To POS Tag or Not to POS Tag: The Impact of POS Tags on Morphological Learning in Low-Resource Settings

Sarah Moeller   Ling Liu   Mans Hulden
University of Colorado Boulder
`first.last@colorado.edu`

Abstract

Part-of-Speech (POS) tags routinely appear as features in morphological tasks. POS taggers are often one of the first NLP tools developed for low-resource languages. However, as NLP expands to new languages it cannot assume that POS tags will be available to train a POS tagger. This paper empirically examines the impact of POS tags on two morphological tasks with the Transformer architecture. Each task is run twice, once with and once without POS tags, on otherwise identical data from ten well-described languages and five under-documented languages. We find that the presence or absence of POS tags does not have a significant bearing on the performance of either task. In joint segmentation and glossing, the largest average difference is an .09 improvement in F_1-scores by removing POS tags. In reinflection, the greatest average difference is 1.2% in accuracy for published data and 5% for unpublished data. These results are indicators that NLP and documentary linguistics may benefit each other even when a POS tag set does not yet exist for a language.

1   Introduction

Parts of speech (POS), also known as word classes or lexical categories, communicate information about a word, its morphological structure and inflectional paradigm, and its potential grammatical role in a clause. POS tagging is a well-studied problem in NLP. It is one of the first tasks undertaken for a new data set and a POS tagger is often one of the first NLP resources built for low-resource languages (Yarowsky and Ngai, 2001; Cox, 2010; De Pauw, 2012; Baldridge and Garrette, 2013; Duong, 2017; Anastasopoulos, 2019; Millour and Fort, 2019; Eskander et al., 2020b). Although this priority on early POS tagging may be simply due to the relative ease of building a POS tagger, it seems to reflect an assumption that POS
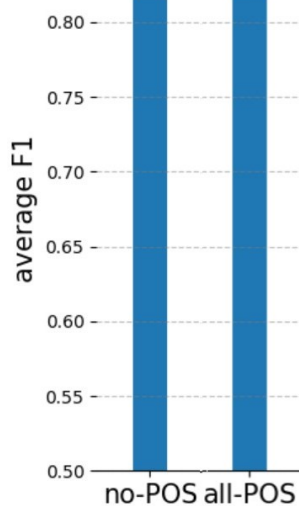

Figure 1: Average F_1-scores on joint segmentation and glossing on interlinear glossed texts from fieldwork in five languages found that POS-tags have little and irregular impact.

tags simplify or improve other NLP tasks (Krauwer, 2003). As far as we are aware, this assumption has not been methodically tested.

This paper examines the impact of POS tags on morphological learning, an important area for low-resource languages, many of which are more morphologically complex than English, Mandarin, or other large-resource languages. Morphological learning can help reduce the out-of-vocabulary problem in morphologically complex languages, especially in low-resource settings. Morphological learning also holds high priority in documentary and descriptive linguistics as a necessary foundation for further descriptive work. We focus on two related tasks that involve morphological learning: joint morpheme segmentation/glossing and morphological reinflection. Joint segmentation and glossing segments a word into its component morphemes and glosses the segments. Reinflection gen-
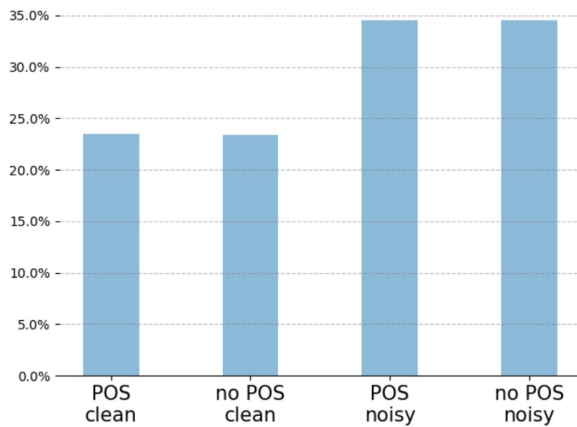
966

Figure 2: During reinflection generation of four interlinear field corpora and four "cleaned" versions of those corpora the presence or absence of POS tags does not make a significant or consistent difference in accuracy of inflected forms.

erates unseen inflected word forms from morphological features based on a language's inflectional patterns. Since lexical categories (POS) are identified partly by morphological structure, it seems reasonable to assume the reverse – that knowing a word's part of speech makes it easier for a model to analyze its morphological structure. For example, knowing that a word is a noun in English makes it extremely unlikely that a final substring *(e)n* could be a participial affix (e.g. "oven" - NOUN; cf. "driven" - VERB). On the other hand, POS tags may be providing redundant information when, for example, an affix that marks a morphosyntactic feature is identical across all categories where that feature appears (e.g. the Russian morpheme /-i/ 'PL' is identical for for plural nouns and plural verb agreement). However, these hypotheses must be tested before claiming either one.

The impact of (not) having POS tags has perhaps not been examined closely in part because it seems safe to assume that POS tags or a POS tagger will be available. However, as NLP expands its reach to new languages, POS tags may not be readily available. In fact, the lexical categories present in the language may not even be described yet when data becomes available. In documentary and descriptive linguistics, the description and tagging of lexical categories takes a relatively low priority compared to its place in NLP (cf. Bird and Chiang (2012)'s workflow). Yet interlinear glossed texts (IGT) are often the largest available annotated resource for a low-resource language—and sometimes the only available resource.

The impact of POS tags on computational morphology may hold implications for linguistic theory as well. The nature of lexical categories (Rauh, 2010), the criteria for identifying them (Croft, 2000), and even their very reality as a universal property of language (Gil, 2005) are not entirely settled among linguists. If the morphological structure of unseen words can be analyzed and generated without reference to lexical categories, then perhaps such categories should not be considered an inherent property of the lexicon (Rauh et al., 2016).

This paper describes experiments that were run on corpora differing only in the presence or absence of POS tags. The results, which are generalized in Figures 1 and 2, indicate that POS tags do not have significant impact on computational morphological learning. Section 2 presents related work in lexical categories, POS-tagging, segmentation and glossing, and (re)inflection. Sections 3 and 4 describe the corpora and the NLP architecture used. The segmentation and glossing task and results are presented in Section 5. The reinflection task and results are presented in Section 6. Implications of both experiments are discussed in Section 7.

