Cross-lingual Inflection as a Data Augmentation Method for Parsing

Alberto Muñoz-Ortiz, Carlos Gómez-Rodríguez and David Vilares
Universidade da Coruña, CITIC
Departamento de Ciencias de la Computación y Tecnologías de la Información
Campus de Elviña s/n, 15071
A Coruña, Spain
{alberto.munoz.ortiz, carlos.gomez, david.vilares}@udc.es

Abstract

We propose a morphology-based method for low-resource (LR) dependency parsing. We train a morphological inflector for target LR languages, and apply it to related rich-resource (RR) treebanks to create cross-lingual (x-inflected) treebanks that resemble the target LR language. We use such inflected treebanks to train parsers in zero- (training on x-inflected treebanks) and few-shot (training on x-inflected and target language treebanks) setups. The results show that the method sometimes improves the baselines, but not consistently.

1 Introduction

Dependency parsers (Dozat et al., 2017; Ma et al., 2018; Strzyz et al., 2019) already achieve accurate results for certain setups (Berzak et al., 2016). Yet, they require large amounts of data to work, which hurts low-resource (LR) scenarios. In this line, authors have studied how to overcome this problem.

On data augmentation, recent approaches have replaced subtrees of sentences to generate new ones (Vania et al., 2019; Dehouck and Gómez-Rodríguez, 2020). On cross-lingual learning, authors have explored delexicalized approaches from rich-resource (RR) treebanks. (McDonald et al., 2011; Falenska and Çetinoğlu, 2017). Wang and Eisner (2018) permuted constituents of distant treebanks to generate synthetic ones that resembled the target language. Vilares et al. (2016); Ammar et al. (2016) merged treebanks to train multilingual parsers that sometimes could outperform the equivalent monolingual version, which has applications for less-resourced parsing. In the context of multilingual representations, Mulcaire et al. (2019) trained a zero-shot parser on top of a polyglot language model, relying on merged RR treebanks too.

In other matters, morphological inflection (Cotterell et al., 2016; Pimentel et al., 2021) generates words from lemmas and morphological feats (e.g.
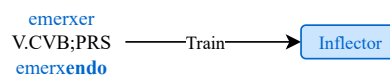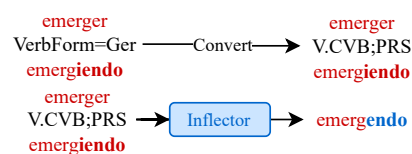
Figure 1: X-inflection process for a target LR language (Galician) using a source RR treebank (Spanish).

look → looking). Also, it is known that morphology helps parsing and that morphological complexity relates to the magnitude of the improvements (Dehouck and Denis, 2018). Yet, as far as we know, there is no work on cross-lingual morphological inflection as a data augmentation method for parsing. Here, we propose a technique that lies in the intersection between data augmentation, cross-lingual learning, and morphological inflection.

Contribution We introduce a method that uses cross-lingual morphological inflection to generate 'synthetic creole' treebanks, which we call *x-inflected* treebanks. To do so, we require a *source* language treebank from a closely-related language (for which lemmas and morphological feats are available), and a morphological inflection system trained for the *target* language. This way, we expect to generate x-inflected treebanks that should resemble *to a certain extent* the target language (see Figure 1). The goal is to improve the parser's performance for languages for which little or no annotated data are available, but for which we can train an accurate morphological inflection system that can be later applied to a related RR treebank and resemble the target language. The

code is available at `https://github.com/amunozo/x-inflection`.

2   Preliminaries

We now describe the basics of our work:

Datasets   We use UniMorph (UM; McCarthy et al., 2020) for morphology and Universal Dependencies (UD; Zeman et al., 2020)

Key concepts   We call *inflector* a morphological system that produces a word form from an input lemma and a set of morphological feats in a given language. We call *target UD treebank* each of the LR treebanks where we test our approach. We call *source UD treebanks* the RR treebanks related to a target LR treebank, used to create a cross-lingual inflected treebank, aka *x-inflected treebank*, which results from applying an inflector over the lemmas and feats of a source UD treebank.

