QP2 Informal Proposal

Sarah Payne
Stony Brook University
sarah.payne@stonybrook.edu

Abstract

Morphological inflection is a fundamental task in subword NLP, popularized by the recent SIGMORPHON shared tasks; it has both practical and cognitive implications. State-of-the-art inflection systems have reported extremely high average accuracy across languages, even with relatively little data. However, average accuracy on inflection tasks is also highly variable,

1 Inflection as an NLP Task

In the inflection task, the input is a lemma and a feature- bundle, and we aim to predict the respective inflected word-form.

The training input in SIGMORPHON's shared-tasks is a random split of the available (lemma,form,features) triplets such that no triplet occurring in the train-set occurs in the test-set

Goldman et al. (2022) In such a setting, models can short-cut their way to better predictions in cases where forms from the same lemma appear in both the train and test data. This may allow models to memorize lemma- specific alternations that make morphological in- flection a challenging task to begin with.

(**TODO:** More thoroughly define the task here)

2 Adversarial Evaluation

Many subfields of NLP and machine learning in general suggested hard splits as means to improve the probing of models' ability to solve the underlying task, and to make sure models do not simply employ loopholes in the data. The addition of unanswerable questions to question answering benchmarks (Rajpurkar et al., 2018), or the addition of expert-annotated minimal pairs (Gardner et al., 2020). Narayan et al. (2017) suggested using the WEBSPLIT data, where models are required to split and rephrase complex sentences associated with a meaning representation over a knowledge-base. Aharoni and Goldberg (2018) found that some facts appeared in both train and test sets and provided a harder split denying models the ability to use memorized facts. Aharoni and Goldberg (2020) also suggested a general splitting method for machine translation such that the domains are as disjoint as possible. In semantic parsing, Finegan-Dollak et al. (2018) suggested a better split for parsing natural language questions to SQL queries by making sure that queries of the same template do not occur in both train and test, while Lachmy et al. (2022) split their HEXAGONS data such that any one visual pat- tern used for the task cannot appear in both train and test. Furthermore, Loula et al. (2018) adversarially split semantic parsing for navigation data to assess their models' capability to use compositionality. In spoken language understanding Arora et al. (2021) designed a splitting method that will account for variation in both speaker identity and linguistic content.

In general, concerns regarding data splits and their undesired influence on model assessments led Gorman and Bedrick (2019) to advocate random splitting instead of standard ones. A common modification is re-splitting the data such that the test set is more challenging and closer to the intended use of the models in the wild (Søgaard et al., 2021). As the performance on morphological inflection models seems to have saturated on high scores, a similar rethinking of the data used is warranted.

3 Previous Work

3.1 Goldman et al

Goldman et al. (2022) propose a re-evaluation of morphological inflection models in which train-test splits are formed by splitting by *lemma* rather than by inflected form. When they evaluate the top 3 systems on SIGMORPHON's 2020 shared task, they find that the lemma-split leads to an average drop in accuracy by about 30 percentage points, with the effect being the most significant for low-resourced languages with a drop as high as 95 points. Even high-resourced languages, however, lose about 10 percentage points on average. Goldman et al. argue that their results clearly show that generalizing inflection to unseen lemmas is far from being solved.

So far, the best-performing models have been neural sequence-to-sequence models (Kann and Schütze, 2016; Canby et al., 2020)

There is a view that morphological inflection is a relatively simple task that is essentially already solved, as reflected in the saturation of the results over the years and declining submissions to the shared tasks. However, is it really solved or can the good performance be attributed to some artifacts in the data?

In this work we propose to construct more difficult datasets for morphological inflection by splitting them such that the test set includes no forms of lemmas appearing in the train set.

When re-splitting, we kept the same proportions of the form-split data, we split the inflection tables: 70%, 10%, 20% for the train, dev, and test sets. In terms of examples the proportions may vary as not all tables are of equal size. In prac- tice, the averaged train set size in examples terms was only 3.5% smaller in the lemma-split data, on average.

Used all 90 languages in the SIGMORPHON 2020 shared task.

