Is Typology-Based Adaptation Effective for Multilingual Sequence Labelling?

Ahmet Üstün **Arianna Bisazza** **Gosse Bouma** **Gertjan van Noord**

University of Groningen

{a.ustun, a.bisazza, g.bouma, g.j.m.van.noord}@rug.nl

Abstract

Recent work has shown that a single multilingual model with typologically informed parameter sharing can improve the performance in dependency parsing on both high-resource and zero-shot conditions. In this work, we investigate whether such improvements are also observed in the POS, NER and morphological tagging tasks.

1 Introduction

Recent studies have shown that state-of-the-art neural models can successfully be trained on multiple languages simultaneously for various NLP tasks such as named-entity recognition (Mulcaire et al., 2019; Rahimi et al., 2019), morphosyntactic tagging (Tsai et al., 2019), dependency parsing (Ammar et al., 2016; Kondratyuk and Straka, 2019), and machine translation (Johnson et al., 2017; Aharoni et al., 2019). Multilingual approaches have the following benefits over monolingual counterparts: (1) Multilingual models learn better generalization by sharing model parameters among multiple source languages which enables cross-lingual transfer and zero-shot learning. (2) Maintaining a single multilingual model is easier than a large set of language-specific models which increase the total size and system complexity.

However, a multilingual model trained on a high number of languages can face the "transfer - interference trade-off" (Johnson et al., 2017; Arivazhagan et al., 2019; Conneau et al., 2020). This trade-off leads multilingual models to outperform monolingual baselines on low/zero-resource languages (*positive transfer*), but to underperform on high-resource languages due to the lack of language specific capacity (*negative interference*). Moreover, multilingual transfer could give mixed results when the model is trained on a diverse set of source languages in terms of script, phonology, morphology, syntax, and semantics.

Early work on multilingual learning has been applied in POS and morphological tagging (Gillick et al., 2016; Tsai et al., 2019), NER (Mulcaire et al., 2019; Rahimi et al., 2019), and UD parsing (Ammar et al., 2016; Kondratyuk and Straka, 2019; Üstün et al., 2020). Gillick et al. (2016) showed that a compact multilingual model operating on bytes could reach similar or better performance in POS and NER. Mulcaire et al. (2019) and Wu and Dredze (2019) showed that multilingual language model pretraining based on ELMo (Peters et al., 2018) and BERT (Devlin et al., 2019), improves the performance on POS, NER and UD tasks, including zero-shot settings. More recently, Kondratyuk and Straka (2019) trained a massively multilingual UD parser, UDify, on 75 languages using multilingual BERT (mBERT), which shows competitive or improved performance, but mostly on low-resource languages.

Finally, Üstün et al. (2020) have proposed a multilingual dependency parser, UDapter, with language-specific and typologically informed parameter sharing. This parser learns task-specific adapters (Houlsby et al., 2019), which are inserted into mBERT, via language embeddings by using contextual parameter generation (Platanios et al., 2018) to balance maximum sharing and language-specific capacity in a single multilingual model. Furthermore, language embeddings are defined as a function of a large set of linguistically curated and predicted typological features obtained from the URIEL (Littell et al., 2017) database. While typology features had been used before (Ammar et al., 2016), this work was the first to report improvements on both high-resource and zero-shot results, while not using gold POS tags.

In this work, we extend UDapter to three other sequence labelling tasks, namely: POS tagging, NER and morphological (MORPH) tagging. We

	ar	eu	fi	he	hi	ko	ru	sv	tr	zh	be*	fo*	kk*	olo*	sa*	yue*	HR-AVG	LR-AVG
POS tagging (Accuracy)																		
UDify	96.3	94.5	96.4	96.5	96.7	95.3	98.5	97.6	93.4	93.8	94.6	80.0	85.0	74.5	41.4	68.1	95.9	58.0
UDapter	96.8	95.7	97.3	97.1	97.4	96.5	98.9	98.4	95.1	95.2	96.9	79.6	83.4	76.6	42.2	66.3	96.8	58.4
Morphological Tagging (F1)																		
UDify	95.7	92.0	94.7	94.9	97.0	99.6	97.6	97.6	90.9	98.9	92.3	58.4	52.2	59.9	40.7	-	95.9	44.4
UDapter	96.8	95.4	96.7	96.2	97.7	99.8	98.3	98.3	94.8	99.2	92.0	56.3	52.5	62.4	20.6	-	97.3	44.3
Named Entity Recognition (F1)																		
UDify	87.9	91.3	90.9	84.7	86.8	87.3	88.6	94.4	91.9	78.8	80.2	78.3	72.4	-	53.8	76.4	88.3	67.4
UDapter	89.4	92.7	92.4	86.7	89.5	88.8	89.8	95.8	93.1	81.3	80.4	81.0	65.4	-	45.6	75.2	89.9	65.6
Dependency Parsing (LAS)																		
UDify	80.12	76.4	85.1	84.4	89.3	78.0	89.0	86.2	62.7	77.8	80.1	68.6	61.9	42.1	19.4	30.5	83.0	50.4
UDapter	84.42	83.3	89.0	88.8	92.0	85.9	92.2	90.3	69.6	83.2	79.3	69.2	60.7	43.3	22.2	32.8	87.3	51.3

Table 1: Results of UDapter (Üstün et al., 2020) and UDify (Kondratyuk and Straka, 2019) trained per task on 13 languages. '*' indicates the zero-shot experiments. HR-AVG is calculated over all 13 training languages, and LR-AVG is calculated over all low-resource languages used in the evaluations.

present results on a large set of languages, including the zero-shot setting, and compare them to the parsing results presented in (Üstün et al., 2020).