3   X-inflection as data augmentation

Character-level models, such as the ones used for morphological inflection, identify shared morphemes across languages with overlapping alphabets (Lee et al., 2017; Vania, 2020). Thus, if two languages share a significant amount of lemmas, n-grams or inflections, an inflector for the first language could maybe produce noisy-but-useful inflected forms for lemmas and feats available for the second language. We hypothesize that this idea can be used for syntactic data augmentation in LR scenarios. Under the assumption that an inflector is available for our target LR language (easier than annotating syntactic data), we could use it to transform a related RR treebank, obtaining silver syntactic data that, despite lexical and grammatical imperfections, could help boost performance.

Our method consists of three steps: (i, §3.1) training an inflector for a given target language using UM data, (ii, §3.2) *x-inflecting* the source UD treebank, cross-lingually applying the inflector trained in (i), and (iii, §3.3) training the *x-inflected* parsers. We summarized the process in Figure 1.

3.1   Building the inflectors

We train the inflectors using the Wu et al. (2018) model, and leave all the hyperparameters at their default value. It is a seq2seq model that uses a hard monotonic attention mechanism to identify what parts of the input the model should focus on to generate the correct output string. It offers a good trade-off between speed and accuracy, compared with other alternatives that we tested in early experiments (Wu et al., 2021). We train the models on UM data, and for each language, we shuffle and split it 80-10-10 for the training, development and test sets (so lemmas are distributed)[1].

3.2   Building the *x-inflected* treebanks

This step requires to: (i) transform the feats column of the source UD treebank into a readable format by the inflector (i.e., UM format), to then (ii) apply the inflector to generate the x-inflected word forms, and (iii) format the output into an x-inflected treebank (i.e., going back to the UD format).

Transform UD feats into UM feats   To x-inflect the source treebank, we first need to convert the morphological feats of the UD treebanks to the UM schema, using the converter by McCarthy et al. (2018).

In early experiments, we also trained inflectors directly on UD feats (following §3.1), but the results showed that x-inflected parsers trained this way performed worse, so we discarded it.

More specifically, the selected converter creates a mapping between both schemata. Yet, annotation errors and missing values in both schemata, together with disagreements between them, makes the process non-trivial. To counteract this, the approach introduces a language-dependent post-editing process, which consists in an iterative process that analyzes those forms and lemmas present both in UD and UM, comparing their annotations, and creating rules to refine the mappings between schemata. However, this extra refinement process is only available for some languages.

X-inflecting treebanks   The lemmas and UM-transformed feats of the source UD treebank are sent to the target LR language inflector. The x-inflection is not applied to all elements, only to those lemmas of the source UD treebank whose part-of-speech is contained in the UM data of our target language (e.g., verbs or nouns, see details in Appendix A). Then, these x-inflected forms replace the original forms in the source UD treebank.

3.3   Training the *x-inflected* parsers

We train the parsers with a graph-based (GB) model (Dozat et al., 2017). It contextualizes words with

---

[1]For languages containing files for different dialects (e.g. Livvi), we concatenated all the forms prior to splitting.

Group	LR	ISO	RR
Iberian	Galician	glg	Spanish, Catalan, Portuguese
N. Germanic	Faroese	fao	Norwegian (nb), Norwegian (nn), Swedish, Icelandic, Danish
Finno-Ugric	Hungarian	hun	Finnish, Estonian
West Slavic	Czech	ces	Polish, Slovak
South Slavic	Slovenian	slv	Bulgarian, Croatian, Serbian
Romance	Latin	lat	Spanish, Romanian, French, Catalan, Italian
Baltic	Lithuanian	lit	Latvian
Celtic	Welsh	cym	Irish, Scottish Gaelic
Finnic	Livvi	olo	Finnish, Estonian
Finno-Permic	North Sami	sme	Finnish, Estonian

Table 1: LR and RR languages used in our experiments. Some LR treebanks come from RR languages (Czech, Latin) to have more samples.
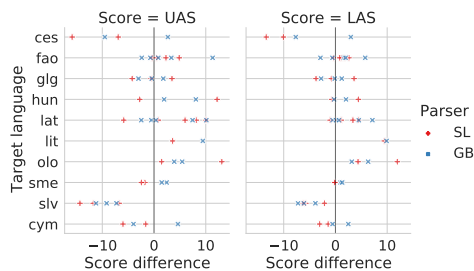

