Overview of GUA-SPA at IberLEF 2023: Guarani-Spanish Code Switching Analysis

Overview de GUA-SPA en IberLEF 2023: Análisis de Cambio de Código en Guaraní-Español

Luis Chiruzzo,[1] **Marvin Agüero-Torales,**[2,3] **Gustavo Giménez-Lugo,**[4] **Aldo Alvarez,**[5]
Yliana Rodríguez,[1] **Santiago Góngora,**[1] **Thamar Solorio**[6,7]

[1]Universidad de la República, Montevideo, Uruguay
[2]Universidad de Granada, Granada, Spain
[3]Global CoE of Data Intelligence, Fujitsu, Madrid, Spain
[4]Universidade Tecnologica Federal do Paraná, Curitiba, PR, Brazil
[5]Universidad Nacional de Itapúa, Encarnación, Paraguay
[6]University of Houston, Houston, TX, USA
[7]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates
luischir@fing.edu.uy, maguero@correo.ugr.es, gustavogl@utfpr.edu.br, aldo.alvarez@fiuni.edu.py,
yrodriguez@fhuce.edu.uy, sgongora@fing.edu.uy, thamar.solorio@gmail.com

Abstract: We present the first shared task for detecting and analyzing code-switching in Guarani and Spanish, GUA-SPA at IberLEF 2023. The challenge consisted of three tasks: identifying the language of a token, NER, and a novel task of classifying the way a Spanish span is used in the code-switched context. We annotated a corpus of 1500 texts extracted from news articles and tweets, around 25 thousand tokens, with the information for the tasks. Three teams took part in the evaluation phase, obtaining in general good results for Task 1, and more mixed results for Tasks 2 and 3.
Keywords: Code-switching, Guarani, Spanish, NER.

Resumen: Presentamos la primera competencia de detección y análisis de cambio de código en guaraní y español, GUA-SPA en IberLEF 2023. La competencia constó de tres tareas: identificar el idioma de cada token, NER, y una nueva tarea de clasificar la forma en que se usan los segmentos en español en el contexto del cambio de código. Anotamos un corpus de 1500 textos extraídos de artículos de prensa y tweets, alrededor de 25 mil tokens, con la información necesaria para las tareas. Tres equipos participaron de la fase de evaluación, obteniendo en general buenos resultados para la tarea 1, y resultados más dispares para las tareas 2 y 3.
Palabras clave: Cambio de código, Guaraní, Español, NER.

1 Introduction

This paper presents the GUA-SPA shared task for code-switching analysis in Guarani and Spanish at IberLEF 2023 (Jiménez-Zafra, Rangel, and Montes-y Gómez, 2023).

Guarani is a South American indigenous language that belongs to the Tupi-Guarani family. Spanish is a Romance language that belongs to the Indo-European family, and both languages have been in contact in the South American region for about 500 years (Rodríguez Gutiérrez and Núñez Méndez, 2018), resulting in many interesting linguistic varieties with different levels of mixture. Both, Guarani and Spa-

nish, are Paraguay's official languages (Ley de Lenguas 4251, 2010). According to the most recent census in Paraguay,[1] most of the country's population speaks at least some Guarani, and there is a high prevalence of Guarani-Spanish bilingualism in urban areas, while Guarani monolingualism is more limited to rural areas.

Bilingual speakers often make use of their two languages at the same time, mixing them in different ways, in a linguistic phenomenon called code-switching (Joshi, 1982). This phe-

---

[1]https://censo2022.ine.gov.py/noticias-actividades/21-de-febrero-dia-internacional-de-la-lengua-materna/

Luis Chiruzzo, Marvin Agüero-Torales, Gustavo Giménez-Lugo, Aldo Alvarez, Yliana Rodríguez, Santiago Góngora, Thamar Solorio

nomenon is very frequent in situations where two or more languages are in contact, for example, in linguistic borders, or in countries with large immigrant populations. In Paraguay, this has resulted in several identified language varieties that combine Guarani and Spanish (Kallfell, 2016). Two of the most common varieties of Guarani spoken in Paraguay are called Jopara and Jehe'a (both are Guarani words for describing a mix of things), and they often incorporate many Spanish loans of words or entire phrases, with varying levels of morphosyntactic adaptation to Guarani grammar (Chiruzzo et al., 2022).

