The ITML Submission to the IberLEF2023 Shared Task on Guarani-Spanish Code Switching Analysis

Robert Pugh[1,*], Francis Tyers[1]

[1]*Department of Linguistics, Indiana University, Bloomington, IN, U.S.A.*

Abstract

This paper describes experiments in the automated analysis of Guarani-Spanish code-switched text as part of a shared task at IberLEF2023. The submission includes results for all three tasks: (1) language identification, (2) named-entity classification, and (3) Spanish code classification. A CRF trained on text features and several neural network approaches using pre-trained multilingual representations are evaluated. We find that fine-tuning the multilingual representations using unlabeled monolingual Guarani data is beneficial for all the three tasks, and that multi-task training achieves the best results for task 2. The systems described here achieved first place in all three tasks. Interestingly, we did not see a performance boost by replacing multilingual BERT with a pre-trained language model that specifically targets indigenous languages of the Americas.

Keywords

Guarani, Spanish, code switching, multilingual BERT, transformers

1. Introduction

Code switching, or the use of two or more languages within a single conversation or utterance, is very common among multilingual individuals [1], and as such the ability to automatically process, parse, and analyze code-switched language is an area of growing interest in the field of Natural Language Processing (NLP). With multiple workshops over the years dedicated to the "Computational processing of linguistic code-switching" [2], research in this area has explored topics such as language identification at both the token and the utterance/sentence level [3, 4, 5, 6, 7], code-switch point prediction [8, 9], and named entity and part-of-speech tagging [10, 11, 12], among others.

As most of the tasks just listed are examples of identifying a sequence of labels for a sequence of input tokens, they can be modeled with sequence labeling models such as Conditional Random Fields [13], which until recently has been the favorite algorithm for tasks such as word-level language identification and part-of-speech tagging [4]. Neural sequence models, such as Long Short Term Memory networks (LSTMs) [14], have also proven to be quite effective [15]. More recently, since the advent of the Transformer architecture [16], large pre-trained language models such as BERT [17] have been shown to offer a powerful solution to a number of NLP problems including sequential word-based labeling. In particular, multilingual pre-trained

IberLEF 2023, September 2023, Jaén, Spain

*Corresponding author.

pughrob@iu.edu (R. Pugh); ftyers@iu.edu (F. Tyers)

https://robertpugh.me (R. Pugh); https://cl.indiana.edu/~ftyers (F. Tyers)

models such as multilingual BERT (mBERT), trained on 104 languages, has been shown to be effective at a number of cross-lingual learning scenarios, and can be extended to produce reasonable performance even on languages it was not trained on [18, 19]. The success of these pre-trained multilingual models, however, is not uniform across languages and tasks, and performance for a number of low-resource languages does not necessarily benefit from the use of these models [20, 21]. In the experiments below, we primarily focus on the effectiveness of leveraging representations from pre-trained multilingual models on a set of tasks analyzing code switching between a high- (Spanish) and low-resource (Guarani) language.

We describe experiments and results for all three tasks of the Guarani-Spanish Code Switching Analysis [22] (The team name associated with the submissions described in this paper is `ITML` (Inclusive technologies for marginalised languages), and the user name that appears on the leaderboard is `pughrob`) at IberLEF2023 [23], namely (1) word-level language identification, which involves classifying each token in the input as one of either Spanish, Guarani, mixed, foreign, named entity, or other (e.g. punctuation and emojis), (2) named-entity classification, i.e. the further classification of named entities as a person, organization, or location, and (3) Spanish code classification, which involves determining whether a Spanish word or set of words constitute a complete code-switch, or are loan words that maintain the syntactic patterning of Guarani.

The remainder of the paper is laid out as follows. In Section 2 we give a brief introduction to Guarani; in Section 3 we describe the datasets used; in Section 4 we describe the approaches we took; and finally in Sections 5 and 6 we give results and concluding remarks respectively.

2. Guarani

Guarani is a language of the Tupi-Guarani family spoken by around 6.5 Million people, mostly in Paraguay, Argentina, Bolivia, and Brazil [24]. Along with Spanish, Guarani is an official language of Paraguay.