The models used include:

- **Base LSTM:** character-based seq2seq model with a 1-layer bi-directional LSTM Encoder and a 1-layer unidirectional LSTM Decoder

- **chr-trm:** the character-level transformer baseline of Wu et al. (2021)

- **DeepSpin:** the system is composed of 2 bi-directional LSTM encoders with bi-linear gated Attention, one for the lemma characters and one for the features characters, and a unidirectional LSTM Decoder for generating the outputs. The innovation in the architecture is the use of sparsemax (Martins and Astudillo, 2016) instead of softmax in the attention layer. (Peters and Martins, 2020)

- **CULing**: another transformer, but with restructuring so that the model learns to inflect from any given cell in the inflection table rather than solely from the lemma. (Liu and Hulden, 2020)

All systems see a drop in performance, with average around 30 points and the lowest being 14 points for DeepSpin-02, which fares better for low-resource languages. (**TODO:** do this calculation). The average performance per language family seems to be controlled by training data availability. For example, Germanic languages show average drop of 23 points, while for Niger-Congo languages the drop is 39 points on average. The major drops in performance that contributed the most to the overall gap between the splits are in those low-resourced language. Re- markably, for some systems and languages the drop can be as high as 95 points. On the other hand, on high-resourced languages with 40,000 training ex- amples or more, all systems didn't lose much. here is no evidence for specific families being eas- 867 ier for inflection when little data is provided

4 Kodner et al

Kodner et al. (2022) emphasize generalization along diferent dimenstions by evaluating test items with unseen lemmas and unseen features separately under small and large training conditions. Across the six submitted systems and two baselines, the prediction of inflections with unseen features proved challenging. This was true even for languages for which the forms were in principle predictable, which suggests that further work is needed in designing systems tat capture the various types of generalization needed for the worlds languages.

Generalization, the ability to extend patterns from known to unknown items, is a critical part of mor- phological competence. Morphological sparsity

In principle, there are at least two kinds of generalization which can be evaluated in our UniMorph-based test paradigm: generalization to unseen lemmas, and generalization to unseen inflectional categories (i.e., unseen feature sets). Contrasting seen and unseen lemmas and categories yields four different test conditions: 1) prediction of the form of a novel combination of a seen lemma and seen feature set, 2) prediction given a seen lemma but novel feature set, 3) prediction given a seen feature set but novel lemma, and 4) the prediction of a form when both the lemma and feature set are novel.

5 Introduction

The SIGMORPHON shared task (Cotterell et al., 2016, 2017, 2018; McCarthy et al., 2019; Vylomova et al., 2020; Pimentel et al., 2021; Kodner and Khalifa, 2022; Goldman et al., 2023)

UniMorph (McCarthy et al., 2020; Batsuren et al., 2022)

Goldman et al. (2022)

Kodner et al. (2023c)

Kodner and Khalifa (2022)

Kodner et al. (2023b)

Kodner et al. (2023a)

Morphological inflection is a fundamental task in sub-word NLP, popularized by the recent SIGMORPHON shared tasks; it has both practical and cognitive applications.

Morphological inflection is a popular task in subword NLP with both practical and cogni- tive applications.

In the domain of Morphology, Inflection is a fundamental and important task that gained a lot of traction in recent years, mostly via SIG- MORPHON's shared-tasks.

For years now, state-of-the- art systems have reported high, but also highly variable, performance across data sets and lan- guages. We investigate the causes of this high performance and high variability; we find sev- eral aspects of data set creation and evaluation which systematically inflate performance and obfuscate differences between languages. To improve generalizability and reliability of re- sults, we propose new data sampling and eval- uation strategies that better reflect likely use- cases. Using these new strategies, we make new observations on the generalization abilities of current inflection systems.

With average ac- curacy above 0.9 over the scores of all lan- guages, the task is considered mostly solved using relatively generic neural seq2seq models, even with little data provided. In this work, we propose to re-evaluate morphological in- flection models by employing harder train-test splits that will challenge the generalization ca- pacity of the models. In particular, as op- posed to the naïve split-by-form, we propose a split-by-lemma method to challenge the per- formance on existing benchmarks. Our exper- iments with the three top-ranked systems on the SIGMORPHON's 2020 shared-task show that the lemma-split presents an average drop of 30 percentage points in macro-average for the 90 languages included. The effect is most significant for low-resourced languages with a drop as high as 95 points, but even high- resourced languages lose about 10 points on average. Our results clearly show that general- izing inflection to unseen lemmas is far from being solved, presenting a simple yet effective means to promote more sophisticated models.

