Multidimensional Affective Analysis for Low-resource Languages: A Use Case with Guarani-Spanish Code-switching Language

Marvin M. Agüero-Torales[1*], Antonio G. López-Herrera[1] and David Vilares[2]

[1*]Department of Computer Science and Artificial Intelligence, University of Granada, Calle Daniel Saucedo Aranda, s/n, Granada, 18071, Granada, Spain.
[2]Department of Computer Science and Information Technologies, Universidade da Coruña, CITIC, Campus de Elviña, s/n, A Coruña, 15008, A Coruña, Spain.

*Corresponding author(s). E-mail(s): maguero@correo.ugr.es;
Contributing authors: lopez-herrera@decsai.ugr.es;
david.vilares@udc.es;

Abstract

Introduction: This paper focuses on text-based affective computing for Jopara, a code-switching language that combines Guarani and Spanish.* **Methods:** First, we collected a dataset of tweets primarily written in Guarani and annotated them for three widely-used dimensions in sentiment analysis: (a) emotion recognition, (b) humor detection, and (c) offensive language identification. Then, we developed several neural network models, including large language models specifically designed for Guarani, and compared their performance against off-the-shelf multilingual and Spanish pre-trained models for the aforementioned dimensions. **Results:** Our experiments show that language models incorporating

*This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s12559-023-10165-0. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/acceptedmanuscript-terms

Guarani during pre-training or pre-fine-tuning consistently achieve the best results, despite limited resources (a single 24GB GPU and only 800K tokens). Notably, even a Guarani BERT model with just two layers of Transformers shows a favorable balance between accuracy and computational power, likely due to the inherent low-resource nature of the task. **Conclusions:** We present a comprehensive overview of corpus creation and model development for low-resource languages like Guarani, particularly in the context of its code-switching with Spanish, resulting in Jopara. Our findings shed light on the challenges and strategies involved in analyzing affective language in such linguistic contexts.

Keywords: natural language processing, sentiment analysis, affective analysis, code-switching, low-resource languages

MSC Classification: 68-02 , 68T50 , 68T07 , 91D30

1 Introduction

Most Latin American countries have indigenous languages, many of them endangered, which coexist with the dominant ones, Spanish and Portuguese being the main ones. In the case of some languages (e.g., Guarani or Nahuatl), the use of indigenous and colonizing languages is intermixing in a profound way. The development and democratization of technologies for indigenous languages have recently gained interest, especially in the context of natural language processing (NLP), however, the results are still scarce and incipient [1–3].

Different factors favor the motivation to develop resources and models for these types of languages, such as for example: (i) the challenge to preserve them in the digital world if no applications are developed, (ii) the risk of speakers abandoning low-resource mother tongues for rich-resource ones if they are unable to take the advantages of language technologies, or (iii) the need to overcome the biases of sociolinguistic analyses that consider mainly privileged languages.

In this paper, we will focus on Guarani. The Guarani language (`grn`),[1] with about 8M speakers, belongs to the Tupi-Guarani family. It has a moderately favorable position in relation to other Latin American indigenous languages: it is an official language in Paraguay and Bolivia, as well as in the province of Corrientes (Argentina). It is an official language of Mercosur, a South American trade block of several countries, and a spoken non-official language in Southwestern Brazil. Guarani is found in code-switching contexts. This is due to its use being closely intertwined with other languages (Spanish, but also English and Portuguese), resulting in Jopara, a code-switching language between Guarani and Spanish [4]. Therefore, we will be referring to both terms

[1] The `grn` is the ISO 639-3 language code for Guarani. Its ISO 639-1 code is `gn`.

throughout the text. Table 1 shows examples written in Jopara and compares them with standard (or pure) Guarani and Spanish.[2]

- *grn*
  - Ikatútapa reikuaa mbohapy kuatiarogue peteĩ aravópe?
  - Nahániri. Che ha'e ndéve nda ñemoarandúiha che akãrasýrõ.
- *jopara*
  - Ikatútapa reaprendé mbohapy páhina una órape?
  - Nahániri. Che ha'e ndéve nda studiáiha che akãrasýrõ.
- *spa*
  - ¿Vas a poder aprenderte tres páginas en una hora?
  - No. Te dije que yo no estudio si me duele la cabeza.
- *eng*
  - Are you going to be able to learn three pages in an hour?
  - No. I told you I don't study if I have a headache.