In this task, we propose to analyze a set of texts extracted from Paraguayan tweets and news articles, where it is usual to see these Jopara or Jehe'a varieties, and also the use of Spanish sentences that include Guarani loanwords. These texts were annotated by the organisers to create the shared task dataset, which is available online.[2] Using this dataset, we proposed three tasks: language identification (Task 1), NER (Task 2), and a novel classification task for the way Spanish spans are used in the code-switched context (Task 3).

The paper is structured as follows: in Section 2 we mention the previous work on code-switching analysis and Guarani processing; in Section 3 we detail the three tasks that make up this shared task; in Section 4 we present the corpus created for this work; in Section 5 we present the competition, the systems built by the participant teams and their results; and in Section 6 we detail the conclusions of this work.

2 Background

There have been a number of competitions focusing on detection and analysis of code-switching, starting in Solorio et al. (2014) with language identification in code-switched data for some language pairs, including Spanish-English. Later on these competitions started to include more complex tasks in code-switched contexts, such as NER (Aguilar, Kar, and Solorio, 2020) and machine translation into the code-switched languages (Chen et al., 2022). As far as we know, this is the first time Guarani, or any other American indigenous language, is present as

subject of a code-switching analysis competition, the first competition of NER in Guarani, and also the first dataset with this kind of annotations built for an American indigenous language.

However this is not the first work on Guarani identification. Agüero-Torales, Vilares, and López-Herrera (2021) tried to detect tweets in Guarani with heuristics using unigrams and bigrams and other out-of-the-box language identifiers that include Guarani, such as Polyglot (Al-Rfou', Perozzi, and Skiena, 2013), fastText (Joulin et al., 2016) and NLTK's (Bird, Klein, and Loper, 2009) implementation[3] of the method introduced by Cavnar and Trenkle (1994). Góngora, Giossa, and Chiruzzo (2021) used a Naïve Bayes model and a word-based heuristic to detect Guarani tweets. Another big effort in language identification is HeLI-OTS (Jauhiainen, Jauhiainen, and Lindén, 2022), which indeed includes Guarani. Their authors were also participants of this shared task, as we will further describe in Section 5.2.

Guarani is considered a low resource language (Joshi et al., 2020) because, despite having millions of speakers, it does not have many digital resources to work with, its written use online is scarce, and it has been mostly under-researched from the NLP perspective. The situation of this and other American indigenous languages could change in the future as there are now some initiatives to build resources for these languages (Mager et al., 2021), but there is still a long way to go. Spanish, on the other hand, belongs to the small set of very resource-rich languages (Joshi et al., 2020), which is good for this competition as there are many tools for Spanish that could be leveraged to see how well they work in this context.

3 Tasks

The code-switched text detection and analysis tasks proposed in this challenge include the identification of the language used in each span of text, the named entities mentioned in the text, and the way Spanish is used. The challenge was structured as the following three tasks.

---

[2]https://github.com/pln-fing-udelar/gua-spa-2023

[3]https://www.nltk.org/_modules/nltk/classify/textcat.html

3.1 Task 1: Language identification in code-switched data

Given a text (sequence of tokens), label each token of the sequence with one of the following categories:

- **gn**: It is a Guarani token.

- **es**: It is a Spanish token.

- **ne**: It is part of a named entity (either in Guarani or in Spanish).

- **mix**: The token is a mixture between Guarani and Spanish. For example a verb with a Spanish root that has been transformed into Guarani morphology, like: *osuspendeta* (he/she will suspend).

- **foreign**: Used for tokens that are in languages other than Guarani or Spanish.

- **other**: Used for other types of tokens that are invariant to language, like punctuation, emojis and URLs.