Figure 2: Score differences between x-inflected versions using different source treebanks and the baseline, for both sequence labeling (+) and graph-based (x) parsers.
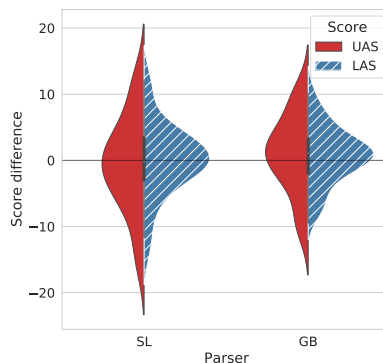

Figure 3: ΔUAS/LAS between the x-inflected models and the baseline for the sequence labeling (SL) and the graph-based (GB) parser.

bidirectional LSTMs (Hochreiter and Schmidhuber, 1997) and computes head and dependent representations for each word. Then, a biaffine transformation of such vectors is used to find the highest scoring parse tree. We also study a sequence labeling (SL) parser (Strzyz et al., 2019) as a lower bound. This parser can be seen as a vanilla biLSTM that only needs softmaxes (instead of a biaffine attention module) to predict syntactic labels, using 2-planar encodings (Strzyz et al., 2020), that are naturally decoded into to a dependency tree and work more robustly on low-resource setups (Muñoz-Ortiz et al., 2021).

4 Experiments

We test both (i, §4.1) zero-shot and (ii, §4.2) few-shot setups. For evaluation, we use unlabeled (UAS) and labeled attachment scores (LAS). Appendix E reports the hardware and costs.

Data We use 10 LR and 21 RR treebanks. Although our method can be applied to any pair of treebanks, the availability of UM and UD resources (in the sense of having LR languages in UM and related RR languages in UD) restricts our empirical analysis to Indo-European and Uralic languages (see Table 1). Yet, we have reasonable diversity and degrees of morphological inflection. For our empirical analysis, we use a relaxed definition of the concept LR for Czech and Latin (as the treebanks used are LR but there are RR treebanks for them in UD), and of the concept RR for Scottish Gaelic (as the treebank used is larger than the Welsh one but not RR). See Appendix B for the details.

4.1 Experiment 1: Zero-shot setup

We test if our method improves parsing accuracy under the assumption that there is no available training data in the target language, but there is an UD treebank for a related language, and enough UM

data to train an inflector for the target language. Although the selected LR treebanks have a training set, we here do not use them, but we will in §4.2.

Setup For each target LR treebank, we first pair them with related source UD treebanks (from 1 to 5)[2], such that they all belong to the most restricted phylogenetic group for which UD data is available. We then train our x-inflected parsers and evaluate them on the corresponding target LR treebank. We compare the results against a baseline consisting in models trained on the source RR treebanks.

Results Figure 2 shows the results for the zero-shot setup (full results in Appendix C). The differences in performance are inconsistent: for some target LR treebanks the x-inflected models always obtain improvements, e.g. Livvi, for some others the models only obtain decreases, e.g. Slovenian, and for some others there is a mix, e.g. Faroese.

Figure 3 shows the distributions of LAS and UAS differences (Δ) against the baseline versions. For the SL parser, the distribution is centered in 0,

---

[2]Depending on the resource availability in UM and UD.

with the occurrence of some extreme results. For the GB parser, we see less extreme results and a distribution centered slightly above 0. This suggests that our method could be more effective for the GB approach, but we do not have clear evidence.

To shed light on what factors might affect the results, Table 2 shows the Pearson correlation coefficient (PCC) of the LAS and UAS differences between the x-inflected models and the baselines; *with respect to* features such as the number of forms and lemmas seen in UM training data, feature and lemma overlap between the target and source UD treebanks, or the number of UD training sentences. Although small, the results show some correlations e.g. for the number of forms and lemmas of the UM data ($0.3 - 0.5$).

Figure 4: ΔLAS between the models trained on the original and x-inflected groups *with respect to* the model trained on the LR treebank, for both parsers. UAS shows the same tendency as seen in Appendix D.

| | GB | | SL | |
	UAS	LAS	UAS	LAS
# UM target language forms	0.31	0.34	**0.49**	**0.47**
# UM target language lemmas	0.32	0.32	**0.50**	**0.42**
# UD source treebank training sents.	0.30	0.17	0.27	0.22
% Morph. feats shared between treebanks	-0.24	-0.24	-0.14	-0.06
% Lemmas shared between treebanks	-0.35	-0.32	-0.20	-0.18

Table 2: PCC of ΔLAS/UAS between the x-inflected models and the baselines vs different dimensions. Bold numbers represents p-values < 0.05.