Guarani is an agglutinative and concatenative language. Agglutination implies many morphemes per surface form, with each morpheme having a single meaning (lexical or grammatical), in contrast to fusional languages. The concatenation of morphemes in Guarani words evokes phonological processes on the morphological boundaries. The morphology of the language has both derivational and inflectional affixes; it uses suffixes, prefixes, circumfixes and incorporation for word formation.

With high levels of bilingualism and language contact, code switching is very common, with a particular name, *jopara*, used to refer to use of Guarani that includes a substantial number of loans from Spanish [25, 26]. Guarani that does not contain unadapted loans is often referred to as *guaraniete*. The systems presented in this paper deal primarily with the automated processing of text produced in the former, highly mixed fashion.

(1) *¿Mo'o opyta baño?*
Where is bathroom?
GN GN ES
"Where is the bathroom?"

In (1) there is an example of code switching where the first two words are in Guarani, while the third is in Spanish. In this example, the word *baño* 'bathroom' appears alone, but Spanish words can also appear with Guarani morphology. For example in (2) from [27] the Spanish verb *socorrer* 'help' is inflected with two Guarani affixes *o-* the third person agreement morph and *-ta* the future tense morph.

(2) *¿Máva piko o socorréta ichupe?*
Who QST SG3 help-FUT him/her?
GN GN GN MIX GN
"Who will help her?"

3. Data

The dataset, provided by the shared task organizers, consists of a total of 1,500 sentences, split into a train/dev/test split of 1,140/180/180, in CONLL-U format. The labels for each of the three tasks are combined into a single tag. Each token has a language identification label, one of *es* (Spanish), *gn* (Guarani), *ne* (Named Entity), *mix* (combination of Guarani and Spanish, such as a Spanish root with Guarani morphology), *foreign* (neither Spanish nor Guarani), or *other* (e.g. punctuation, url, or emoji). Each token tagged as *ne* has an additional label corresponding to the named entity classification shared task, either *per* (person), *loc* (location), or *org* (organization). These labels follow the BIO tagging schema. For tokens tagged as *es*, a subset of them are further tagged with one of two "Spanish code classifications", which describe the extent to which spans of Spanish words are loans that behave syntactically more like Guarani words ("unadapted loan" *ul*), or constitute a shift to Spanish, including Spanish syntax ("change in code" *cc*). Importantly, while in theory every span of words labeled as *es* belongs to one of these two categories, only a subset of the spans were labeled in the data. Figure 1 shows an example sentence from the training data.

```
#train4: Atyha salón Mangoré Gobernación Misiones-pe .
1       Atyha   gn
2       salón   es-b-ul
3       Mangoré ne-b-per
4       Gobernación     ne-b-org
5       Misiones-pe     ne-i-org
6       .       other
```

Figure 1: An example from the training data in CONLL-U format. The labels for all three tasks are combined into a single label column. *gn*="Guarani"; *es*="Spanish"; *ne*="Named entity"; *ul*="Unadapted Loan"; *per*="Person"; *org*="Organization". See [22] for a more in-depth description of the data.

4. Experiments

We experimented with 5 approaches to modeling the three tasks: a CRF model trained on naive textual features, and 4 neural network systems that fine-tuned the representations of multilingual pre-trained transformer models for the three tasks. For the transformer-based models, we used the MaChAmp toolkit [28].

4.1. CRF

As a way to benchmark the performance of relatively simple and computationally-inexpensive model architecture, we trained a Conditional Random Field using the same set of naive textual features for all three tasks. For each word in the input, we use the lower-cased word, and case information (e.g. whether the word is in title case). Following substantial research finding the utility of characters in language identification [29, 30, 31], we also extract a number of character-based features to train on, including prefixes and suffixes up to length 5, and character bigrams and trigrams. For a given word, these features were calculated for itself as well as the immediately-adjacent words. We trained the CRF on these features using `pycrfsuite` (https://github.com/scrapinghub/python-crfsuite), training a separate model for each task.

