), Natural Language Inference (Conneau et al., 2018; Singh et al., 2019; Huang et al., 2019; Doval et al., 2019), Question Answering (Liu et al., 2019; Lee and Lee, 2019; Lewis et al., 2019), Text-classification (Bel et al., 2003; Shi et al., 2010; Mihalcea et al., 2007; Prettenhofer and Stein, 2010; Xu et al., 2016; Chen et al., 2018) etc. The key contribution of this small and focused work is that, as far as we are aware, it is the first paper which evaluates the performance of CLT on the *Constituency Parsing* task.

The key reason behind CLT not been applied to the Constituency-parsing task so far is the unavailability of universally annotated datasets in multiple languages. There are numerous constituency tree-banks available in a diverse range of languages. But unlike *Dependency Parsing* tree-banks which are mostly annotated with the *UD Annotations* (McDonald et al., 2013), in case of *Constituency Parsing* various existing tree-banks have their own independent tag annotations, thus making the application of multilingual approaches to it as impossible. However, (Han et al., 2014) proposed a *Universal Phrase tag-set* with 9 common Phrase-tags. Furthermore, (Han et al., 2014) also provides a mapping table to map tags of popular constituency tree-banks (including all treebanks used by us in our experiments) to these *Unversal Phrase Tags*.

We used this mapping table to replace all tags within all tree-banks utilized by us during experiments, with the universal tags. Subsequently we trained and evaluated all approaches (including baseline and proposed CLT based approaches) on these *Universally Tagged* tree-bank versions.

Action	Description
NT(X)	Opens a non-terminal node 'X' and puts it on top of *Stack*. eg: NT(VP)==>(VP
SHIFT	Removes topmost token from the *Buffer* B and pushes onto Stack
REDUCE	Repeatedly pops completed sub-trees or terminal symbols from the stack until an open non-terminal is encountered, and then this open NT is popped and used as the label of a new constituent that has the popped sub-trees as its children. This new completed constituent is pushed onto the stack as a single composite item.

Table 1: Action Set for *Discriminative RNNG* (Dyer et al., 2016)

2 Cross-lingual Discriminative RNNG

Discriminative RNNGs is a transition based constituency parser comprising of three key components namely *Stack* S which stores the incomplete parse-tree, *Buffer* B which stores the sentence tokens and the set of all possible actions A. At every time-step t, the algorithm chooses the best action $a_t \in A$, given the current state of stack S_t, buffer B_t and history of actions $a_{<t}$. Depending upon the chosen action a_t, the Stack and Buffer are updated accordingly. The process is continued until the Buffer becomes empty and Stack consists of completed parse-tree.

Table 1 describes the actions within action-set A for the *Discriminative RNNG (DiscRNNG)*. At any time-step t, RNNGs use a stack-LSTM (Dyer et al., 2015) to encode the current state of Stack S_t and use simple RNN to encode the current state of Buffer B_t and action-history $a_{<t}$. Given S_t, B_t and $a_{<t}$, the probability vector P_t comprising probabilities of all actions within A at time-step t is computed by applying equation 1.

$$P_t = softmax(r^T u_t + b) \quad (1)$$

Vector u_t is vector representing the entire model-state at time t. u_t is computed by applying equation 2.

$$u_t = tanh(W[S_t; B_t; a_{<t}] + c) \quad (2)$$

The Cross-lingual variant of this Discriminative RNNG parser evaluated by us, has same architecture as original model with two distinctions. **1.)** Multilingual BERT based Word-embeddings (Wu and Dredze, 2019) are used instead of monolingual Word-embeddings during the Buffer and Stack encodings (S_t and B_t), to ensure cross-lingual transfering. Such mBERT based embeddings are calculated in same way as in (Kondratyuk and Straka, 2019) (Appendix B describes the computation process in details). These multilingual embeddings are fixed during the training of the parser. **2.)** We fed-in the linguistic-typology features of the language being parsed, along with Stack, Buffer and Action-history encoding while predicting the best action at time t. Hence, the cross-lingual RNNG model predicts the probability vector P_t by applying equation 3 (instead of equation 1).

$$P_t = softmax(r^T [u_t; Z] + b) \quad (3)$$

Here $Z \in R^{|Z|}$ is a *Linguistic-typology* vector. Each value within Z represents a single typology-feature from *WALS* (Haspelmath, 2009) database having specific value as integer for the language being parsed. Missing features for any language is assigned *zero* indicating no dominant value for it. We refer to this model as *Cross-lingual RNNG parser with Linguistic Typology* (CL-RNNG-w-Typo) in this work.