4.2 Experiment 2: Few-shot setup

Experiment 1 did not show consistent improvements. However, we question whether the reason for our x-inflected models not consistently improving over the baseline could be that having some annotated data in the target language would help better guide the learning process, or that we are simply not taking advantage of x-inflecting more than one language treebank. In this line, previous studies have shown that training on harmonized treebanks, i.e. treebanks with the same annotation guidelines but coming from different languages, could improve performance over the corresponding monolingual model (Vilares et al., 2016), which has applications to less-resourced languages (Ammar et al., 2016).

Setup To test this, we train models on many x-inflected treebanks and evaluate them on the corresponding target LR test sets. Here, we also consider merging the available training data for the target LR language, to have a better understanding of how our approach behaves in few-shot setups. Particularly, we combine all the x-inflected treebanks from the phylogenetic groups described for the previous experiment (see again Table 1), instead of training

separate models for each one. We compare the performance against two baselines: (i) models trained on the target LR language training set, and (ii) models trained on a merged training treebank composed of the training set of the target LR treebank and the original training sets of the source treebanks of the related languages (but without x-inflecting them).

Results Figure 4 shows the LAS differences between the merged original and x-inflected models *with respect to the* models that are only trained on data coming from the target LR language (UAS results in Appendix D). For the GB parser, all models trained on merged (original or x-inflected) treebanks perform better than their counterparts trained only on the LR treebank, suggesting that adding data from similar languages helps the parsers. However, merging non-x-inflected treebanks sometimes outperforms merging x-inflected treebanks (e.g. Livvi, Lithuanian, and Latin). For the SL parser merging treebanks is not always beneficial compared to training only on the LR training set. We see that the models trained on harmonized (original or x-inflected) treebanks improve only half of the times. Yet, we see some interesting patterns. For instance, when the x-inflection benefits a sequence-labeling model, it also benefits the graph-based one for the same merged treebank, and *vice versa*. Overall, merging x-inflected treebanks is the best option for 6 out of 10 models, although in many cases the differences are small.

5 Discussion

The results show that the proposed method is able to improve parsing results for some treebank pairs under both zero- and few-shot setups, but it also obtains decreases for other pairs. Due to the high number of factors involved, we were unable to clearly

isolate those that are beneficial and those that are harmful. However, we identified some reasons that could partially explain the behaviour of the method:

- PCCs from Table 2 show that having more UM data is beneficial to obtain better parsing performance, so better inflectors create better x-inflected treebanks.

- Conversion between UM and UD schemata is non-trivial and dependent on the language pair (see McCarthy et al. (2018) for a detailed analysis), and thus incorrect feature conversions could express different morphological information and mislead the parser.

- Although both UD and UM aim to follow a universal annotation schema, not all languages are annotated exactly in the same way, expressing similar morphological phenomena with slightly different features or omitting some of them. Therefore, even when the conversion between schemata is correct, the annotation discrepancies between languages may confuse the inflector, which again, would output a word whose form would express different morphological information than the original form.

6   Conclusion

By cross-inflecting a rich-resource UD treebank using an inflector from a low-resource related language, we can obtain silver, syntactically annotated data to train dependency parsers. Although containing noise and grammatical imperfections, we aimed to test whether the approach could boost performance. The results show that it is possible to obtain improvements (but also decreases) both for zero- and few-shot setups.

About this, we could not clearly identify what aspects make the approach succeed or fail. Although we identified moderate correlations between scores and the amount of available UM data for the target language, we hypothesize that other aspects that are hard to measure could be playing a role: (i) incorrect/incomplete feature conversion from UM to UD schemata that might make the cross-lingual inflections carry different information that the inflections in the original language, or (ii) unknown input features for a given inflector due to differences in exhaustiveness between the UM and UD annotations.

References

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.

Yevgeni Berzak, Yan Huang, Andrei Barbu, Anna Korhonen, and Boris Katz. 2016. Anchoring and agreement in syntactic annotations. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2215–2224, Austin, Texas. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The sigmorphon 2016 shared task—morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 10–22.