4.2. Fine-tuning multilingual BERT (mBERT)

The use of pre-trained models, which have learned potentially useful textual representations and can be fine-tuned for a specific task with annotated data, has proven effective across a wide range of tasks in NLP. In particular, multilingual models such as multilingual BERT (mBERT) [17] and XLM and [32] have been shown to perform at or near state of the art results in some “low-resource language" scenarios.

Although Guarani was not included in the mBERT training data, mBERT was trained on Spanish and a number of languages related to Spanish. Thus the resulting representations may still be valuable for dealing with Spanish and Guarani data. We fine-tune the mBERT representations (specifically, the `bert-base-multilingual-cased` model) by adding task-specific decoders for each of the three tasks and updating the mBERT weights at the same time as the decoder weights. For the language identification task, the decoder is a fully-connected layer on each of the tokens output by the transformer model. For the remaining two tasks, which involve span classification, we use a CRF decoder in order to ensure that the output conforms to the BIO tagging schema.

4.3. Multitask learning with mBERT (mBERT-MTT)

Given the relative low data volume for the three tasks, we were interested in investigating whether multitask training, wherein we update the mBERT parameters and the parameters of all three task-specific decoders simultaneously. The motivation for this is the apparent overlap among the three tasks. Both task 2, named-entity classification, and task 3, Spanish code classification, are related to task 1 (language identification), since this involved the identification of both named entities and Spanish words in text. It therefore seems reasonable that updates to the parameters for any of these three tasks could improve performance on another.

4.4. Two-stage fine-tuning of mBERT with unlabeled data (mBERT-2SF)

Since, as mentioned above, Guarani was not included in the mBERT training data, we also tried two-stage fine tuning [33], where mBERT is first fine-tuned on the entirety of Guarani wikipedia using a masked language modeling task. The weights in the resulting language model are then further fine-tunes for each specific task. We extracted a plaintext version of the Guarani Wikipedia from a database dump (Wikipedia only, i.e. excluding wikibooks and wiktionary) using the `WikiExtractor.py` script from the Guampa toolkit [34].

4.5. IndT5: a pre-trained model targeting indigenous languages of the Americas (IndT5)

Since multilingual pre-trained models generally underperform on languages that they were not trained on, [35] attempted to fine-tune one such model on a number of indigenous languages of the America. Using the masked language modeling task, the transformer model was trained on data from 10 indigenous languages of the Americas and Spanish. Subsequent experiments showed modest performance improvements on a machine translation task as a result. We hypothesized that using this model in place of standard mBERT, and further updating its parameters during training (task and language-specific fine-tuning), would result in increased performance.

5. Results

The results of our 5 systems on the three tasks can be found in Tables 1, 2, and 3, respectively.

With the exception of the CRF model, all of our systems out-performed the baseline with respect to the weighted F1 score.

The CRF model did perform better than the baseline on the language identification task, but not on the other two tasks. This fact is likely the result of feature selection, in that we primarily chose features based on the language identification literature, and used the same feature sets for all three tasks. The exceptionally poor performance of this model on task 3, Spanish code classification, is likely due to the fact possible that the CRF model features, which only include a one-word context on either side, do not contain enough information to make syntactic determinations (at least perhaps not without ample training examples, which this task certainly did not have).

The *mBERT*, *mBERT-2SF*, *mBERT-MTT*, and *IndT5* systems all performed relatively similarly, and all would have achieved first place in the shared task for all three tasks. The two-stage fine-tuning approach (*mBERT-2SF* yielded the best results for the language identification and Spanish code classification tasks, whereas the multi-task approach (*mBERT-MTT*) achieved the highest weighted F1 score for the named entity classification task. Since named entity classification is closely related to named entity identification, we suspect that simultaneously training on language identification allows the model to learn patterns from the related task with more labeled data, improving performance. Interestingly, however, we do not observe a similar improvement when using multi-task training on the Spanish code classification task.