Fig. 1 Part of a conversation written in Jopara and its equivalent in Guarani (**grn**), Spanish (**spa**), and English (**eng**). Spanish words and lemmas are shown in red in the Jopara example, and in green their correspondence with the other languages. The sample text **grn** is known as 'pure' or 'standard' (sometimes 'academic') Guarani with often a high rate of neologisms [5]. The text **jopara**, which is also sometimes called "jehe'a", i.e., when the mixture has an elaborated adaptation of the Spanish loanwords into Guarani [6], and it is closer to how native speakers of Guarani communicate in their daily life [7, 8].

In this context, sentiment analysis [SA; 9], the analysis of the emotional content of a text, is a popular application of NLP systems [10, 11]. The aim of these systems is to identify the feeling expressed by the writer in a text, rather than the emotion experienced by the reader. For example, since the requirements of different applications lead to defining different subtasks of sentiment, we can name the most usual ones, such as emotion recognition, e.g., assigning the text a label such as 'joy' or 'sad' [12]; or polarity classification, e.g., classifying texts into positive, neutral or negative [13]. Within the context of subjective analysis automation, to the best of our knowledge related work on Guarani is scarce. Yet, developing SA tools and resources for this language would have important implications in low-resource NLP, and more importantly, would contribute to democratizing the use of social analysis language technologies for Guarani speakers.

In this context, our work falls in the intersection between cognitive computation, affective computation, and code-switching, three intertwined research areas. Code-switching is a complex linguistic phenomenon that requires cognitive flexibility and effort [14], involving the ability to switch between different linguistic systems. This occurs on both the part of the sender and the receiver of the message. Thus, developing language technologies with the ability to

[2]Some tokens are a mixture of n-grams of Guarani and Spanish characters, e.g., '*study*' would be '*estudio*' (spa), '*ñemoarandú*' (grn) and '*studiá*' (jopara).

automatically understand such environments contributes directly to the design of new cognitive computation methods. In addition, affective analysis, which refers to the broader concept of analyzing emotions, sentiments and other emotional dimensions such as emotion recognition, mood detection, and subjective assessments, has also been long considered an interesting NLP testbed in the context of cognitive computation [15]. In our case, by examining how emotions, humor, and offensive language are expressed in Guarani/Jopara, we shed light under a unified setup on how to apply cognitive computation mechanisms underlying multidimensional affective processing, in the context of a code-switching language and low-resource modeling [16]. In addition, by relying on artificial neural networks and evaluating large language models for Guarani, we contribute to the development of biologically inspired language architectures that help us develop text analysis methods to automatically understand and process Guarani.

Contribution

A framework for affective analysis of Guarani/Jopara is introduced. First, we present three Twitter corpora annotated according to multiple affective dimensions: emotion recognition, humor detection, and identification of offensive language. Second, we develop different models and study how they perform in this low-resource setup. Specifically, we consider (i) models without pre-training, (ii) pre-trained language models that only have access to Guarani at the fine-tuning phase, and (iii) language models specifically trained on Guarani texts, either from scratch or using pre-initialized with the weights of multilingual models, to what we will be referring here as pre-fine-tuning. Notably, (iii) was trained with limited resources (800K tokens for a 24GB GPU), for a truly low-resource scenario, where the amount of data *and* physical resources are scarce. The goal of this is to test if this approach can be beneficial for these setups, and potentially to other languages. Note that this work extends Chapter 6 of the doctoral thesis of [17]. The diagram summarizing the methodology followed in this work can be seen in Figure 2.

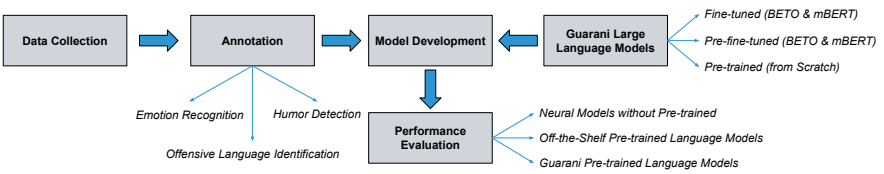

Fig. 2 Summary of the methodology.