Examples:

(1) *che kuerai de pagar 6000 gs. por una llamada de 40 segundos . son aliados del gobierno parece ustedes*
    Could be tagged as:
    che/gn kuerai/gn de/es pagar/es 6000/other gs./other por/es una/es llamada/es de/es 40/other segundos/es ./other son/es aliados/es del/es gobierno/es parece/es ustedes/es

(2) *Ministerio de Salud omombe'u ko'ã káso malaria ojuhúva importado Guinea Ecuatorial guive ha oîma jesareko ohapejokóvo jeipyso .*
    Could be tagged as:
    Ministerio/ne de/ne Salud/ne omombe'u/gn ko'ã/gn káso/mix malaria/es ojuhúva/gn importado/es Guinea/ne Ecuatorial/ne guive/gn ha/gn oîma/gn jesareko/gn ohapejokóvo/gn jeipyso/gn ./other

The metrics for Task 1 are accuracy, weighted precision, weighted recall and weighted F1. The main metric is weighted F1.

We implemented the following baseline for this task: choose the most frequent category for each word in the training corpus, or choose **other** if the word is not in the training corpus.

3.2 Task 2: Named entity classification

Given a text (sequence of tokens), identify the named entities as spans in the text, and classify each one with a category: person, location or organization.
Examples:

(3) [$_{ORG}$ *Ministerio de Salud*] *omombe'u ko'ã káso malaria ojuhúva importado* [$_{LOC}$ *Guinea Ecuatorial*] *guive ha oîma jesareko ohapejokóvo jeipyso* .

(4) [$_{PER}$ *Ministra de Hacienda Lea Giménez*] *he'i oñepromulga léi capitalidad ary 2014*

For this task, the tokens must be marked using the BIO notation, with the following labels: **ne-b-per**, **ne-b-loc**, **ne-b-org**, **ne-i-per**, **ne-i-loc**, **ne-i-org**. The metrics for Task 2 are precision, recall and F1, either labeled or unlabeled. The criterion for finding a named entity is exact match. The main metric is labeled F1.

The baseline for this task is the following: take as an entity span any sequence of tokens labeled as **ne** by the baseline for Task 1, then choose the most frequent category (**per**, **org**, **loc**) for the first word of the sequence in the training corpus. If it is an unknown word, choose the category **per**.

3.3 Task 3: Spanish code classification

Given a text (sequence of tokens), identify spans of text in Spanish and label them in one of these categories:

- change in code (**CC**): the text keeps all the characteristics of Spanish.

- unadapted loan (**UL**): the Spanish text could be partially adapted in some ways to Guarani syntax, but it is not fully merged into Guarani, in particular it does not present orthographic transformations.

Examples:

(5) *che kuerai* [$_{CC}$ *de pagar 6000 gs. por una llamada de 40 segundos . son aliados del gobierno parece ustedes*]

(6) *Okañývo pe Policía Nacional ha orekóva caso omomarandúvo Fiscalía , peteî* [$_{UL}$ *investigación ámbito penal*] .

Luis Chiruzzo, Marvin Agüero-Torales, Gustavo Giménez-Lugo, Aldo Alvarez, Yliana Rodríguez, Santiago Góngora, Thamar Solorio

For this task, the tokens must be marked using the following BIO labels: `es-b-cc`, `es-b-ul`, `es-i-cc`, `es-i-ul`. The metrics for Task 3 are precision, recall and F1, either labeled or unlabeled. The criterion for finding a Spanish span is exact match. The main metric is labeled F1.

The baseline for this task is analogous to the one for Task 2: take as a Spanish span any sequence of tokens labeled as `es` by the baseline for Task 1, then choose the most frequent category (`cc`, `ul`) for the first word of the sequence in the training corpus. If it is an unknown word, choose the category `cc`.