System description	Precision	Recall	F1
Baseline			0.733
CRF	0.788	0.804	0.795
mBERT	0.934	0.936	0.935
mBERT-MTT	0.917	0.925	0.917
mBERT-2SF	**0.937**	**0.940**	**0.938**
IndT5	0.930	0.931	0.930

Table 1
Weighted Precision, Recall, and F1 scores on the test data for Task 1, language identification. The baseline for this task selects the most frequent category for each word in the training corpus, and selects *other* if the word is not in the training data. The best-performing model on this task uses 2-stage fine-tuning, by first fine-tuning mBERT on Guarani text with the masked language modeling objective, and subsequently fine-tuning on the task itself. This is the only task for which the CRF model performs better than the baseline. *CRF*=conditional random field with textual features; *mBERT*=fine-tuning multilingual BERT; *mBERT-MTT*=multi-task training, i.e. updating weights for all three tasks simultaneously; *mBERT-2SF*=2-stage fine-tuning of multilingual BERT; *IndT5*=fine-tuning the IndT5 transformer model.

System description	Precision	Recall	F1
Baseline			0.495
CRF	0.537	0.360	0.431
mBERT	0.708	0.680	0.693
mBERT-MTT	0.700	**0.724**	**0.712**
mBERT-2SF	**0.739**	0.670	0.703
IndT5	0.693	0.690	0.691

Table 2
Weighted Precision, Recall, and F1 scores on the test data for Task 2, named entity classification. The baseline for this task chooses the most frequent named entity class from the training data based on the first word in the span, and chooses *per* if it was not seen in the training data. This was the only task that seemed to benefit from multi-task training, though the 2-stage fine-tuning model also performs relatively similarly, and achieving better weighted Precision. We did not submit the results of *mBERT-MTT* to the shared task in time for the official evaluation, thus our official submission includes only the results of *mBERT-2SF*. *CRF*=conditional random field with textual features; *mBERT*=fine-tuning multilingual BERT; *mBERT-MTT*=multi-task training, i.e. updating weights for all three tasks simultaneously; *mBERT-2SF*=2-stage fine-tuning of multilingual BERT; *IndT5*=fine-tuning the IndT5 transformer model.

Curiously, although we see improved performance when fine-tuning on Guarani text first, the Ind5T model, which was fine-tuned in the same way (using masked language modeling) specifically on indigenous languages of the Americas including Guarani, consistently underperforms the mBERT-based models (with the exception of the language identification task, where it only out-performs the *mBERT-MTT* model). We suspect that this is due to two factors. First, the languages used for fine-tuning the IndT5 model, despite all being from the Americas, despite a shared contact with Latin-American Spanish, do not have significant genetic or typological overlap. Second, the dataset sizes used in fine-tuning are small with respect to large language

model training. We hypothesize that a similar approach would be more effective if either (1) the languages selected for fine-tuning were more similar (e.g. a model fine-tuned exclusively on languages in long-standing areal contact such as those of Mesoamerica, or members of the same language family), or (2) there were significantly more data for each language during fine-tuning.

System description	Precision	Recall	F1
Baseline			0.220
CRF	0.375	0.015	0.029
mBERT	0.504	0.292	0.370
mBERT-MTT	0.439	0.302	0.358
mBERT-2SF	**0.528**	**0.322**	**0.400**
IndT5	0.470	0.233	0.311

Table 3
Weighted Precision, Recall, and F1 scores on the test data for Task 3, Spanish code classification. The baseline for this task chooses the most frequent class from the training data based on the first word of the span, and chooses *cc* if not seen in the training data. Here, the 2-stage fine-tuning model again achieves the highest weighted F1 score. *CRF*=conditional random field with textual features; *mBERT*=fine-tuning multilingual BERT; *mBERT-MTT*=multi-task training, i.e. updating weights for all three tasks simultaneously; *mBERT-2SF*=2-stage fine-tuning of multilingual BERT; *IndT5*=fine-tuning the IndT5 transformer model.

6. Conclusion

We have presented results of a CRF and a number of neural network sequence models using pre-trained multilingual transformers for analyzing Guarani-Spanish code-switched texts. Our results suggest that using a pre-trained multilingual model such as mBERT and using monolingual data to fine-tune it before training it on the task of interest, can be effective for languages that the model was not trained on, but for which a substantial volume of text data exists. In future work, we are interested in expanding this approach by leveraging the output of a morphological analyzer (such as [36] for Guarani) to fine-tune on the task of morphological analysis and/or generation.