The remainder of the paper is structured as follows. Section 2 first presents the related work in the area of sentiment analysis for low-resource languages and then focuses towards explaining the related work on NLP and sentiment

analysis for the specific case of Guarani. Section 3 describes the collected dataset and its annotation for the three proposed categories. Section 4 presents models and their particularities to address the problem at hand. Next, Section 5 discusses the results on the proposed dataset for each of the three annotated dimensions. Finally, we summarize our work in Section 6.

2 Related Work

Low-resource languages are those that lack (labeled or unlabeled) large monolingual or parallel corpora to build specific NLP applications. The following is a review of sentiment analysis (as well as other relevant NLP) research involving low-resource languages in general and then more specifically involving Latin American indigenous languages.

2.1 Sentiment analysis for low-resource languages

Like other NLP tasks, sentiment analysis has focused on rich-resource languages. The main limitation is that, as suggested above, it is hard to have access to digitalized text for low-resource languages even from web sources, and consequently it is difficult to develop sentiment analysis corpora and systems [18]. Yet, with the plethora of social media platforms out there, it is possible to find exceptions. [19] created a corpus of 4K online articles paragraphs and lexica for Estonian (a Uralic language with 1M speakers) and tested both lexicon-based and machine learning approaches for polarity classification. In a related line, [20] proposed a machine translation (MT) and a lexicon-based method for sentiment analysis in Irish (also called Gaelic, an Indo-European Celtic language with 1.7M speakers). In this case, the accuracy of the transliteration of the Irish tweets into English was 6% lower than that achieved by the manually created lexicon. In contrast, [21] proposed a large-scale Twitter dataset for Urdu, an Indo-Aryan language and lingua franca of Pakistan with 230M speakers, for sentiment analysis. The dataset contains 1.1M of tweets and was collected through the Twitter Search API over two months, using emojis to annotate the dataset following [22], where each emoji was assigned a sentiment score and an emotion category. On a smaller scale, [23] collected 1K Urdu political tweets for trinary polarity classification. They also performed an analysis of textual similarities, manifold learning, and part-of-speech (PoS) tagging, and proposed a technique to extract sentiment lexicon from the corpus.

Related to this, [24] introduced the first publicly available large-scale dataset for polarity classification, consisting of tweets written in widely spoken Nigerian languages. The multilingual dataset covers four languages: Hausa, Igbo, Yorùbá (30K annotated tweets were annotated for each language), and Nigerian-Pidgin (14K tweets). It is worth noting that, all of them commonly exhibit code-switching with English. They also evaluated standard multilingual pre-trained language models and transfer strategies on this dataset. They found out that language-specific models, such as AfriBERTa [25], performed the best.

Related work also goes beyond social media posts. For instance, [26] exploited topic modeling techniques to classify sentiment from song lyrics in Manipuri (a Tibeto-Burman language, a lingua franca in Manipur, India, with 1.6M speakers). They extracted topics from 150 lyrics samples labeled as 'happy' or 'sad', being able to obtain the underlying sentiment of lyrics.

2.2 Sentiment analysis for Latin American indigenous languages

In this section, we describe related work in the context of Latin American indigenous languages and SA. [27] and [28] introduced methods and lexica to carry out polarity classification in many languages, including indigenous ones, such as Quechua. They built sentiment lexica propagating from parallel corpus seed words, projecting English to other languages. [29] presented a study on the automatic generation of paraphrases of sentiment predicates for the specific case of Quechua, and also built a dictionary of Quechua sentiment verbs. More related to the content of our work, [30] developed a sentiment analysis corpus from Twitter for Paraguayan Spanish. This corpus has more Spanish words (3,802), compared to the number of Guarani and Jopara ones (80) [30, p. 40, Table II].