4 Corpus

The corpus of the task consists of 1500 texts, about 25 thousand tokens, split it in 76%-12%-12% for train-dev-test. The data contains sentences extracted from news articles (Chiruzzo et al., 2022; Chiruzzo et al., 2020) and tweets (Agüero-Torales, López-Herrera, and Vilares, 2023; Agüero-Torales, Vilares, and López-Herrera, 2021; Góngora, Giossa, and Chiruzzo, 2022; Ríos, Amarilla, and Lugo, 2014). The characteristics of this corpus are shown in Table 1. As seen in the table, the corpus contains more Guarani tokens than any of the other classes, for example there are on average three Guarani tokens per each Spanish token. Although there are few mixed tokens, around 2% of the corpus, they are a very interesting class to analyze as it presents particular challenges.

	Train	Dev	Test	Total
Texts	1140	180	180	1500
gn	7698	1241	1193	10132
es	5058	812	815	6685
mix	388	52	47	487
ne	2510	414	331	3255
foreign	129	14	8	151
other	3220	456	463	4139
Total tokens	19003	2989	2857	24849
ne-per	663	85	81	829
ne-org	562	95	89	746
ne-loc	240	58	33	331
Total ne spans	1465	238	203	1906
es-ul	636	80	95	811
es-cc	621	70	112	803
Total es spans	1257	150	207	1614

Table 1: Composition of the dataset.

The annotation was done in two phases: A first phase consisted in four annotators annotating a sample of 54 sentences (around 750 tokens). The aim of the first phase was to unify criteria and gauge how hard it was to make the annotations. The four annotators obtained a Fleiss Kappa inter-annotator agreement of 0.836 for Task 1, which indicates very substantial agreement. Only two of the annotators annotated the data for Task 2 during this phase, and we calculated their agreement using average F1, obtaining 0.926 between the two annotators, which we also consider a substantial agreement. We came up with the idea for Task 3 during this pilot annotation, but it was considerably harder to annotate, with many ambiguous cases that were difficult to generalize.

In the final phase of annotation, six annotators fluent in Spanish and with some knowledge of Guarani labeled data for Tasks 1 and 2, but only the three annotators with more knowledge of Guarani annotated the data for Task 3.

After the annotation was over we estimated the annotator agreement in Task 3 in the following way: we took a subsample of 90 texts that contained at least one token in Spanish, 30 for each annotator that participated in the task. The three annotators completed the annotation of the 90 texts, and we compared the annotations as the F1 measure of one annotator against another. The average F1 obtained in this way was 0.758 for the labeled F1 considering exact match which, although not as substantial as the agreement for Task 2, was still quite high.

5 Competition

The competition ran between March 22 and June 7 (2023) on the CodaLab platform[4] (Pavao et al., 2022). A total of 20 users registered to participate, but the number of participants that submitted results was considerably lower: three of them participated both in the development and the evaluation phase, and the other two participated in a single phase.

5.1 Phases

The competition consisted of three phases:

Development phase: from March 22 to May 23. This phase started with the publication of the training and development sets.

---

[4]https://codalab.lisn.upsaclay.fr/competitions/11030

During this phase the participants could submit their predictions for the development set and get the correspondent score for each task. Each participant could make up to 200 submissions. There were 50 submissions.

Evaluation phase: from May 24 to June 6. This phase started with the publication of the test set. In this phase the participants could submit the predictions of their final systems and get the correspondent score for each task. Each participant could make up to 10 submissions. There were 19 submissions.

Post-Evaluation phase: from June 7 onward. This phase started after the end of the competition. The CodaLab page remains available for everyone who wants to test additional systems, download the training, development and test sets, and check the shared-task information.

5.2 Systems Descriptions

Although we had only three participants in the evaluation phase, overall the followed approaches were notoriously diverse. We briefly describe those systems:

The user **aalgarra** submitted results only during the development phase, obtaining good results. The chosen approach is based on finetuning a Spanish RoBERTa-base trained on BNE finetuned for CAPITEL NER dataset (Fandiño et al., 2022), to which a bi-directional LSTM layer is added. The target is obtained by assigning the category to the first token of each word and applying a mask in the loss calculation for the following tokens of each word. The training was performed in two phases: an initial phase with frozen RoBERTa weights, with 30 epochs and a learning rate of $1e-3$, and a second phase training the full network during 2 epochs with a learning rate of $5e-5$.