These instructions are for authors submitting papers to *ACL conferences using LaTeX. They are not self-contained. All authors must follow the general instructions for *ACL proceedings,[1] and this document contains additional instructions for the LaTeX style files.

The templates include the LaTeX source of this document (`acl_latex.tex`), the LaTeX style file used to format it (`acl.sty`), an ACL bibliography style (`acl_natbib.bst`), an example bibliography (`custom.bib`), and the bibliography for the ACL Anthology (`anthology.bib`).

References

Roee Aharoni and Yoav Goldberg. 2018. Split and rephrase: Better evaluation and stronger baselines. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 719–724, Melbourne, Australia. Association for Computational Linguistics.

Roee Aharoni and Yoav Goldberg. 2020. Unsupervised domain clusters in pretrained language models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7747–7763, Online. Association for Computational Linguistics.

Siddhant Arora, Alissa Ostapenko, Vijay Viswanathan, Siddharth Dalmia, Florian Metze, Shinji Watanabe, and Alan W. Black. 2021. Rethinking end-to-end evaluation of decomposable tasks: A case study on spoken language understanding. In *Interspeech*.

Khuyagbaatar Batsuren, Omer Goldman, Salam Khalifa, Nizar Habash, Witold Kieraś, Gábor Bella, Brian Leonard, Garrett Nicolai, Kyle Gorman, Yustinus Ghanggo Ate, Maria Ryskina, Sabrina Mielke, Elena Budianskaya, Charbel El-Khaissi, Tiago Pimentel, Michael Gasser, William Abbott Lane, Mohit Raj, Matt Coler, Jaime Rafael Montoya Samame, Delio Siticonatzi Camaiteri, Esaú Zumaeta Rojas, Didier López Francis, Arturo Oncevay, Juan López Bautista, Gema Celeste Silva Villegas, Lucas Torroba Hennigen, Adam Ek, David Guriel, Peter Dirix, Jean-Philippe Bernardy, Andrey Scherbakov, Aziyana Bayyr-ool, Antonios Anastasopoulos, Roberto Zariquiey, Karina Sheifer,

[1] `http://acl-org.github.io/ACLPUB/formatting.html`

Sofya Ganieva, Hilaria Cruz, Ritván Karahóǧa, Stella Markantonatou, George Pavlidis, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Candy Angulo, Jatayu Baxi, Andrew Krizhanovsky, Natalia Krizhanovskaya, Elizabeth Salesky, Clara Vania, Sardana Ivanova, Jennifer White, Rowan Hall Maudslay, Josef Valvoda, Ran Zmigrod, Paula Czarnowska, Irene Nikkarinen, Aelita Salchak, Brijesh Bhatt, Christopher Straughn, Zoey Liu, Jonathan North Washington, Yuval Pinter, Duygu Ataman, Marcin Wolinski, Totok Suhardijanto, Anna Yablonskaya, Niklas Stoehr, Hossep Dolatian, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Aryaman Arora, Richard J. Hatcher, Ritesh Kumar, Jeremiah Young, Daria Rodionova, Anastasia Yemelina, Taras Andrushko, Igor Marchenko, Polina Mashkovtseva, Alexandra Serova, Emily Prud'hommeaux, Maria Nepomniashchaya, Fausto Giunchiglia, Eleanor Chodroff, Mans Hulden, Miikka Silfverberg, Arya D. McCarthy, David Yarowsky, Ryan Cotterell, Reut Tsarfaty, and Ekaterina Vylomova. 2022. UniMorph 4.0: Universal Morphology. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 840–855, Marseille, France. European Language Resources Association.