2 Related Work

Work on POS tagging has led to the development of several related resources in NLP and linguistics including numerous methods for automatic tagging (e.g. Kupiec (1992); Toutanova and Johnson (2008)) as well as tag sets. The most popular tag set for English was developed by the Penn Treebank Project (Taylor et al., 2003). A universal POS tag set was proposed by Petrov et al. (2012) and has been widely adopted. It closely follows traditional linguistic conventions for common lexical categories as can be seen by comparing to the Leipzig Glossing Rules (Institute, 2008) which also has recommended tags for less common categories.

Many NLP models have been applied to segmentation and glossing of low-resource languages but they often tackle just one of the two tasks, e.g. segmentation only (Ruokolainen et al., 2014; Wang et al., 2016; Kann et al., 2018; Mager et al., 2020; Sorokin, 2019; Eskander et al., 2020a). Automatic morpheme segmentation was introduced by Harris (1970) and much earlier segmentation research implemented unsupervised learning (Goldsmith, 2001; Creutz and Lagus, 2002; Poon et al., 2009). Published linguistic descriptive data is used

as training data usually after some preprocessing. Glossing-only experiments make the assumption that data is already segmented into morphemes. For example, McMillan-Major (2020) trained a conditional random field (CRF) model to produce a gloss line for several high-resource languages and three low-resource languages. The low-resource language data came from interlinearized data that was polished for publication. McMillan-Major (2020) and some other experiments such as Samardzic et al. (2015) use information from lines of interlinearized texts such as translation and POS tags.

Computational approaches to morphological inflection or reinflection have been developed by Durrett and DeNero (2013); Nicolai et al. (2015); Liu and Mao (2016); Cotterell et al. (2017); Kann and Schütze (2016); Aharoni and Goldberg (2017), etc. Some of the work was developed as part of the SIGMORPHON Shared Tasks.[1] Our work partly replicates the CoNLL-SIGMORPHON reinflection shared tasks (Cotterell et al., 2016, 2017, 2018a). Sequence-to-sequence neural network models have been very successful at handling the morphological (re)inflection task, even in low-resource conditions with model improvement designed to tackle the situation (Kann et al., 2017; Silfverberg et al., 2017; Sharma et al., 2018; Makarov and Clematide, 2018; Anastasopoulos and Neubig, 2019; Wu and Cotterell, 2019; Liu, 2021). The Transformer (Vaswani et al., 2017a) is the model architecture which produces the current state-of-the-art performance on this task (Vylomova et al., 2020; Wu et al., 2020; Liu and Hulden, 2020b,a). Therefore, we use the Transformer for all the experiments in this paper.

This paper is an expansion of a section in Moeller et al. (2020). The experimental setup and SIGMORPHON languages are the same as that work, but it does not look at what happens when POS tags are available in the field data. We expanded the re-inflection task to field corpora. we also ran the SIGMORPHON experiments 5 times instead of one time. The addition of the segmentation and glossing was inspired by Moeller and Hulden (2021).

3 Data

We use published data in ten languages and unpublished data in five low-resource languages. The published and unpublished data is used for the mor-

Language	POS
Adyghe	N, ADJ
Arabic	N, V, ADJ
Basque	V
Finnish	N, V, ADJ
German	N, ADJ
Persian	V
Russian	N, V, ADJ
Spanish	N, V
Swahili	N, V, ADJ
Turkish	N, V, ADJ

Table 1: SIGMORPHON languages and the lexical categories found in the data.

phological reinflection but only the unpublished data for segmentation and glossing.

3.1 SIGMORPHON Data

For the morphological reinflection task we use datasets that were released for the CoNLL-SIGMORPHON 2018 shared task 1 (Cotterell et al., 2018b). We selected 10 languages that belong to different families and are typologically diverse with regards to morphology. The languages and the inflected lexical categories available for the shared task are listed in Table 1. The language family and morphological typology for each language is available on the UniMorph official website.[2] Only the listed lexical categories were POS-tagged.

3.2 Interlinear Glossed Texts

The manually-annotated interlinear glossed texts (IGT) were created in documentary and descriptive projects for five low-resource and under-documented languages. The corpora represent a range of documentary field projects rather than a range of language typology, although they do represent three different language families on four continents. It is difficult to find corpora of under-documented languages with (enough) POS tags to conduct our POS experiments precisely because of the low priority of POS-tagging in documentary and descriptive linguistics. We were unable to use half of the field corpora available to us for this reason. However, because we are interested in leveraging NLP for fieldwork, we felt it is important to work with the noisy field data, rather than use (often morphologically simpler) high-resource

---

Language	Tokens	POS-tagged		Inflected
Alas	4.5k	3845	86%	623
Lamkang	101k	46,557	46%	n/a
Lezgi	14k	13,636	96%	843
Manipuri	12k	2067	17%	3,260
Natügu	16.5k	10,994	66%	1,954

Table 2: The approximate total number token counts in the field data does not include multiple-word-expressions (when parsed as such) and ignores personal nouns and digits. The amount of segmented, glossed, and POS-tagged tokens is shown as a number and a percentage of the corpus. Of those the inflected words were usable for the reinflection.

languages with reduced data size.[3]

The corpora were compiled during projects that each had their own priorities and workflow and this resulted in the differing amounts of annotation shown in Table 2.[4] Only the tokens that were segmented, glossed, and POS-tagged could be used. The POS tags were provided by the annotators. For the reinflection task, the data was further limited to inflected forms. The collection of inflected forms was automatically extracted and grouped based on the gloss of the root morpheme (noisy version). We happened to have cleaned versions for the re-inflection task and include those for the sake of completeness. The cleaned versions were created from the noisy versions that had been checked by language experts.[5]

It is worth noting that the Lamkang (used only for the segmentation and glossing study), Manipuri, and Natügu corpora are the result of many years of work and these extended projects eventually led to significant POS tagging. Two other large and completely segmented/glossed corpora could not be included because the lexical categories had not been tagged. The Lezgi project used POS tags at an early stage because the research was focused on verb tenses (Donet, 2014). All POS tags in the smaller Alas corpus, and many in the Lezgi corpus, were added specifically for our research.