Linguistic typology knowledge is successfully utilised for the cross-lingual dependency-parsing task by numerous researchers such as (Naseem et al., 2012; Täckström et al., 2013; Barzilay and Zhang, 2015; Wang and Eisner, 2016a; Rasooli and Collins, 2017; Ammar, 2016; Wang and Eisner, 2016b) to facilitate cross-lingual transfer. This inspired us to include linguistic typology knowledge for the cross-lingual Constituency-parsing task indeed.

3 Experiments

We conducted numerous experiments to evaluate the *CL-RNNG-w-Typo* model in both *Few-shot* (Wang et al., 2019) and *Zero-shot* (Socher et al., 2013) settings[1].

3.1 Baselines

We compared the performance of *CL-RNNG-w-Typo* parser with following baselines.

[1]Source Code, Mappings and Model-weights at www.github.com/XXXX

Language	Tree-bank	Family
English	Penn tree-bank (Marcus et al., 1993)	Germanic
Swedish (sd)	Talbanken05 (Nivre et al., 2006)	Germanic
French (fr)	FrenchTreebank (Abeillé et al., 2003)	Romance
Spanish (es)	Spanish UAM Treebank (Moreno et al., 1999)	Romance
Japanese (jp)	Tüba-J/S (Kawata and Bartels, 2000)	Altic
Arabic (ab)	Arabic PENN Treebank (Bies and Maamouri, 2003)	Afro-asiatic
Hungarian (hg)	Hungarian Szeged Treebank (Treebank)	Uralic

Table 2: List of source languages and their corpra used during experimentation.

Language	Tree-bank	Family
German (de)	Negra Treebank (Skut et al., 1997)	Germanic
Danish (da)	Arboretum Treebank (Bick, 2003)	Germanic
Italian (it)	ISST Treebank (Montemagni et al., 2003)	Romance
Catalan (ct)	Catalan AnCora Treebank (Taulé et al., 2008)	Romance
Korean (kr)	Korean Penn Treebank (Han et al., 2002)	Altic
Heberew (hb)	(Sima'an et al., 2001)	Afro-asiatic
Estonian (est)	Estonian Arborest Treebank (Bick et al.)	Uralic
Hindi (hi)*	Hindi-Urdu Treebank (Bhat et al., 2017)	Indo-aryan
Vietnamese (vt)*	Vietnamese Treebank (Nguyen et al., 2009)	Austroasiatic

Table 3: List of target languages and their corpra used during experimentation.

Model	de	da	it	ct	kr	hb	est	hi	vt
CL-RNNG-Mono	70.09	72.64	64.48	59.35	61.47	60.79	55.35	51.17	50.06
CL-RNNG-Poly	66.9	66.11	66.98	69.35	65.37	66.91	66.05	58.83	58.76
CL-RNNG-w-LangID	66.9	67.42	68.01	69.25	66.35	68.16	66.1	58.3	58.84
CL-RNNG-w-Typo	67.97	67.66	67.92	71.15	66.7	69.35	67.43	60.37	60.35

Table 4: F1 Score in *Few-shot* learning settings.

Model	de	da	it	ct	kr	hb	est	hi	vt
CPE-PLM	41.36	43.89	45.72	46.12	50.15	45.4	44.03	39.86	43.72
CL-RNNG-Mono	68.13	70.14	61.99	56.85	58.91	57.82	52.61	48.66	47.92
CL-RNNG-Poly	64.43	64.13	64.5	66.37	63.32	64.99	63.5	56.2	56.21
CL-RNNG-w-LangID	64.85	64.72	65.15	67.05	63.87	66.07	64.28	56.29	56.71
CL-RNNG-w-Typo	65.83	65.75	66.08	68.19	64.26	66.87	65.07	57.5	58.48

Table 5: F1 Score in *Zero-shot* learning settings.