Marc Canby, Aidana Karipbayeva, Bryan Lunt, Sahand Mozaffari, Charlotte Yoder, and Julia Hockenmaier. 2020. University of Illinois submission to the SIGMORPHON 2020 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 137–145, Online. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Arya D. McCarthy, Katharina Kann, Sabrina J. Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. The CoNLL–SIGMORPHON 2018 shared task: Universal morphological reinflection. In *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 1–27, Brussels. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2017. CoNLL-SIGMORPHON 2017 shared task: Universal morphological reinflection in 52 languages. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 1–30, Vancouver. 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, Berlin, Germany. Association for Computational Linguistics.

Catherine Finegan-Dollak, Jonathan K. Kummerfeld, Li Zhang, Karthik Ramanathan, Sesh Sadasivam, Rui Zhang, and Dragomir Radev. 2018. Improving text-to-SQL evaluation methodology. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 351–360, Melbourne, Australia. Association for Computational Linguistics.

Matt Gardner, Yoav Artzi, Victoria Basmov, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, Nitish Gupta, Hannaneh Hajishirzi, Gabriel Ilharco, Daniel Khashabi, Kevin Lin, Jiangming Liu, Nelson F. Liu, Phoebe Mulcaire, Qiang Ning, Sameer Singh, Noah A. Smith, Sanjay Subramanian, Reut Tsarfaty, Eric Wallace, Ally Zhang, and Ben Zhou. 2020. Evaluating models' local decision boundaries via contrast sets. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1307–1323, Online. Association for Computational Linguistics.

Omer Goldman, Khuyagbaatar Batsuren, Salam Khalifa, Aryaman Arora, Garrett Nicolai, Reut Tsarfaty, and Ekaterina Vylomova. 2023. SIGMORPHON–UniMorph 2023 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 117–125, Toronto, Canada. Association for Computational Linguistics.

Omer Goldman, David Guriel, and Reut Tsarfaty. 2022. (un)solving morphological inflection: Lemma overlap artificially inflates models' performance. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 864–870, Dublin, Ireland. Association for Computational Linguistics.

Kyle Gorman and Steven Bedrick. 2019. We need to talk about standard splits. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2786–2791, Florence, Italy. Association for Computational Linguistics.

Katharina Kann and Hinrich Schütze. 2016. MED: The LMU system for the SIGMORPHON 2016 shared task on morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 62–70, Berlin, Germany. Association for Computational Linguistics.

Jordan Kodner and Salam Khalifa. 2022. SIGMORPHON–UniMorph 2022 shared task 0: Modeling inflection in language acquisition. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 157–175, Seattle, Washington. Association for Computational Linguistics.

Jordan Kodner, Salam Khalifa, Khuyagbaatar Batsuren, Hossep Dolatian, Ryan Cotterell, Faruk Akkus, Antonios Anastasopoulos, Taras Andrushko, Aryaman Arora, Nona Atanalov, Gábor Bella, Elena Budianskaya, Yustinus Ghanggo Ate, Omer Goldman, David Guriel, Simon Guriel, Silvia Guriel-Agiashvili, Witold Kieraś, Andrew Krizhanovsky, Natalia Krizhanovsky, Igor Marchenko, Magdalena Markowska, Polina Mashkovtseva, Maria Nepomniashchaya, Daria Rodionova, Karina Scheifer, Alexandra Sorova, Anastasia Yemelina, Jeremiah Young, and Ekaterina Vylomova. 2022. SIGMORPHON–UniMorph 2022 shared task 0: Generalization and typologically diverse morphological inflection. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 176–203, Seattle, Washington. Association for Computational Linguistics.

Jordan Kodner, Salam Khalifa, Sarah RB Payne, and Zoey Liu. 2023a. Re-evaluating the evaluation of neural morphological inflection models. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

Jordan Kodner, Salam Khalifa, and Sarah Ruth Brogden Payne. 2023b. Exploring linguistic probes for morphological generalization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8933–8941, Singapore. Association for Computational Linguistics.

Jordan Kodner, Sarah Payne, Salam Khalifa, and Zoey Liu. 2023c. Morphological inflection: A reality check. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6082–6101, Toronto, Canada. Association for Computational Linguistics.