Alas [btz] (Alas-Kluet, Batak Alas, Batak Alas-Kluet) is an Austronesian language spoken by 200,000 people on the Indonesian island of Sumatra (Eberhard et al., 2020). Its morphology features reduplication, infixation, and circumfixation. The POS set in the corpus is: **ADJ**, **ADV**, AUX, CARDNUM, CLF, **CONJ**, COP, DEM, DISTRNUM, EXISTMRKR, INTERJ, **N**, NPROP, ORDNUM, **PREP**, **PRO**, PRT, QUANT, REFL, RELPRO, **V**, VD, **VI**, **VT**.[6]

Lamkang [lmk] is a Northern Kuki-Chin language of the Tibeto-Burman family with an estimated 4 to 10 thousand speakers primarily in Manipur, India but also in Burma (Thounaojam and Chelliah, 2007). Its morphology tends toward agglutination with many stem-stem patterns to signal syntactic categories. The corpus is accessible through the Computational Resources for South Asian Languages (CoRSAL) digital archive at the University of North Texas.[7] The POS tag set is: ADN, ADVL, DEM, CONN, COORDCONN, COP, INTERJ, N, NPR, NUM, ORDNUM, POSTP, PRON, PTC, QUANT, SUBO, UNK, V, VC, VI, VT.

Lezgi [lez] (Lezgian) is a highly agglutinative language belonging to the Lezgic branch of the Nakh-Daghestanian (Northeast Caucasian) family. It is spoken by over 400,000 speakers in Russia and Azerbaijan (Eberhard et al., 2020). It features overwhelmingly suffixing agglutinative morphology. The POS tag set is: **ADJ**, **ADV**, CARDNUM, CONN, COORDCONN, **DEM**, DET, INDFPRO, INTERJ, INTERROG, MSD, MULTIPNUM, **N**, **NPROP**, **NUM**, ORDNUM, **PERS**, **POSS**, POST, PREP, **PRO**, PROFORM, PRT, PTCP, **RECP**, SUBORDCONN, **V**,

---

[3]We investigated the Online Database of Interlinear Text (ODIN) since the AGGREGATION project at University of Washington has projected POS tags from English, but as yet, we have not found a corpus of comparative size to the smallest field corpus. Perhaps because we focused on finding more polysynthetic languages in order to balance the diversity of morphological types and because preprocessing the ODIN format is time-consuming.

[4]Rights holders gave informed consent to use the data for this research and links are provide to the corpora that are publicly available.

[5]Inflection data available at: https://github.com/LINGuistLIU/IGT

[6]All POS were used for the segmentation and glossing task. Tags in boldface indicate POS that are inflected and were therefore used in the reinflection task.

[7]https://digital.library.unt.edu/explore/collections/SAALT/

VERBPRT, **VF**, **VNF**, VOC.

Manipuri [mni] (Meitei, Meetei) is a Tibeto-Burman language spoken by nearly two million people, primarily in the state of Manipur, and is one of India's official languages. It nonetheless has been classified as vulnerable to extinction (Moseley, 2010). It is a tonal language with weakly suffixing, agglutinative morphology (Chelliah, 1997). The corpus is at CoRSAL.[8] The POS set is: **ADV**, **INTERJ**, **N**, **PROFORM**, **UNK**, **V**.

Natügu [ntu] belongs to the Reefs-Santa Cruz group in the Austronesian family and is spoken by about 4,000 people in the Temotu Province of the Solomon Islands. It has mainly agglutinative morphology with complex verb structures (Næss and Boerger, 2008). The corpus is stored at SIL Language & Culture Archives.[9] The POS tags set is: **A-D-P2**, **ADJ**, **ADV**, CLAUSE, **CONJ**, **DEM**, **DET**, **GEN**, **GERUND**, INTERROG, INTJ, **N**, **N.(KX.CL)**, NCOMP, **NEG**, **NOM1**, NP, NP(COMP), NPROP, **NUM**, **ORD**, **PARTICLE**, **PCLF**, **PER-SPRO**, **PHRASE**, **PN**, **POSSPRO**, **PREP**, **PRO**, **RPRN**, **SUBR**, **UNK**, **V**, **VI**, VP, **VT**, Z-GERUND.

4 Models

For simple comparisons, we chose a single neural model architecture for both tasks. The tasks were trained with the Transformer (Vaswani et al., 2017b), the current state-of-the-art neural model architecture for morphological tasks (Vylomova et al., 2020; Liu and Hulden, 2020b). We used the implementation of the Transformer model in the Fairseq toolkit (Ott et al., 2019)[10] with character-level transduction (Wu et al., 2020) for morphology learning in low-resource settings. Following (Wu et al., 2020), we employ $N = 4$ layers for the encoder and the decoder, each with 4 self-attention heads. The embedding size for the encoder and decoder is 256, and the hidden layer size is 1024. We use a dropout rate of 0.3 for encoding and beam search with a width of 5 at decoding time. The Adam algorithm (Kingma and Ba, 2014) ($\beta_1 = 0.9$, $\beta_2 = 0.98$) is used to optimize the cross entropy loss with label smoothing (Szegedy et al., 2016) of 0.1. All models have been trained on an NVIDIA

GP102 [TITAN Xp] GPU for 10k maximum updates with a batch size of 400.

5 POS for Segmentation and Glossing

The first study asks whether POS tags makes a significant impact on automated morpheme segmenting and glossing. The experiment tests and compares two models on data that is identical except for the presence/lack of POS tags.

We chose morpheme segmentation and glossing because it is a high-priority and early step in documenting and describing new languages. Segmenting words into morphemes and glossing (strictly translating) them is usually the first task undertaken after new data has been transcribed. Therefore, it is important to study how to provide and improve automated assistance for field linguists. Automatic systems could greatly benefit the analysis of endangered languages and combat the "annotation bottleneck" caused by current manual methods (Simons and Lewis, 2013; Holton et al., 2017; Seifart et al., 2018).