1.) Chart-based CPE-PLM (Kim et al., 2020b): Its a state of the art neural unsupervised constituency parser which only utilises the syntactic knowledge encoded within a transformer based language-model such as BERT (Devlin et al., 2018), XLM-R (Conneau et al., 2019) etc. to construct a parse-tree. We re-implemented the model and used it as our baseline within *Zero-shot* settings.

2.) Cross-lingual RNNG Parser trained on single source language (CL-RNNG-Mono): Its the same model as *CL-RNNG-w-Typo* except that it does not use the linguistic-typology knowledge and is trained on a single source language English.

3.) Cross-lingual RNNG Parser trained of multiple source languages (CL-RNNG-Poly): It is the same model as *CL-RNNG-Mono*, but trained on a mixed polyglot corpus of high-resource source languages rather than a single source language English.

4.) Cross-lingual RNNG Parser with Language-id (CL-RNNG-w-LangID): It has same architecture as *CL-RNNG-w-Typo* model, with typology vector been replaced by the one-hot language-id vector representing the language being parsed.

3.2 Dataset

Tables 2 and 3 list all the *Source* and *Target* languages as well as their tree-bank corpra, the universally-tagged versions of which were used for the experimentation. Appendix A outlines the mapping-table used to replace the original annotations in these tree-banks to the universal-tag annotations (mapping provided by (Han et al., 2014)). We evaluated the CL-RNNG models on each of the target languages listed in Table 3 independently. For each experiment, the source-language training corpus size is always fixed to 700,000 tokens to ensure controlled experiment-settings.

We created the source-language training-corpus for *CL-RNNG-Mono* parsers by randomly sampling sentences from the English-PTB corpus (one at a time), until the token-size becomes approximately equal to 700,000. On the other hand, to create the source-language training-corpus for all *CL-RNNG-Poly* models, we randomly sampled sentences from each of the seven source-language corpra listed in table 2 until the token-size becomes approximately equal 100,000, concatenated all these sampled datasets and randomly shuffled the order, thus ensuring that all seven source-languages listed in table 2 are equally represented in the training-corpus.

Few-shot learning settings require a handful of training examples in the target language. We extracted this small target-language training-set by randomly sampling sentences from the train-set of copra listed in table 3 until the token-size becomes approximately equal to 3000. This is inspired by (Ammar et al., 2016) who used the same yardstick to evaluate their dependency parser.

3.3 Typology and Hyper-parameters

Appendix C will outline all the hyper-parameters used during the training. Typology vector Z includes feature-values of all word-order and constituency features in WALS (Haspelmath, 2009) database excluding trivially redundant features as excluded by (Takamura et al., 2016).

4 Results and Inference

Tables 4 and 5 outlined overall F1 scores obtained on all target-languages, within the *Few-shot Learning* and *Zero-shot* learning settings. Results in table 5 show that for *Zero-shot* settings, all CLT approaches significantly outperformed the CPE-PLM. It is inline with trends observed for other NLP tasks, where even a simple CLT based approach to the respective task always significantly outperforms the most complex unsupervised approaches.

In general, it is evident in Tables 4 and 5 that all models perform marginally better in Few-shot rather that in Zero-shot settings. In both the settings, for languages Danish (da) and German (de), the *Cl-RNNG-Mono* outperformed other polyglot models. The reason being that these languages belong to the same language-family as English namely *Germanic* and are indeed typologically very close to the source-languages of *Cl-RNNG-Mono* namely *en*. Whereas, it under-performed *CL-RNNG-Poly* on the other target languages namely *it*, *ct*, *est*, *hb* and *kr*. It is also evident that all model achieved a lower score on target-languages *hi* and *vt*, as compared to other target-languages. The reason being that these languages belong to families *Indo-aryan* and *Austro-asiatic* respectively and are typologically very distinct from all source languages listed in Table 2.