Mathieu Dehouck and Pascal Denis. 2018. A framework for understanding the role of morphology in Universal Dependency parsing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2864–2870, Brussels, Belgium. Association for Computational Linguistics.

Mathieu Dehouck and Carlos Gómez-Rodríguez. 2020. Data augmentation via subtree swapping for dependency parsing of low-resource languages. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3818–3830, Barcelona, Spain (Online). International Committee on 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. Association for Computational Linguistics.

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. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Jason Lee, Kyunghyun Cho, and Thomas Hofmann. 2017. Fully character-level neural machine translation without explicit segmentation. *Transactions of the Association for Computational Linguistics*, 5:365–378.

Xuezhe Ma, Zecong Hu, Jingzhou Liu, Nanyun Peng, Graham Neubig, and Eduard Hovy. 2018. Stack-pointer networks for dependency parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1403–1414, Melbourne, Australia. Association for Computational Linguistics.

Arya D. McCarthy, Christo Kirov, Matteo Grella, Amrit Nidhi, Patrick Xia, Kyle Gorman, Ekaterina Vylomova, Sabrina J. Mielke, Garrett Nicolai, Miikka Silfverberg, Timofey Arkhangelskiy, Nataly Krizhanovsky, Andrew Krizhanovsky, Elena Klyachko, Alexey Sorokin, John Mansfield, Valts Ernštreits, Yuval Pinter, Cassandra L. Jacobs, Ryan Cotterell, Mans Hulden, and David Yarowsky. 2020. UniMorph 3.0: Universal Morphology. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3922–3931, Marseille, France. European Language Resources Association.

Arya D. McCarthy, Miikka Silfverberg, Ryan Cotterell, Mans Hulden, and David Yarowsky. 2018. Marrying Universal Dependencies and Universal Morphology. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 91–101. Association for Computational Linguistics.

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.

Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. Low-resource parsing with crosslingual contextualized representations. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 304–315, Hong Kong, China. Association for Computational Linguistics.

Alberto Muñoz-Ortiz, Michalina Strzyz, and David Vilares. 2021. Not all linearizations are equally data-hungry in sequence labeling parsing. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 978–988, Held Online. INCOMA Ltd.

Tiago Pimentel, Maria Ryskinaì, Sabrina MielkeZ, Shijie WuZ Eleanor Chodroff, Brian LeonardB, Garrett Nicolaiá, Yustinus Ghanggo AteÆ, Salam Khalifaè, Nizar Habashè, Charbel El-KhaissiS, et al. 2021. Sigmorphon 2021 shared task on morphological reinflection: Generalization across languages. *SIGMORPHON 2021*, page 154.

Michalina Strzyz, David Vilares, and Carlos Gómez-Rodríguez. 2019. Viable dependency parsing as sequence labeling. 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 717–723, Minneapolis, Minnesota. Association for Computational Linguistics.

Michalina Strzyz, David Vilares, and Carlos Gómez-Rodríguez. 2020. Bracketing encodings for 2-planar dependency parsing. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2472–2484, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Clara Vania. 2020. *An examination of keystroke dynamics for continuous user authentication*. Ph.D. thesis, Queensland University of Technology.

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. 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.

Dingquan Wang and Jason Eisner. 2018. Synthetic data made to order: The case of parsing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1325–1337, Brussels, Belgium. Association for Computational Linguistics.

Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1901–1907, Online. Association for Computational Linguistics.

Shijie Wu, Pamela Shapiro, and Ryan Cotterell. 2018. Hard non-monotonic attention for character-level transduction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4425–4438, Brussels, Belgium. Association for Computational Linguistics.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Ĥórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Ethan Chi, Yongseok Cho, Jinho Choi, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Krister Lindén, Nikola Ljubešić, Olga Loginova, Andry Luthfi, Mikko Luukko, 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, Hiroshi Matsuda, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayọ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Arzucan Özgür, Balkız Öztürk Başaran, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Benoît Sagot, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh

Shohibussirri, Dmitry Sichinava, Einar Freyr Sigurðsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Steinþór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Samson Tella, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, Hanzhi Zhu, and Anna Zhuravleva. 2020. Universal dependencies 2.7. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.