We also found that IndT5, a model trained specifically to handle indigenous languages of the Americas and which was trained on Guarani, generally under-performed the other models in our experiments. We suspect this is due to the relatively low volume of data available for the 10 languages it was trained on compounded by the fact that the languages do not have enough typological similarity for them to learn much from one another (without significantly larger datasets). The code and configuration files are provided to facilitate replicability of the experiments described here (https://github.com/Lguyogiro/es-gn-analysis).

References

[1] P. Gardner-Chloros, Code-switching, Cambridge university press, 2009.

[2] T. Solorio, S. Chen, A. W. Black, M. Diab, S. Sitaram, V. Soto, E. Yilmaz, A. Srinivasan, Proceedings of the fifth workshop on computational approaches to linguistic code-switching, in: Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching, 2021.

[3] T. Solorio, E. Blair, S. Maharjan, S. Bethard, M. Diab, M. Ghoneim, A. Hawwari, F. AlGhamdi, J. Hirschberg, A. Chang, P. Fung, Overview for the first shared task on language identification in code-switched data, in: Proceedings of the First Workshop on Computational Approaches to Code Switching, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 62–72. URL: https://aclanthology.org/W14-3907. doi:10.3115/v1/W14-3907.

[4] G. Molina, F. AlGhamdi, M. Ghoneim, A. Hawwari, N. Rey-Villamizar, M. Diab, T. Solorio, Overview for the second shared task on language identification in code-switched data, in: Proceedings of the Second Workshop on Computational Approaches to Code Switching, Association for Computational Linguistics, Austin, Texas, 2016, pp. 40–49. URL: https://aclanthology.org/W16-5805. doi:10.18653/v1/W16-5805.

[5] V. Ramanarayanan, R. Pugh, Automatic token and turn level language identification for code-switched text dialog: An analysis across language pairs and corpora, in: Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, 2018, pp. 80–88.

[6] J. R. I. Smith, Sinhala-English language detection in code-mixed data, Ph.D. thesis, 2020.

[7] A. Minocha, F. Tyers, Subsegmental language detection in Celtic language text, in: Proceedings of the First Celtic Language Technology Workshop, Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 76–80. URL: https://aclanthology.org/W14-4612. doi:10.3115/v1/W14-4612.

[8] H. Elfardy, M. Al-Badrashiny, M. Diab, Code switch point detection in arabic, in: Natural Language Processing and Information Systems: 18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013, Salford, UK, June 19-21, 2013. Proceedings 18, Springer, 2013, pp. 412–416.

[9] T. Solorio, Y. Liu, Learning to predict code-switching points, in: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Honolulu, Hawaii, 2008, pp. 973–981. URL: https://aclanthology.org/D08-1102.

[10] S. Ghosh, S. Ghosh, D. Das, Part-of-speech tagging of code-mixed social media text, in: Proceedings of the second workshop on computational approaches to code switching, 2016, pp. 90–97.

[11] T. Solorio, Y. Liu, Part-of-Speech tagging for English-Spanish code-switched text, in: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Honolulu, Hawaii, 2008, pp. 1051–1060. URL: https://aclanthology.org/D08-1110.

[12] G. Aguilar, T. Solorio, From English to code-switching: Transfer learning with strong morphological clues, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8033–8044. URL: https://aclanthology.org/2020.acl-main.716. doi:10.18653/v1/2020.acl-main.716.

[13] J. Lafferty, A. McCallum, F. C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001).