Based on these trends it can be inferred that the CLT based parsers perform better when the source and target languages are typologically closer. Furthermore, it can infer that the polyglot training training increases the Cross-lingual transferring ability of the CL-RNNG models to the unseen target-language (typologically distinct from its source languages) as it allows the model to better generalize over a diverse set of languages. Both of these trends are also observed for CLT based approaches to other NLP tasks as well.

Results also show that, the *CL-RNNG-w-Typo* outperformed *CL-RNNG-w-LangID* and *CL-RNNG-Poly* models for all the target-languages in both settings. Hence, it can be inferred that feeding the linguistic-typology knowledge does indeed improve cross-lingual tranferring ability of the parser.

5 Conclusion

This is the first work which evaluated a cross-lingual transfer learning approach to the Constituency-parsing task. We proved that CLT significantly outperforms Unsupervised approaches. Future work would involve extrinsic evaluation of CL constituency parsing on numerous downstream NLP tasks.

References

Anne Abeillé, Lionel Clément, and François Toussenel. 2003. Building a treebank for french. In *Treebanks*, pages 165–187. Springer.

Waleed Ammar. 2016. *Towards a Universal Analyzer of Natural Languages*. Ph.D. thesis, Ph. D. thesis, Google Research.

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.

Regina Barzilay and Yuan Zhang. 2015. Hierarchical low-rank tensors for multilingual transfer parsing. Association for Computational Linguistics.

Nuria Bel, Cornelis HA Koster, and Marta Villegas. 2003. Cross-lingual text categorization. In *International Conference on Theory and Practice of Digital Libraries*, pages 126–139. Springer.

Riyaz Ahmad Bhat, Rajesh Bhatt, Annahita Farudi, Prescott Klassen, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Ashwini Vaidya, Sri Ramagurumurthy Vishnu, et al. 2017. The hindi/urdu treebank project. In *Handbook of Linguistic Annotation*, pages 659–697. Springer.

Eckhard Bick. 2003. Arboretum, a hybrid treebank for danish. In *Proceedings of TLT 2003 (2nd Workshop on Treebanks and Linguistic Theory, Växjö*, pages 9–20.

Eckhard Bick, Heli Uibo, and Kadri Muischnek. Preliminary experiments for a cg-based syntactic tree corpus of estonian.

Ann Bies and Mohamed Maamouri. 2003. Penn arabic treebank guidelines. *Draft: January*, 28:2003.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *1st Meeting of the North American Chapter of the Association for Computational Linguistics*.

Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. 2018. Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6:557–570.

Michael Collins. 2003. Head-driven statistical models for natural language parsing. *Computational linguistics*, 29(4):589–637.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

Alexis Conneau, Guillaume Lample, Ruty Rinott, Adina Williams, Samuel R Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. Xnli: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053*.

Zeman Daniel, Popel Martin, Straka Milan, Hajic Jan, Nivre Joakim, Ginter Filip, Luotolahti Juhani, Pyysalo Sampo, Petrov Slav, Potthast Martin, et al. 2017. Conll 2017 shared task: Multilingual parsing from raw text to universal dependencies. In *CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, volume 1, pages 1–19. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yerai Doval, Jose Camacho-Collados, Luis Espinosa Anke, and Steven Schockaert. 2019. Meemi: A simple method for post-processing cross-lingual word embeddings. *arXiv preprint arXiv:1910.07221*.

Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah A Smith. 2015. Transition-based dependency parsing with stack long short-term memory. *arXiv preprint arXiv:1505.08075*.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A Smith. 2016. Recurrent neural network grammars. *arXiv preprint arXiv:1602.07776*.

Aaron Li-Feng Han, Derek F Wong, Lidia S Chao, Yi Lu, Liangye He, and Liang Tian. 2014. A universal phrase tagset for multilingual treebanks. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pages 247–258. Springer.