Although adding POS tagging as a high-priority task would add to that bottleneck, if the tags have a significant and positive impact on automating segmentation and glossing, then linguists may receive long-term benefits from the addition to their workflow. Therefore, we explore the impact of POS tags at very low-resource settings and the impact of POS tags when a new field project takes time to tag some, but not all, tokens. This is also why we chose noisy field corpora, rather than published, polished corpora which are not like the data that linguists typically work with. We are interested in how POS tags influence segmentation and glossing in the earliest work with a new language.

5.1 Experimental Setup

Three Transformer models were trained. The English example in (1) shows the input and output of models 1, 2, and 3. Model 1, shown in (1a), has no POS tags. Models 2 and 3 have a POS tags, as shown in (1b). Model 2 has POS tags on every word but Model 3 includes POS tags only for some words, simulating projects unable to complete POS-tagging.

(1) a. **INPUT 1:** t a x e s

    b. **INPUT 2/3 :** t a x e s N

    c. **OUTPUT:** tax#levy -es#PL

---

[8] https://digital.library.unt.edu/explore/collections/MDR
[9] https://www.sil.org/resources/search/language/ntu
[10] https://fairseq.readthedocs.io/en/latest/

Language	1%	3%	6.5%	10%	20%	30%	40%	100%
Alas	.00	.02	.02	.03	.05	.05	.04	-.09
Lamkang	.05	.08	.07	.07	.08	.08	.08	-.01
Lezgi	.03	-.01	.02	.04	.03	.03	.03	.02
Manipuri	-.01	.00	.01	.00	.00	.01	.00	.00
Natügu	.01	.03	.02	.03	.02	.03	.04	.00

Table 3: The difference in F_1 scores with/out POS tags when training segmentation and glossing on increasing amounts of annotated data, as percentages of total available training data. Negative scores indicates that adding POS tags improve results.

Language	0%	1%	3%	6.5%	10%	20%	30%	40%	100%
Alas	.6902	.6448	.6415	.6546	.6517	.6627	.6647	.6708	**.6968**
Lamkang	.8573	.8074	.8195	.8298	.8332	.8482	.8527	.8524	**.8645**
Lezgi	.7501	**.7564**	.7542	.7529	.7505	.7498	.7480	.7471	.7317
Manipuri	.8903	.8885	.8877	.8882	.8889	.8874	.8896	.8897	**.8921**
Natügu	.8995	.8748	.8782	.8864	.8855	.8932	.8999	.8965	**.9006**

Table 4: F_1 scores on segmenting and glossing when trained on all data with increasing percentages of POS tags.

All three models are trained on all the available training data. Models 1 and 2 are also trained on different proportions of training data in order to simulate very small corpora. These proportions of training data start at 1% and are gradually increased to 40% of available training data.

Even when POS tags are included in interlinear field data, it is rarely completed as Table 2 clearly indicates.In order to simulate this reality Model 3 was trained on all the available training data but the proportion of inputs with POS tags was gradually and randomly increased.

The training/development/test split is 8/1/1. All models are trained and evaluated on a 10-fold cross-validation. The folds were trained twice, once with and once without POS tags; no other changes were made to the data. All folds were evaluated on a single, consistent held-out test set. Since we wanted to simulate a realistic field situation where the system is segmenting and glossing newly transcribed but unannotated text, the test inputs do not include POS tags.

5.2 Segmentation and Glossing Results

POS tags have no consistent positive or negative effect on automated segmentation and glossing in low-resource settings. The overall impact of POS tags is not significant. Table 3 shows the differences when F_1-scores without POS tags are subtracted from the F_1-scores with POS tags, with various amounts of training data. The largest difference is just under .1 points.

A few interesting observations can be made that should be explored with more languages. Manipuri shows the smallest differences overall; it also has the fewest POS-tagged words and the smallest tag set. The largest differences are seen in the Alas and Lamkang corpora. Alas also has a relatively small amount of POS-tagged words, but it has quite a large tag set. As the size of the Alas training data increases, the impact of POS tags becomes more pronounced, suggesting that perhaps a relatively large POS tag set may have a greater effect on results in medium settings. Lamkang has the largest amount of POS-tagged words, but of those, a significant number were tagged as UNK. It is not clear whether the UNK tag is limited to categories that have not been fully analyzed or if it is a default tag that covers a diverse set of words. The difference made by adding POS tags all but disappears when all the Lamkang data is trained, suggesting that a smaller data set is more impacted by a large tag set or inconsistent annotations.

Overall, increasing the number of POS tags in the training data has minimal impact. Table 4 shows the F_1-scores when the amount of POS tags in the data is gradually increased. For example, at 30%, one of three random training instances have a POS tag. In most cases, having incomplete POS-tagged data hurts performance compared to have POS tags on all words or none at all. The system either performs worse, or, in the case of Lezgi, makes very small improvement (.0063 points). Except for Lezgi, as more POS tags are added, the system

tends to improve slightly but never matches the best scores.

6 POS for Reinflection

The second study asks whether POS tags make a significant impact on learning inflectional patterns and generating unseen inflected forms. We chose the morphological re-inflection task because it is easy to reproduce and to compare with the original SIGMORPHON shared task. Eliciting and analyzing a language's inflectional patterns is a recommended next step after morpheme segmentation and glossing (Bird and Chiang, 2012). The inflectional pattern of a lexeme or a lexical category is also known as a morphological paradigm. Learning morphological paradigms can be viewed in terms of filling in, or generating, the missing forms of a paradigm table by generalizing over inflectional patterns (Ackerman et al., 2009; Ahlberg et al., 2014, 2015; Liu and Hulden, 2017; Malouf, 2017; Silfverberg et al., 2018; Silfverberg and Hulden, 2018).