The Universidade da Coruña team (Muñoz-Ortiz and Vilares, 2023) (user **amunozo**) participated on all three tasks. They implemented their solution as a Multi-task learning problem where pre-trained encoder-decoder models were fine-tuned based on a hard parameter-sharing approach. They introduced three models with a common pre-trained LLM encoder fine-tuned on all three tasks and different decoders for each task. Their best-performing model on Task 1 is single-nllb model (NLLB Team et al., 2022) where the encoder is trained only on Task 1 and the decoder is a softmax on the output layer, while the systems for Tasks 2 and 3 decode the output as Conditional Random Fields (CRF). Their submitted systems use the mtl–beto-gn (Agüero-Torales, López-Herrera, and Vilares, 2023) model for Task 2 and the mtl-nllb model (NLLB Team et al., 2022) for Task 3. A post-processing heuristic is added to ensure the formatting of the final results.

The University of Helsinki team (Jauhiainen, Jauhiainen, and Lindén, 2023) (user **tsjauhia**) participated only on Task 1 obtaining the second place, very close to the winning team (0.0242 F1 points away). They used a system consisting of their own HeLI-OTS language identifier (Jauhiainen, Jauhiainen, and Lindén, 2022) based on the HeLI method (Jauhiainen, Lindén, and Jauhiainen, 2016), that includes Guarani. Besides making some internal configurations to HeLI-OTS, they also added ad hoc rules, such as a text preprocessing pipline for Guarani and pattern-based logic to detect mixed Spanish-Guarani words.

The Inclusive Technologies for Marginalised Languages (ITML) team (Pugh and Tyers, 2023) (user **pughrob**) from Indiana University obtained the first place on every task. The authors evaluate a CRF model trained on text features and several neural network approaches using pre-trained multilingual representations, different fine-tuned BERT and T5 models. The team got the first position in the competition for the three tasks, after finding that fine-tuning the multilingual representations with unlabeled monolingual Guarani data is beneficial for all three tasks, and that multi-task training achieves the best results for Task 2.

5.3 Results

Table 2 shows the results of the development phase, while Table 3 shows the final results of the evaluation phase. We also had an ad-

User	Task 1 Weighted F1	Task 2 Labeled F1	Task 3 Labeled F1
pughrob	**0.9424**	**0.7604**	**0.4370**
aalgarra	0.9304	0.7047	0.2937
tsjauhia	0.9053	-	-
amunozo	0.8221	0.1986	-
baseline	*0.7032*	*0.4218*	*0.2331*

Table 2: Results of the development phase.

User	Task 1	Task 2	Task 3
	Weighted F1	Labeled F1	Labeled F1
pughrob	**0.9381**	**0.7028**	**0.3836**
tsjauhia	0.9139	-	-
amunozo	0.8500	0.4153	0.1939
baseline	*0.7325*	*0.4946*	*0.2195*

Table 3: Results of the evaluation phase.

ditional submission by user **pakapro** for the evaluation phase that just tagged every token as *other*, hence it is not included in the tables.

In general, participants got very good results for Task 1, beating the baseline by a good margin. The categories where the participants got the best performance for Task 1 were `gn`, `es`, and `other`, being the last one a very easy class to predict, containing mostly punctuation signs and URLs. Conversely, the hardest classes to classify were `foreign` and `mix`, which also had the fewest examples. The `mix` category is particularly interesting, as it comprises words that combine Guarani and Spanish content in some way, for example a Spanish verb with some adapted Guarani verbal morphology, or a Spanish noun with a Guarani case marker suffix. There are only 47 examples of `mix` words in the test corpus, and three of those were correctly identified by all three participants: *oñepresenta*, *salud-pe*, and *cartista-pe*. The first one was in the training data, but not the others, so the systems were able to actually generalize at least to these simple cases. Of the three participants, the University of Helsinki team was the only one that reported including particular rules for dealing with `mix` words by analyzing affixes, and they also got the best precision for this class, although the ITML achieved better recall.