2 Experiments

We modify UDapter (Üstün et al., 2020), by adding simple task-specific softmax layers for POS, MORPH tagging, and NER. Parameters of task-specific layers are modified by typologically informed language embeddings as in the original model. As a baseline, we use UDify (Kondratyuk and Straka, 2019) trained on our set of 13 languages (see below). Both UDapter and UDify are trained separately for each of the three tasks with the same hyper-parameter setting as in Üstün et al. (2020) for 30 epochs.

For training languages, following Kulmizev et al. (2019), we choose Arabic (ar), English (en), Basque (eu), Finnish (fi), Hebrew (he), Hindi (hi), Italian (it), Japanese (ja), Korean (ko), Russian (ru), Swedish (sv), Turkish (tr) and Chinese (zh). In the zero-shot setup, both models are evaluated on languages used in Üstün et al. (2020) if annotations exist in the datasets. Table 1 shows results on a subset of languages including Belarusian (be), Faroese (fo), Kazakh (kk), Livvi (olo), Sanskrit (sa) and Cantonese (yue). We selected this subset so that each language is from the same family as one of the high-resource languages. For POS and MORPH tagging, we use UD 2.3 (Nivre et al., 2018) and for NER, we use WikiANN (Pan et al., 2017) which was divided into train/dev/test by Rahimi et al. (2019).

3 Results and Conclusion

Results including dependency parsing scores from Üstün et al. (2020) are given in Table 1. Language-specific adaptation of a multilingual model with typological features increases the performance in all tasks for high-resource (HR) languages. The increase in MORPH tagging and parsing is larger than for POS and NE.

However, for low-resource (LR) languages, we find a mixed picture. In POS tagging and parsing, UDapter outperforms UDify but the improvements are rather small compared to HR. In MORPH tagging, we observed a very small decrease on average but a large variation, mostly due to the inconsistent annotations in LR languages. For UDapter, jointly predicting separate morphological attributes besides the unfactored tag string (Inoue et al., 2017), helps the model to learn infrequent tags, but overall performance is still below the baseline. Finally, in NER, unlike the other three tasks, UDapter has significantly lower performance. A possible explanation is that while POS tagging and parsing have more structured and universal representation which allows the model to exploit multilingual sharing even for unseen words, NER requires better word-level representations where typological features may not have benefits. Moreover, the low representation quality of LR languages in mBERT (Wu and Dredze, 2020) may have a stronger negative effect on these tasks.

This is work in progress. We are currently studying the reasons of the different trends and we plan to provide further analysis during the presentation.

References

Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively multilingual neural machine translation. 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 3874–3884.

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.

Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George Foster, Colin Cherry, Wolfgang Macherey, Zhifeng Chen, and Yonghui Wu. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges. *CoRR*, abs/1907.05019.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

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.

Dan Gillick, Cliff Brunk, Oriol Vinyals, and Amarnag Subramanya. 2016. Multilingual language processing from bytes. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1296–1306.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799.

Go Inoue, Hiroyuki Shindo, and Yuji Matsumoto. 2017. Joint prediction of morphosyntactic categories for fine-grained Arabic part-of-speech tagging exploiting tag dictionary information. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 421–431.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

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.

Artur Kulmizev, Miryam de Lhoneux, Johannes Gontrum, Elena Fano, and Joakim Nivre. 2019. Deep contextualized word embeddings in transition-based and graph-based dependency parsing - a tale of two parsers revisited. 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 2755–2768.

Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 8–14.

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.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Rogier Blokland, Victoria Bobicev, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Katarína

Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta Gonzáles 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, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Kamil Kopacewicz, Natalia Kotsyba, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, 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 Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguy˜ên Thị, Huy`ên Nguy˜ên Thị Minh, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédayọ Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Slav Petrov, Jussi Piitulainen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roșca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Zsolt Szántó, Dima Taji, Yuta Takahashi, Takaaki Tanaka, Isabelle 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, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Seyi Williams, Mats Wirén, Tsegay Woldemariam, Tak-sum Wong, Chunxiao Yan, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, Manying Zhang, and Hanzhi Zhu. 2018. Universal dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1946–1958.

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.

Emmanouil Antonios Platanios, Mrinmaya Sachan, Graham Neubig, and Tom Mitchell. 2018. Contextual parameter generation for universal neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 425–435.

Afshin Rahimi, Yuan Li, and Trevor Cohn. 2019. Massively multilingual transfer for NER. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 151–164.

Henry Tsai, Jason Riesa, Melvin Johnson, Naveen Arivazhagan, Xin Li, and Amelia Archer. 2019. Small and practical BERT models for sequence labeling. 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 3632–3636.

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. 2020. Udapter: Language adaptation for truly universal dependency parsing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. 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 833–844.

Shijie Wu and Mark Dredze. 2020. Are all languages created equal in multilingual BERT? In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 120–130.