The experiments in this section partly replicates the CoNLL-SIGMORPHON 2018 shared task 1 of morphological reinflection. Reinflection consists of generating unknown inflected forms, given a related inflected form $f(\ell, \vec{t}_{\gamma_1})$ and a target morphological feature vector $\vec{t}_{\gamma_2}$. Thus, it corresponds to learning the mapping $f : \Sigma^* \times \mathcal{T} \to \Sigma^*$. The goal is then to produce the inflected form $f(\ell, \vec{t}_{\gamma_2})$. An inflected form is generated when the model is given a related inflected form and the target morphological features (which are essentially glosses of affixes) of the inflected form to be generated. In previous work, POS tags have been included by default as part of the morphological features. That is, they have been assumed to be helpful and to be available.

6.1 Experimental Setup

The models were trained on individual languages in three different data sets. The first data set is the published Unimorph inflectional data in ten languages. The second data set is inflected word forms extracted from unpublished IGT in four languages; the third is the clean, or corrected, versions of the second data set. The Unimorph data was extracted from published data and is the "cleanest". Its inflected forms and morphological features were double-checked and the forms provided were selected to provide a balanced picture of the language's morphological structure. The inflected forms extracted from the IGT contains only inflected forms attested in original texts which are transcribed samples of natural oral speech. The noisy version was automatically grouped into paradigms based on the assumption that identical glosses of root morphemes signified the same lemma, and therefore the same morphological paradigm. The clean data was made by asking language experts to examine the noisy data and regroup paradigms when root morphemes were incorrectly glossed. They also corrected typos and morphological features that were incorrectly glossed.

For the Unimorph data, the original SIGMORPHON training/validation/test splits were kept. The prepared *medium* setting of 1,000 training examples was used. This setting was chosen because of the three possible settings (100, 1k and 10k), it is the closest in size to number of inflected word forms extracted from the four IGT corpora, which provided between 600 and 3,000 training examples. An 8/1/1 training/development/tests split was used for the IGT data.

6.2 Reinflection Results

Five reinflection models with random seeds were trained on each data set. All models were trained twice, once with and once without POS tags on the input. Crosswise pairs were compared by subtracting the results with POS tags from the results without POS, giving 25 accuracy scores per language. Figures 3 and 4 show the average and range of differences between the two.

The range of differences shows that POS tags do not have a consistently positive or negative impact. Only two languages show a clear tendency to be impact in one way. In Natügu, POS tags improve accuracy while in Adyghe, they decreases accuracy.

The average difference in accuracy on any data set is rarely more than 1 percentage point. As the data becomes less polished, the impact of POS tags increases slightly and the range of differences grows noticeably. The largest average difference (∼5 percentage points) seen in the noisy data from field IGT. This indicates that time invested in polishing existing IGT data may give a better return than time spent on POS-tagging. For the SIGMORPHON languages, the largest mean difference is barely over 2 points and for the clean IGT-extracted data the largest mean difference is about 3 points.
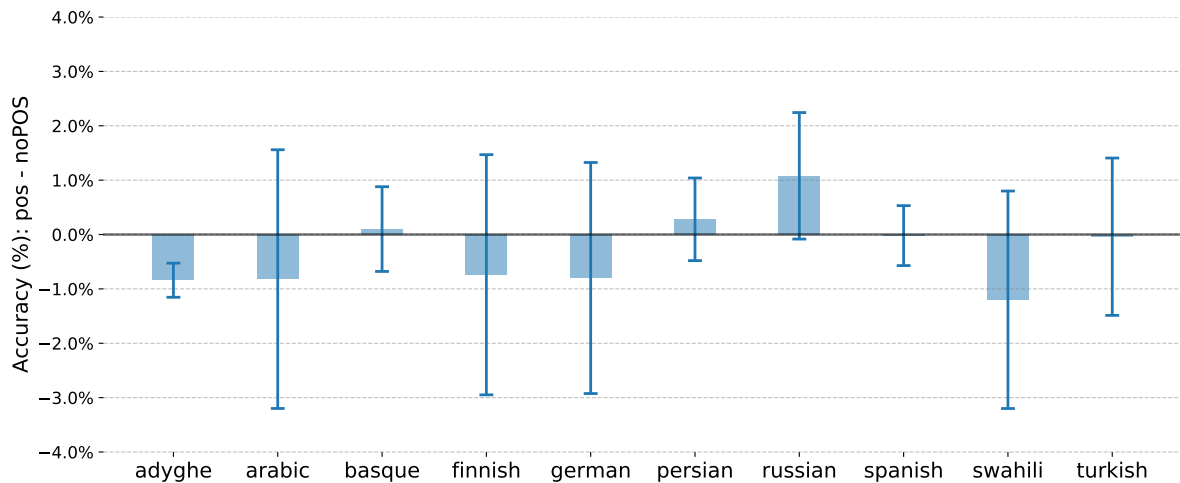
Figure 3: The difference in accuracy with/out POS on the reinflection task with SIGMORPHON languages. Negative scores indicates that adding POS tags improves results. The bar shows the mean of the differences and lines indicate the range of the mean plus or minus the standard deviation.
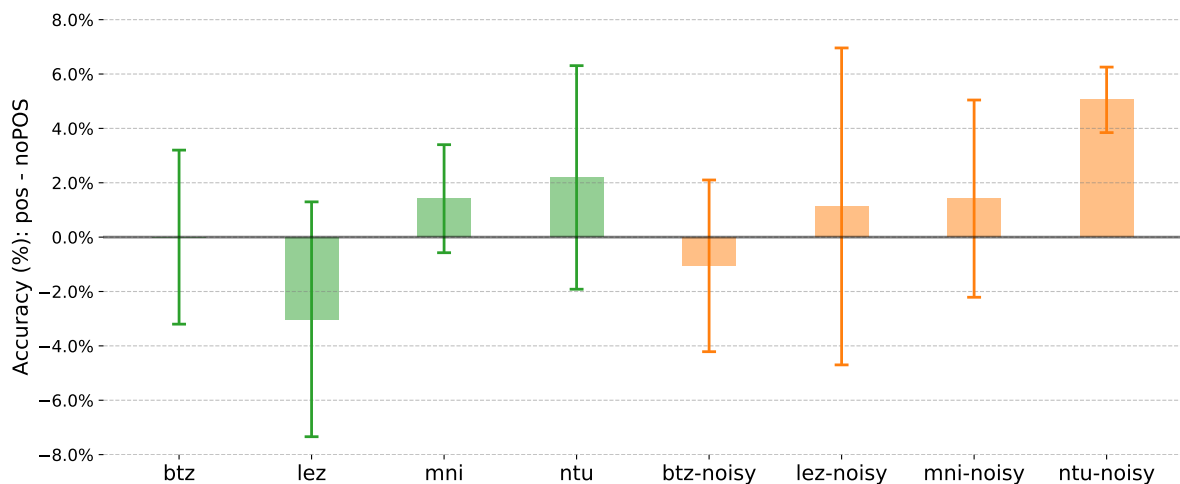

Figure 4: The difference in accuracy with/out POS on the reinflection task with cleaned and noisy field data. Negative scores indicates that adding POS tags improve results. The bar shows the mean of the differences and line indicates the range of the mean plus or minus the standard deviation.