Royi Lachmy, Valentina Pyatkin, Avshalom Manevich, and Reut Tsarfaty. 2022. Draw Me a Flower: Processing and Grounding Abstraction in Natural Language. *Transactions of the Association for Computational Linguistics*, 10:1341–1356.

Ling Liu and Mans Hulden. 2020. Leveraging principal parts for morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 153–161, Online. Association for Computational Linguistics.

João Loula, Marco Baroni, and Brenden Lake. 2018. Rearranging the familiar: Testing compositional generalization in recurrent networks. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 108–114, Brussels, Belgium. Association for Computational Linguistics.

Andre Martins and Ramon Astudillo. 2016. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1614–1623, New York, New York, USA. PMLR.

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 Twelfth Language Resources and Evaluation Conference*, pages 3922–3931, Marseille, France. European Language Resources Association.

Arya D. McCarthy, Ekaterina Vylomova, Shijie Wu, Chaitanya Malaviya, Lawrence Wolf-Sonkin, Garrett Nicolai, Christo Kirov, Miikka Silfverberg, Sabrina J. Mielke, Jeffrey Heinz, Ryan Cotterell, and Mans Hulden. 2019. The SIGMORPHON 2019 shared task: Morphological analysis in context and cross-lingual transfer for inflection. In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 229–244, Florence, Italy. Association for Computational Linguistics.

Shashi Narayan, Claire Gardent, Shay B. Cohen, and Anastasia Shimorina. 2017. Split and rephrase. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 606–616, Copenhagen, Denmark. Association for Computational Linguistics.

Ben Peters and André F. T. Martins. 2020. One-size-fits-all multilingual models. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 63–69, Online. Association for Computational Linguistics.

Tiago Pimentel, Maria Ryskina, Sabrina J. Mielke, Shijie Wu, Eleanor Chodroff, Brian Leonard, Garrett Nicolai, Yustinus Ghanggo Ate, Salam Khalifa, Nizar Habash, Charbel El-Khaissi, Omer Goldman, Michael Gasser, William Lane, Matt Coler, Arturo Oncevay, Jaime Rafael Montoya Samame, Gema Celeste Silva Villegas, Adam Ek, Jean-Philippe Bernardy, Andrey Shcherbakov, Aziyana Bayyr-ool, Karina Sheifer, Sofya Ganieva, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Andrew Krizhanovsky, Natalia Krizhanovsky, Clara Vania, Sardana Ivanova, Aelita Salchak, Christopher Straughn, Zoey Liu, Jonathan North Washington, Duygu Ataman, Witold Kieraś, Marcin Woliński, Totok Suhardijanto, Niklas Stoehr, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Richard J. Hatcher, Emily Prud'hommeaux, Ritesh Kumar, Mans Hulden, Botond Barta, Dorina Lakatos, Gábor Szolnok, Judit Ács, Mohit Raj, David Yarowsky, Ryan Cotterell, Ben Ambridge, and Ekaterina Vylomova. 2021. SIGMORPHON 2021 shared task on morphological reinflection: Generalization across languages. In *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 229–259, Online. Association for Computational Linguistics.

Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. Know what you don't know: Unanswerable questions for SQuAD. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789, Melbourne, Australia. Association for Computational Linguistics.

Anders Søgaard, Sebastian Ebert, Jasmijn Bastings, and Katja Filippova. 2021. We need to talk about random splits. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1823–1832, Online. Association for Computational Linguistics.

Ekaterina Vylomova, Jennifer White, Elizabeth Salesky, Sabrina J. Mielke, Shijie Wu, Edoardo Maria Ponti, Rowan Hall Maudslay, Ran Zmigrod, Josef Valvoda, Svetlana Toldova, Francis Tyers, Elena Klyachko, Ilya Yegorov, Natalia Krizhanovsky, Paula Czarnowska, Irene Nikkarinen, Andrew Krizhanovsky, Tiago Pimentel, Lucas Torroba Hennigen, Christo Kirov, Garrett Nicolai, Adina Williams, Antonios Anastasopoulos, Hilaria Cruz, Eleanor Chodroff, Ryan Cotterell, Miikka Silfverberg, and Mans Hulden. 2020. SIGMORPHON 2020 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 1–39, Online. 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.