Chunghye Han, Narae Han, Eonsuk Ko, and Martha Palmer. 2002. Korean treebank: Development and evaluation. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*.

Martin Haspelmath. 2009. *The typological database of the World Atlas of Language Structures*. Berlin: Walter de Gruyter.

Haoyang Huang, Yaobo Liang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, and Ming Zhou. 2019. Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks. *arXiv preprint arXiv:1909.00964*.

Katharina Kann, Anhad Mohananey, Samuel Bowman, and Kyunghyun Cho. 2019. Neural unsupervised parsing beyond english. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, pages 209–218.

Yasuhiro Kawata and Julia Bartels. 2000. Stylebook for the japanese treebank in verbmobil. In *Verbmobil-Report 240, Seminar für Sprachwissenschaft, Universität Tübingen*.

Taeuk Kim, Jihun Choi, Daniel Edmiston, and Sang-goo Lee. 2020a. Are pre-trained language models aware of phrases? simple but strong baselines for grammar induction. *arXiv preprint arXiv:2002.00737*.

Taeuk Kim, Bowen Li, and Sang-goo Lee. 2020b. Multilingual chart-based constituency parse extraction from pre-trained language models. *arXiv preprint arXiv:2004.13805*.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing universal dependencies universally. *arXiv preprint arXiv:1904.02099*.

Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A Smith. 2016. What do recurrent neural network grammars learn about syntax? *arXiv preprint arXiv:1611.05774*.

Chia-Hsuan Lee and Hung-Yi Lee. 2019. Cross-lingual transfer learning for question answering. *arXiv preprint arXiv:1907.06042*.

Patrick Lewis, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2019. Mlqa: Evaluating cross-lingual extractive question answering. *arXiv preprint arXiv:1910.07475*.

Jiahua Liu, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2019. Xqa: A cross-lingual open-domain question answering dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2358–2368.

Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, et al. 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.

Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 976–983.

Simonetta Montemagni, Francesco Barsotti, Marco Battista, Nicoletta Calzolari, Ornella Corazzari, Antonio Zampolli, Francesca Fanciulli, Maria Massetani, Remo Raffaelli, Roberto Basili, et al. 2003. The italian syntactic-semantic treebank: Architecture, annotation, tools and evaluation.

Antonio Moreno, Susana López, and Manuel Alcántara. 1999. Spanish tree bank: Specifications, version 5. *Technical paper*.

Khalil Mrini, Franck Dernoncourt, Quan Tran, Trung Bui, Walter Chang, and Ndapa Nakashole. 2019. Rethinking self-attention: Towards interpretability in neural parsing. *arXiv preprint arXiv:1911.03875*.

Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing. The Association for Computational Linguistics.

Phuong Thai Nguyen, Xuan Luong Vu, Thi Minh Huyen Nguyen, Hong Phuong Le, et al. 2009. Building a large syntactically-annotated corpus of vietnamese.

Joakim Nivre, Jens Nilsson, and Johan Hall. 2006. Talbanken05: A swedish treebank with phrase structure and dependency annotation. In *LREC*, pages 1392–1395.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440.

Peter Prettenhofer and Benno Stein. 2010. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1118–1127.

Mohammad Sadegh Rasooli and Michael Collins. 2017. Cross-lingual syntactic transfer with limited resources. *Transactions of the Association for Computational Linguistics*, 5:279–293.

Lei Shi, Rada Mihalcea, and Mingjun Tian. 2010. Cross language text classification by model translation and semi-supervised learning. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1057–1067.

Khalil Sima'an, Alon Itai, Yoad Winter, Alon Altman, and Noa Nativ. 2001. Building a tree-bank of modern hebrew text. *Traitement Automatique des Langues*, 42(2):247–380.

Jasdeep Singh, Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2019. Xlda: Cross-lingual data augmentation for natural language inference and question answering. *arXiv preprint arXiv:1905.11471*.

Wojciech Skut, Brigitte Krenn, Thorsten Brants, and Hans Uszkoreit. 1997. An annotation scheme for free word order languages. *arXiv preprint cmp-lg/9702004*.

Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. 2013. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems*, pages 935–943.

Oscar Täckström, Ryan McDonald, and Joakim Nivre. 2013. Target language adaptation of discriminative transfer parsers.

Hiroya Takamura, Ryo Nagata, and Yoshifumi Kawasaki. 2016. Discriminative analysis of linguistic features for typological study. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 69–76.

Sho Takase, Jun Suzuki, and Masaaki Nagata. 2018. Direct output connection for a high-rank language model. *arXiv preprint arXiv:1808.10143*.

Mariona Taulé, Maria Antònia Martí, and Marta Recasens. 2008. Ancora: Multilevel annotated corpora for catalan and spanish. In *Lrec*.

Hungarian Szeged Treebank. Szeged treebank 2.0: A hungarian natural language database with detailed syntactic analysis. *Hungarian linguistics at the University of Szeged*.

Dingquan Wang and Jason Eisner. 2016a. 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. 2016b. The galactic dependencies treebanks: Getting more data by synthesizing new languages. *Transactions of the Association for Computational Linguistics*, 4:491–505.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2019. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of bert. *arXiv preprint arXiv:1904.09077*.

Zhiyong Wu, Yun Chen, Ben Kao, and Qun Liu. 2020. Perturbed masking: Parameter-free probing for analyzing and interpreting bert. *arXiv preprint arXiv:2004.14786*.

Ruochen Xu, Yiming Yang, Hanxiao Liu, and Andrew Hsi. 2016. Cross-lingual text classification via model translation with limited dictionaries. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 95–104.

Kaiyu Yang and Jia Deng. 2020. Strongly incremental constituency parsing with graph neural networks. *arXiv preprint arXiv:2010.14568*.

Daniel Zeman, Jan Hajic, 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.

Yanpeng Zhao and Ivan Titov. 2021. An empirical study of compound pcfgs. *arXiv preprint arXiv:2103.02298*.

A Mappings

Tables 7 and 8 outlines the *Universal Phrase-tag set* (in the first column), as well as the mappings from the distinct tag-annotations of all the training and test copra tree-banks used by us during experimentation, to these Universal Phrase-tags. These mappings are provided by (Han et al., 2014).

Before each experiment, we used these mapping tables to replace all tags within the train/test tree-banks with the universal tags. Subsequently we trained and evaluated all the baseline and proposed CLT based approaches on these *Universally Tagged* tree-bank versions.

B Word-embeddings

We used the Multilingual BERT based Word-embeddings instead of monolingual Word-embeddings during the Buffer and Stack encodings, to ensure the cross-lingual transferring between source and target languages. For any sentence S being parsed, we computed the embeddings for all the words in S simultaneous.

We inputted the entire sentence S to the BERT's WordPiece Tokenizer to obtain the corresponding token-sequence. Subsequently we fed-in this obtained token-sequence into a pre-trained mBERT model. For any word $w \in S$ we used the outputs of the pre-trained mBERT corresponding to the first wordpiece token of it to compute of its embedding e_w, ignoring the rest of the token. The embedding vector e_w is computed by simply summing-up the outputs of all the layers of the pre-trained BERT model (equation 4).

$$e_w = \sum_j BERT_j \tag{4}$$

These embeddings are then utilised to encode the Stack and Buffer during the parsing. Hence the word-embeddings are distinct for each input sentence, but are not fine-tuned with the parser training.

C Hyper-parameters

Table 6 outlines hyper-permeters used during experiments. These values are obtained by minimizing the training loss on *Development* dataset (Dev set) for *Penn Treebank Corpus* (Marcus et al., 1993).

Typology vector Z includes feature-values of all word-order and constituency features in WALS (Haspelmath, 2009) database excluding trivially redundant features as excluded by (Takamura et al., 2016).