7 Discussion

The number of language we used is not large but a few general observations can be made. For both tasks the impact made by the presence or absence of POS tags is minimal. Still, the best results with a small corpus are achieved when either all or no tokens are POS-tagged, at least for segmentation and glossing. This suggests that having a completely tagged corpus is better than an incompletely tagged corpus, so perhaps limited annotation time might be better spent on more segmentation and glossing.

The size or specificity of the tag set may make a difference in the impact of POS tags. When comparing the tag sets in the CoNLL-SIGMORPHON 2018 shared task data and the IGT from fieldwork,

the difference in the number of lexical categories is significant. The CoNLL-SIGMORPHON 2018 shared task data sets have at most three: noun (N), verb (V), and adjective (ADJ). The IGT corpora have larger tag sets; for example, they may have tags for both finite verb form (VF) and non-finite forms (VNF). The smallest IGT tag set has six categories (Manipuri). That is twice as many POS tags as the SIGMORPHON languages, but still much smaller than the other corpora, which have over 20 unique tags.

However, the difference in results cannot be definitely attributed to tag set size. The IGT tag sets are larger because the goal of descriptive work is to discover fine-grained categories, whereas the Unimorph data use more general categories which

are common for language learning material or general dictionaries. Similar fine-grained distinctions appear in the Penn Treebank tag set and are presumably useful for NLP tasks. Future work could re-tag IGT with more general categories to test how the size and specificity of POS tags on small corpora impact these tasks. This could be fruitful area of research because it might help us predict the usefulness of another linguistic category: the category of morphemes. Morpheme-level categories are similar to POS tags but tagged for individual morphemes. Interestingly, morpheme categories generally take higher priority than word-level tags in documentary and descriptive linguists and are therefore more often available in field data.

Consistency of annotation may be significant. It is likely that the POS tags in the UNIMORPH data were added carefully and correctly, but the field data were likely tagged as the lexical categories were being discovered and described. The differences in results between the two data sets may be due to these factors, but the differences are not huge. So it seems possible that the effect of POS tags may be similar no matter how the POS tags are added. A different approach to POS-tagging, such as training with context might affect results. This possibility points to many future useful experiments. We believe there may be many unresolved issues related to the way the POS tags were added or which POS tags were used. One auxiliary task would be to project POS tags from the target language of the translated sentences that are usually available in IGT even before morpheme segmentation and glosses. Also, metrics for annotation quality could be devised so that its impact is better understood. Linguists need to know as they start annotation how best to perform their earliest analysis and annotation so that they gain optimal benefit from automated help later.

Finally, although a consistent impact by POS tags cannot be seen on morphological learning across all corpora, some corpora did show a more or less consistent impact from the presence or absence of POS tags. Sometimes better results were achieved by removing POS tags, sometimes by adding them. Reinflection in Adyghe and the "clean" version of Lezgi data tend to improve when POS tags are removed while Persian, Russian, and the noisy version of Natügu generally have more accurate results when POS tags are available. In segmentation and glossing, Alas and Lamkang show in some settings nearly .1 points difference when POS tags are added and removed, respectively. With these trends, a more interesting question for these corpora becomes "When are POS tags helpful?" and this should be explored further.

8 Conclusion

We conclude that the presence or absence of POS tags does not have a significant impact on two morphological learning tasks: segmentation and glossing, or reinflection. No clear advantage is gained or lost from POS-tagging on low-resource data. In segmentation and glossing, the greatest average difference is a loss of .09 F_1-score when a large POS tag set is added to a small field corpus. In reinflection, the overall tendency, though slight, is that accuracy decreases when POS tags are added. The greatest average difference is 1.2 percentage points of accuracy for published data, 2.2 points for unpublished "clean" data, and 5 points for unpublished noisy data.

We hypothesize that POS tags do not have a significant impact on these tasks because the information provided by POS tags is implicitly learned. These are, of course, not the only two tasks where POS tags could be leveraged for low-resource languages so we cannot make a definitive statement regarding the impact of POS tags in other NLP tasks with low-resource languages, particularly ones that more syntactic or semantic in nature. Further methodical research needs to be done in order to produce a definitive analysis. However, it does bring into question whether the development of POS taggers and POS tagging should be prioritized less.

Future work should explore how other tasks are impacted by POS tags. The results might influence workflow priorities for documentary and descriptive linguists who want to receive benefit from, or give it to, NLP. When a sophisticated POS tag set and POS taggers are available for a language, leveraging POS tags is trivial. However, as NLP expands into a broader range of languages, the usefulness of POS tags may become an important question because documentary and descriptive linguistics does not currently place a high priority on lexical categories. Discovering a language's lexical categories requires a detailed understanding of the language's syntax—something linguists do not always possess in the early stages of describing a new language.

References

Farrell Ackerman, James P Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. *Analogy in grammar: Form and acquisition*, 54:81.

Roee Aharoni and Yoav Goldberg. 2017. Morphological Inflection Generation with Hard Monotonic Attention. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1:2004–2015.

Malin Ahlberg, Markus Forsberg, and Mans Hulden. 2014. Semi-supervised learning of morphological paradigms and lexicons. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 569–578, Gothenburg, Sweden. Association for Computational Linguistics.