The performance for Tasks 2 and 3 was more mixed, with only one team beating the baselines. Task 3 was particularly hard for the automatic systems, and it was also hard for the annotators. That can be seen as the best performance was only 0.384 for labeled F1. Although it is not completely comparable, if we consider the annotators obtained an average F1 of 0.758 for a subsample of 90 texts, we can say that there is still a lot of room for improvement in this task.

6 Conclusions

In this work, we presented the GUA-SPA shared task for Guarani-Spanish code-switching analysis at IberLEF 2023, the first task of its kind that involves an American indigenous language. The competition consisted of three tasks: language identification, NER, and a novel task of detecting the way Spanish is used in the code-switched context. Three participants submitted their predictions for the evaluation phase, with the ITML team submission being the one with the best results across the three tasks. Although performance for Task 1 was generally good for all participants, the results obtained for Tasks 2 and 3 were more mixed. In particular Task 3, predicting if a Spanish span is used as an unadapted loan or as part of a code change, was the most challenging one both for the human annotators and the participants of the competition. It would be interesting to explore this task more in the future, trying to better specify the two classes and how to deal with the ambiguous cases. This could also be explored for other languages as well, as this phenomenon could be common in any code-switched context. We would also like to incorporate more linguistic varieties of Guarani to the task, for example the Bolivian variety, and also include Portuguese as another possible language for the tokens, since it is another widely spoken language in South America, specifically near the areas where Guarani is spoken. We hope this work contributes to spark the interest in code-switching analysis for more American indigenous languages.

References

Agüero-Torales, M., D. Vilares, and A. López-Herrera. 2021. On the logistical difficulties and findings of jopara sentiment analysis. In *Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching*, pages 95–102, Online, June. Association for Computational Linguistics.

Agüero-Torales, M. M., A. G. López-Herrera, and D. Vilares. 2023. Multidimensional affective analysis for low-resource languages: A use case with guarani-spanish code-switching language. *Cognitive Computation*.

Aguilar, G., S. Kar, and T. Solorio. 2020. LinCE: A centralized benchmark for lin-

guistic code-switching evaluation. *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1803–1813.

Al-Rfou', R., B. Perozzi, and S. Skiena. 2013. Polyglot: Distributed word representations for multilingual NLP. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192, Sofia, Bulgaria, August. Association for Computational Linguistics.

Bird, S., E. Klein, and E. Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* .ºʹReilly Media, Inc.".

Cavnar, W. B. and J. M. Trenkle. 1994. N-gram-based text categorization.

Chen, S., G. Aguilar, A. Srinivasan, M. Diab, and T. Solorio. 2022. CALCS 2021 shared task: Machine translation for code-switched data. *arXiv preprint arXiv:2202.09625*.

Chiruzzo, L., P. Amarilla, A. Ríos, and G. Giménez-Lugo. 2020. Development of a guarani-spanish parallel corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2629–2633.

Chiruzzo, L., S. Góngora, A. Alvarez, G. Giménez-Lugo, M. Agüero-Torales, and Y. Rodríguez. 2022. Jojajovai: A Parallel Guarani-Spanish Corpus for MT Benchmarking. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2098–2107.

Fandiño, A. G., J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, and M. Villegas. 2022. Maria: Spanish language models. *Procesamiento del Lenguaje Natural*, 68.

Góngora, S., N. Giossa, and L. Chiruzzo. 2021. Experiments on a guarani corpus of news and social media. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 153–158.

Góngora, S., N. Giossa, and L. Chiruzzo. 2022. Can we use word embeddings for enhancing Guarani-Spanish machine translation? In *Proceedings of the Fifth*

Workshop on the Use of Computational Methods in the Study of Endangered Languages, pages 127–132, Dublin, Ireland, May. Association for Computational Linguistics.