For each experiment, every model is trained and evaluated five times and the averaged value of results are reported in Tables 4 and 5. The models are implemented in Tensorflow. We used the BERT model bert_multi_cased_L-12_H-768_A-12 provided by `huggingface.com`

Hyper-parameter	**Value**
WE dims	768
S_t,B_t,$a_{<t}$ dims	450
$u^{\beta}{}_t$, $u^{\alpha}{}_t$ dims	450
Dropout prob.	0.01
Bach-size	32
Epochs	150
BERT Model	bert_multi_cased_L-12_H-768_A-12
Learning rate	0.05
Exponential decay	True

Table 6: Hyper-parameters

Universal Phrase Tag	**UPenn**	**Talbanken 05**	**French-Treebank**	**Spanish UAM**	**Tuba-J/S**	**Arabic PENN**	**Hungarian Szeged**
NP	NP, WHNP	CNP, NP	NP	HOUR, NP, QP, SCORE, TITLE	NPper, NPloc, NPtmp, NP, NP.foc	NP, NX, QP, WHNP	NP, QP
VP	VP	CVP, VP	VN, VP, VPpart, VPinf	VP	VP.foc, VP, VPcnd, VPfin	VP	VP, INF, INF0
AJP	ADJP	AP, CAP	AP	ADJP	AP.foc, AP, APcnd	ADJP, WHADJP	ADJP
AVP	ADVP, WHADVP	AVP, CAVP	AdP	ADVP, PRED-COMPL	ADVP.foc, ADVP	ADVP, WHADVP	ADVP, PA, PA0
PP	PP, WHPP	CPP, PP	PP	PP	PP, PP.foc, PPnom, PPgen, PPacc	PP, WHPP	PP
S	S, SBAR, SBARQ, SINV, SQ	CS, S	SENT, Ssub, Sint, Srel, S	CL, S	S, SS	S, SBAR, SBARQ, SQ	S
CONJP						CONJP, NAC	C0
COP		CONJP, CXP	COORD				CP
X	X	NAC, XP			ITJ, GR, err	PRN, PRT, FRAG, INTJ, X, UCP	FP, XP

Table 7: Mappings of Source Treebank copra annotations to Universal Phrase tags

Universal Phrase Tag	**Negra Tree-bank**	**Arboretum Tree-bank**	**ISST Tree-bank**	**Catalan AnCora Tree-bank**	**Korean Penn**	**Heberew Tree-bank**	**Estonian Arborest**	**Hindi-Urdu Tree-bank**	**Viet-namese Tree-bank**
NP	NP, CNP, MPN, NM	Np	SN	sn	NP	NP-gn-(H)	AN, NN	NP, NP-P, NPNST, SC-A, SC-P, NP-P-Pred	NP, WHNP, QP
VP	VP, CVP, VZ, CVZ	vp, acl	IBAR	gv	VP	PREDP, VP, VP-MD, VPINF	VN, INF-N	VP, VP-Pred, V	VP
AJP	AP, AA, CAP, MTA	Ajp	SA	sa	ADJP, DANP	ADJP-gn-(H)		AP, AP-Pred	AP, WHAP
AVP	AVP, CAVP	Dvp	SAVV	sadv, neg	ADVP, ADCP	ADVP	AD	DegP	RP, WHRP
PP	PP, CAC, CPP, CCP	pp	SP, SPD, SPDA	sp		PP			PP, WHPP
S	S, CS, CH, DL, PSEUDO	fcl, icl	F, SV2, SV3, SV5, FAC, FS, FINT, F2	S, S*, S.NF.C, S.NF.A, S.NF.P, S.F.C, S.F.AComp, S.F.AConc, S.F.Acons, S.F.Acond, S.F.R,	S	FRAG, FRAGQ, S, SBAR, SQ			S, SQ, SBAR
CONJP		cp	CP, COMPC	conj.subord, coord					
COP	CO		FC, CO-ORD				PN	CCP, XP-CC	
X	ISU, QL	par	FP, COMPT, COMPIN	interjeccio, mor-fema.verbal, morf.pron	INTJ, PRN, X, LST, XP	INTJ, PRN	P, Q	CP	XP, YP, MDP

Table 8: Mappings of Target Treebank copra annotations to Universal Phrase tags