Malin Ahlberg, Markus Forsberg, and Mans Hulden. 2015. Paradigm classification in supervised learning of morphology. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1024–1029, Denver, Colorado. Association for Computational Linguistics.

Antonios Anastasopoulos. 2019. *Computational Tools for Endangered Language Documentation*. Ph.D. thesis, University Of Notre Dame.

Antonios Anastasopoulos and Graham Neubig. 2019. Pushing the limits of low-resource morphological inflection. 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 984–996, Hong Kong, China. Association for Computational Linguistics.

Jason Baldridge and Dan Garrette. 2013. Learning a Part-of-Speech Tagger from Two Hours of Annotation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-13)*.

Steven Bird and David Chiang. 2012. Machine Translation for Language Preservation. In *Proceedings of COLING 2012*, pages 125–134, Mumbai.

Shobhana Lakshmi Chelliah. 1997. *A Grammar of Meithei*, volume 17 of *Mouton Grammar Library*. De Gruyter Mouton, Berlin, Boston. Publication Title: A Grammar of Meithei.

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. 2018a. 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, Arya D. McCarthy, Katharina Kann, Sebastian Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden. 2018b. 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*, 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.

Christopher Cox. 2010. Probabilistic tagging of minority language data: a case study using Qtag. In S.Th. Gries, S. Wulff, and M. Davies, editors, *Corpus linguistic applications: current studies, new directions.*, pages 213–231. Rodopi, Amsterdam:.

Mathias Creutz and Krista Lagus. 2002. Unsupervised discovery of morphemes. In *Proceedings of the ACL-02 workshop on Morphological and phonological learning-Volume 6*, pages 21–30, Philadelphia, PA. Association for Computational Linguistics.

William Croft. 2000. Parts of speech as language universals and as language-particular categories. In Petra M. Vogel and Bernard Comrie, editors, *Approaches to the Typology of Word Classes*, volume 23 of *Empirical Approaches to Language Typology [EALT]*, pages 65–102. De Gruyter Mouton.

Guy De Pauw. 2012. Resource-Light Bantu Part-of-Speech Tagging. In *Proceedings of the workshop on Language technology for normalisation of less-resourced languages (SALTMIL8/AfLaT2012)*, Istanbul, Turkey. European Language Resources Association (ELRA).

Charles Donet. 2014. The Importance of Verb Salience in the Followability of Lezgi Oral Narratives. Master's thesis, Graduate Institute of Applied Linguistics, Dallas, TX.

Long Duong. 2017. *Natural language processing for resource-poor languages*. Ph.D. thesis, University of Melbourne, Melbourne, Australia.

Greg Durrett and John DeNero. 2013. Supervised Learning of Complete Morphological Paradigms. In *Proceedings of NAACL-HLT*, Atlanta, Georgia. Association for Computational Linguistics.

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

Ramy Eskander, Francesca Callejas, Elizabeth Nichols, Judith Klavans, and Smaranda Muresan. 2020a. MorphAGram, Evaluation and Framework for Unsupervised Morphological Segmentation. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7112–7122, Marseille, France. European Language Resources Association.

Ramy Eskander, Smaranda Muresan, and Michael Collins. 2020b. Unsupervised Cross-Lingual Part-of-Speech Tagging for Truly Low-Resource Scenarios. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 4820–4831. Association for Computational Linguistic.

David Gil. 2005. Word order without syntactic categories: How Riau Indonesian does it. In Andrew Carnie, Sheila Ann Dooley, and Heidi Harley, editors, *Verb First: On the Syntax of Verb-initial Languages*, volume 73 of *Linguistik Aktuell/Linguistics Today*, pages 243–264. John Benjamins Publishing. Google-Books-ID: StIEpRsmDbIC.

John Goldsmith. 2001. Unsupervised learning of the morphology of a natural language. *Computational linguistics*, 27(2):153–198.

Zellig S. Harris. 1970. From Phoneme to Morpheme. In *Papers in Structural and Transformational Linguistics*, pages 32–67. Springer Netherlands, Dordrecht.

Gary Holton, Kavon Hooshiar, and Nicholas Thieberger. 2017. Developing collection management tools to create more robust and reliable linguistic data. In *2nd Workshop on Computational Methods for Endangered Languages*.

Max Planck Institute. 2008. The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses.

Katharina Kann, Ryan Cotterell, and Hinrich Schütze. 2017. One-Shot Neural Cross-Lingual Transfer for Paradigm Completion. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1993–2003, Vancouver, Canada. Association for Computational Linguistics.

Katharina Kann, Jesus Manuel Mager Hois, Ivan Vladimir Meza-Ruiz, and Hinrich Schütze. 2018. Fortification of neural morphological segmentation models for polysynthetic minimal-resource languages. 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 47–57, New Orleans, Louisiana. Association for Computational Linguistics.

Katharina Kann and Hinrich Schütze. 2016. Single-Model Encoder-Decoder with Explicit Morphological Representation for Reinflection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 555–560, Berlin, Germany. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Steven Krauwer. 2003. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. In *Proceedings of SPECOM 2003*, pages 8–15.

Julian Kupiec. 1992. Robust part-of-speech tagging using a hidden Markov model. *Computer Speech & Language*, 6(3):225–242.

Ling Liu. 2021. Computational morphology with neural network approaches. *arXiv preprint arXiv:2105.09404*.

Ling Liu and Mans Hulden. 2017. Evaluation of finite state morphological analyzers based on paradigm extraction from Wiktionary. In *Proceedings of the 13th International Conference on Finite State Methods and Natural Language Processing (FSMNLP 2017)*, pages 69–74, Umeå, Sweden. Association for Computational Linguistics.

Ling Liu and Mans Hulden. 2020a. Analogy models for neural word inflection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2861–2878, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Ling Liu and Mans Hulden. 2020b. 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.

Ling Liu and Lingshuang Jack Mao. 2016. Morphological reinflection with conditional random fields and unsupervised features. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 36–40, Berlin, Germany. Association for Computational Linguistics.

Manuel Mager, Özlem Çetinoğlu, and Katharina Kann. 2020. Tackling the low-resource challenge for canonical segmentation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5237–5250, Online. Association for Computational Linguistics.