Jauhiainen, T., H. Jauhiainen, and K. Lindén. 2022. HeLI-OTS, off-the-shelf language identifier for text. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3912–3922, Marseille, France, June. European Language Resources Association.

Jauhiainen, T., H. Jauhiainen, and K. Lindén. 2023. Tuning heli-ots for guarani-spanish code switching analysis. In *Procesamiento del Lenguaje Natural 71*. SEPLN.

Jauhiainen, T., K. Lindén, and H. Jauhiainen. 2016. HeLI, a word-based back-koff method for language identification. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pages 153–162, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Jiménez-Zafra, S. M., F. Rangel, and M. Montes-y Gómez. 2023. 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*.

Joshi, A. 1982. Processing of sentences with intra-sentential code-switching. In *Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics*.

Joshi, P., S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv preprint arXiv:2004.09095*.

Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov. 2016. Bag of tricks for efficient text classification.

Kallfell, G. 2016. *¿Cómo hablan los paraguayos con dos lenguas?: gramática del jopa-*

Luis Chiruzzo, Marvin Agüero-Torales, Gustavo Giménez-Lugo, Aldo Alvarez, Yliana Rodríguez, Santiago Góngora, Thamar Solorio

ra. Centro de Estudios Antropológicos de la Universidad Católica (CEADUC).

Ley de Lenguas 4251. 2010. Ley Nº 4251 / DE LENGUAS. In *https://www.bacn.gov.py/leyes-paraguayas/2895/ley-n-4251-de-lenguas*.

Mager, M., A. Oncevay, A. Ebrahimi, J. Ortega, A. R. Gonzales, A. Fan, X. Gutierrez-Vasques, L. Chiruzzo, G. Giménez-Lugo, R. Ramos, et al. 2021. Findings of the AmericasNLP 2021 shared task on open machine translation for indigenous languages of the Americas. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 202–217.

Muñoz-Ortiz, A. and D. Vilares. 2023. Lys a coruña at gua-spa@iberlef2023: Multi-task learning with large language model encoders for guarani-spanish code switching analysis. In *Procesamiento del Lenguaje Natural 71*. SEPLN.

NLLB Team, M. R. Costa-jussà, J. Cross, O. Çelebi, M. Elbayad, K. Heafield, K. Heffernan, E. Kalbassi, J. Lam, D. Licht, J. Maillard, A. Sun, S. Wang, G. Wenzek, A. Youngblood, B. Akula, L. Barrault, G. M. Gonzalez, P. Hansanti, J. Hoffman, S. Jarrett, K. R. Sadagopan, D. Rowe, S. Spruit, C. Tran, P. Andrews, N. F. Ayan, S. Bhosale, S. Edunov, A. Fan, C. Gao, V. Goswami, F. Guzmán, P. Koehn, A. Mourachko, C. Ropers, S. Saleem, H. Schwenk, and J. Wang. 2022. No language left behind: Scaling human-centered machine translation.

Pavao, A., I. Guyon, A.-C. Letournel, X. Baró, H. Escalante, S. Escalera, T. Thomas, and Z. Xu. 2022. Codalab competitions: An open source platform to organize scientific challenges. *Technical report.*

Pugh, R. and F. Tyers. 2023. The itml submission to the iberlef2023 shared task on guarani-spanish code switching analysis. In *Procesamiento del Lenguaje Natural 71*. SEPLN.

Ríos, A. A., P. J. Amarilla, and G. A. G. Lugo. 2014. Sentiment categorization on a creole language with lexicon-based and machine learning techniques. In *2014 Brazilian Conference on Intelligent Systems*, pages 37–43. IEEE.

Rodríguez Gutiérrez, Y. and E. Núñez Méndez. 2018. Language contact and the indigenous languages of Uruguay. *Biculturalism and Spanish in contact: sociolinguistic case studies*, pages 217–238.

Solorio, T., E. Blair, S. Maharjan, S. Bethard, M. Diab, M. Ghoneim, A. Hawwari, F. AlGhamdi, J. Hirschberg, A. Chang, et al. 2014. 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*, pages 62–72.