[14] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.
[15] Y. Samih, S. Maharjan, M. Attia, L. Kallmeyer, T. Solorio, Multilingual code-switching identification via LSTM recurrent neural networks, in: Proceedings of the Second Workshop on Computational Approaches to Code Switching, Association for Computational Linguistics, Austin, Texas, 2016, pp. 50–59. URL: https://aclanthology.org/W16-5806. doi:10.18653/v1/W16-5806.
[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[17] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.
[18] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual bert?, arXiv preprint arXiv:1906.01502 (2019).
[19] Z. Wang, S. Mayhew, D. Roth, et al., Extending multilingual bert to low-resource languages, arXiv preprint arXiv:2004.13640 (2020).
[20] A. Virtanen, J. Kanerva, R. Ilo, J. Luoma, J. Luotolahti, T. Salakoski, F. Ginter, S. Pyysalo, Multilingual is not enough: Bert for finnish, arXiv preprint arXiv:1912.07076 (2019).
[21] S. Wu, M. Dredze, Are all languages created equal in multilingual BERT?, in: Proceedings of the 5th Workshop on Representation Learning for NLP, Association for Computational Linguistics, Online, 2020, pp. 120–130. URL: https://aclanthology.org/2020.repl4nlp-1.16. doi:10.18653/v1/2020.repl4nlp-1.16.
[22] L. Chiruzzo, M. Agüero-Torales, G. Giménez-Lugo, A. Alvarez, Y. Rodríguez, S. Góngora, T. Solorio, Overview of GUA-SPA at IberLEF 2023: Guarani-Spanish Code-Switching Analysis, Procesamiento del Lenguaje Natural 71 (2023).
[23] S. M. Jiménez-Zafra, F. Rangel, M. Montes-y Gómez, Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org, 2023.
[24] R. O. Collin, Ethnologue, Ethnopolitics 9 (2010) 425–432.
[25] B. Estigarribia, Guaraní-Spanish Jopara Mixing in a Paraguayan Novel: Does it Reflect a Third Language, a Language Variety, or True Codeswitching?, Journal of Language Contact 8 (2015) 183 – 222. URL: https://brill.com/view/journals/jlc/8/2/article-p183_2.xml. doi:https://doi.org/10.1163/19552629-00802002.
[26] B. Estigarribia, Insertion and Backflagging as Mixing Strategies Underlying Guarani-Spanish Mixed Words, Brill, Leiden, The Netherlands, 2017, pp. 315 – 347. URL: https://brill.com/view/book/9789004322578/B9789004322578_011.xml. doi:https://doi.org/10.1163/9789004322578_011.
[27] M. Ayala de Michelagnoli, Ramona quebranto, el habla desoída, Comuneros PEN Club, Asunción, 1989.
[28] R. van der Goot, A. Üstün, A. Ramponi, I. Sharaf, B. Plank, Massive choice, ample tasks (MaChAmp): A toolkit for multi-task learning in NLP, in: Proceedings of the

16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2021, pp. 176–197. URL: https://aclanthology.org/2021.eacl-demos.22. doi:10.18653/v1/2021.eacl-demos.22.

[29] T. Dunning, Statistical identification of language (1996).

[30] J. D. Zamora, A. F. Bruzón, R. O. Bueno, Tweets Language Identification using Feature Weighting., in: TweetLID@ SEPLN, Citeseer, 2014, pp. 30–34.

[31] P. Lamabam, A short review on the language identification systems of social media post, International Journal of Applied Engineering Research 13 (2018) 10339–10342.

[32] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747. doi:10.18653/v1/2020.acl-main.747.

[33] D. Kondratyuk, Cross-lingual lemmatization and morphology tagging with two-stage multilingual BERT fine-tuning, in: Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology, Association for Computational Linguistics, Florence, Italy, 2019, pp. 12–18. URL: https://aclanthology.org/W19-4203. doi:10.18653/v1/W19-4203.

[34] A. Rudnick, T. Skidmore, A. Samaniego, M. Gasser, Guampa: a toolkit for collaborative translation., in: LREC, 2014, pp. 1659–1663.

[35] E. M. B. Nagoudi, W.-R. Chen, M. Abdul-Mageed, H. Cavusoglu, IndT5: A text-to-text transformer for 10 indigenous languages, in: Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas, Association for Computational Linguistics, Online, 2021, pp. 265–271. URL: https://aclanthology.org/2021.americasnlp-1.30. doi:10.18653/v1/2021.americasnlp-1.30.

[36] A. Kuznetsova, F. Tyers, A finite-state morphological analyser for Paraguayan Guaraní, in: Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas, Association for Computational Linguistics, Online, 2021, pp. 81–89. URL: https://aclanthology.org/2021.americasnlp-1.9. doi:10.18653/v1/2021.americasnlp-1.9.