Peter Makarov and Simon Clematide. 2018. UZH at CoNLL–SIGMORPHON 2018 Shared Task on Universal Morphological Reinflection. In *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 69–75, Brussels. Association for Computational Linguistics.

Robert Malouf. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology*, 27(4):431–458.

Angelina McMillan-Major. 2020. Automating Gloss Generation in Interlinear Glossed Text. In *Proceedings of the Society for Computation in Linguistics*, volume 3.

Alice Millour and Karën Fort. 2019. Unsupervised Data Augmentation for Less-Resourced Languages with no Standardized Spelling. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 776–784, Varna, Bulgaria. INCOMA Ltd.

Sarah Moeller and Mans Hulden. 2021. Integrating Automated Segmentation and Glossing into Documentary and Descriptive Linguistics. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1, pages 86–95, Honolulu, HI. Association for Computational Linguistics.

Sarah Moeller, Ling Liu, Changbing Yang, Katharina Kann, and Mans Hulden. 2020. IGT2P: From Interlinear Glossed Texts to Paradigms. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processi*, pages 5251–5262. Association for Computational Linguistics.

Christopher Moseley, editor. 2010. *Atlas of the World's Languages in Danger*, third edition. UNESCO Publishing, Paris.

Garrett Nicolai, Colin Cherry, and Grzegorz Kondrak. 2015. Inflection Generation as Discriminative String Transduction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 922–931, Denver, Colorado. Association for Computational Linguistics.

Åshild Næss and Brenda H. Boerger. 2008. Reefs–Santa Cruz as Oceanic: Evidence from the Verb Complex. *Oceanic Linguistics*, 47(1):185–212. Publisher: University of Hawai'i Press.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48–53, Minneapolis, Minnesota. Association for Computational Linguistics.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A Universal Part-of-Speech Tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2089–2096, Istanbul, Turkey. European Language Resources Association (ELRA).

Hoifung Poon, Colin Cherry, and Kristina Toutanova. 2009. Unsupervised morphological segmentation with log-linear models. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 209–217. Association for Computational Linguistics.

Gisa Rauh. 2010. *Syntactic Categories: Their Identification and Description in Linguistic Theories*. Oxford University Press.

Gisa Rauh, Jens Fleischhauer, Anja Latrouite, and Rainer Osswald. 2016. Linguistic Categories and the Syntax-Semantics Interface: Evaluating Competing Approaches. In *Explorations of the Syntax-Semantics Interface*, pages 15–55. Düsseldorf.

Teemu Ruokolainen, Oskar Kohonen, Sami Virpioja, and Mikko Kurimo. 2014. Painless semi-supervised morphological segmentation using conditional random fields. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 84–89, Gothenburg, Sweden. Association for Computational Linguistics.

Tanja Samardzic, Robert Schikowski, and Sabine Stoll. 2015. Automatic interlinear glossing as two-level sequence classification. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pages 68–72, Beijing, China. Association for Computational Linguistics.

Frank Seifart, Nicholas Evans, Harald Hammarström, and Stephen C. Levinson. 2018. Language documentation twenty-five years on. *Language*, 94(4):e324–e345.

Abhishek Sharma, Ganesh Katrapati, and Dipti Misra Sharma. 2018. IIT(BHU)–IIITH at CoNLL–SIGMORPHON 2018 Shared Task on Universal Morphological Reinflection. In *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 105–111, Brussels. Association for Computational Linguistics.

Miikka Silfverberg and Mans Hulden. 2018. An encoder-decoder approach to the paradigm cell filling problem. In *Proceedings of the 2018 Conference*

on Empirical Methods in Natural Language Processing, pages 2883–2889, Brussels, Belgium. Association for Computational Linguistics.

Miikka Silfverberg, Ling Liu, and Mans Hulden. 2018. A computational model for the linguistic notion of morphological paradigm. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1615–1626, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Miikka Silfverberg, Adam Wiemerslage, Ling Liu, and Lingshuang Jack Mao. 2017. Data augmentation for morphological reinflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 90–99, Vancouver. Association for Computational Linguistics.

Gary F. Simons and M. Paul Lewis. 2013. The world's languages in crisis: A 20-year update. In Elena Mihas, Bernard Perley, Gabriel Rei-Doval, and Kathleen Wheatley, editors, *Responses to Language Endangerment: In honor of Mickey Noonan. New directions in language documentation and language revitalization*, number 142 in Studies in Language Companion Series, pages 3–20. John Benjamins, Amsterdam.

Alexey Sorokin. 2019. Convolutional neural networks for low-resource morpheme segmentation: baseline or state-of-the-art? In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 154–159, Florence, Italy. Association for Computational Linguistics.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

Ann Taylor, Mitchell Marcus, and Beatrice Santorini. 2003. The Penn Treebank: An Overview. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, Text, Speech and Language Technology, pages 5–22. Springer Netherlands, Dordrecht.

Harimohon Thounaojam and Shobhana L. Chelliah. 2007. The Lamkang language: grammatical sketch, texts and lexicon. *Linguistics of the Tibeto-Burman Area*, 30(1):1–189.

Kristina Toutanova and Mark Johnson. 2008. A Bayesian LDA-based Model for Semi-Supervised Part-of-speech Tagging. In *Proceedings of NIPS*. MIT Press.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017a. Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017b. Attention is All you Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010, Long Beach, California, USA. Curran Associates Inc.

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.

Linlin Wang, Zhu Cao, Yu Xia, and Gerard de Melo. 2016. Morphological Segmentation with Window LSTM Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).

Shijie Wu and Ryan Cotterell. 2019. Exact Hard Monotonic Attention for Character-Level Transduction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1530–1537, Florence, Italy. Association for Computational Linguistics.

Shijie Wu, Ryan Cotterell, and Mans Hulden. 2020. Applying the Transformer to Character-level Transduction. *arXiv:2005.10213 [cs]*.

David Yarowsky and Grace Ngai. 2001. Inducing Multilingual POS Taggers and NP Bracketers via Robust Projection Across Aligned Corpora. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.