2.3 NLP for Guarani and Jopara

This section covers NLP research on Guarani and Jopara languages. For instance, [31] used morphological information to improve a Guarani-Spanish MT system. More particularly, they focused on the analysis of verbs, augmenting morphological features with different versions of the corpus. Along the same line, [2] proposed an MT shared task for Spanish and Guarani, as well as other nine Latin American indigenous languages. [32], first created a Guarani-Spanish parallel set of news (15K sentence pairs), and then performed experiments with neural MT on both directions of the language pair. [5] presented a parallel corpus of 30K Guarani-Spanish text aligned at sentence level for use in MT as a benchmark. Related to MT, [33] created `AmericasNLI`, a dataset that tests natural language inference for 10 Indigenous languages of America (including Guarani). Using this dataset, the authors found out that pre-trained models do not perform well without additional training, and that translation-based models work better than other approaches. On lower-level tasks, [34] presented a finite-state, segmentation-based morphological analyzer for Paraguayan Guarani, which is part of `Apertium`.[3]

As for databases for Guarani, LANGAS [35] offers a database of historical texts, mainly in Guarani (also in Quechua and Tupi), for sociolinguistic studies, with an optional query expansion mechanism to address word variation challenges. Authors outline morphology, modules, and evaluations, providing resources for the language. [36] created a Guarani version of the WordNet database using three different Guarani-Spanish dictionaries with an expansion

[3] A rule-based MT open-source platform, https://github.com/apertium/apertium-grn.

approach to finding Guarani lemmas for concepts in WordNet. The procedure identified lemmas for around 6.5K synsets, including 27% of the base concepts.

On the multimodal side, there are also a few resources available for Guarani. For instance, a large audio dataset called 'Multilingual Spoken Words Corpus' [37, MSWC], covers spoken words in 50 languages (about 6K hours), including Guarani, which is useful for applications such as keyword spotting and spoken term search. As a product, this dataset can be used for voice-enabled consumer devices or call center automation. Also related to this, [38, XLS-R] introduced a large-scale model for speech representation learning in various languages. The model, based on wav2vec 2.0 [39, 40], was trained on spoken audio from 128 languages (about 500K hours publicly available), including Guarani. [41] introduced a generative model to make MT directly between any pair of more than 200 languages. Another generative model proposed by [42] was introduced for language generation.

2.4 Sentiment analysis for Guarani and Jopara

Concerning sentiment analysis, we are aware of a few works. [43] discussed the logistical difficulties of creating resources for truly low-resource languages and more specifically for the case of Jopara. They explained how traditional methods (e.g., collecting texts by keywords), which work for collecting data for rich-resource languages, fail for low-resource setups and specially code-switching setups, where one of the used languages is highly privileged in comparison to the second one, such as the case of Guarani/Jopara. [43] also proposed alternative strategies to collect data for this type of languages and, unlike [30], used them to create the first Guarani-dominant tweets corpus for polarity classification.

Beyond polarity classification, we have no knowledge of prior work on Guarani or Jopara for other text-based affective tasks [44], such as for instance emotion recognition [45, 46], humor detection [47], and offensive and toxic language identification [48]. In general, affect detection has focused on languages for which textual content is easy to collect (see the challenges for collecting Twitter content written in Guarani in the following **S** 3). In this context, although some previous research has developed multilingual approaches [49–51]; for languages such as Guarani and other truly low-resource languages [52, p. 7658, Table 1], it is challenging to develop new NLP corpora and systems. Here, our aim is to fill this gap, providing resources and models for these specific problems.

3 Datasets for text-based affective detection in Jopara

In this section we describe the two challenges that we need to face to create a collection for affective analysis of Guarani tweets: (i) to find tweets that were mostly written in Guarani, avoiding those that are highly contaminated by Spanish, as these languages are closely intertwined, and (ii) finding a sufficient

number of tweets that are relevant to affective tasks such as the ones we addressed in this paper.

Collecting the data

Guarani is an agglutinative language, featuring complex morphology; each word consists of multiple prefixes, stems, and suffixes [53]. On the other hand, Guarani is not included in Twitter's language detection feature, meaning Twitter cannot directly monitor a stream of tweets published in Guarani using its API. To deal with these problems, we queried the Twitter API using specifically Guarani keywords, following [43, JOSA]. With the same idea that [21, 54], who used tweets and emojis as keywords, we used sentiment terms to create our corpora. Note that using emojis in our case is unrealistic due to the poor status of Guarani. We relied on positive and negative words, selecting 763 JOSA's Guarani terms as keywords,[4] and ignoring those that occurred less than three times in their dataset. Usually, these terms are related to the expression of sentiment, mood, and emotion [Figure 2; 55, p. 622]. For instance, a negative term may denote an angry mood, and this in turn may lead to the use of offensive language or jargon by the speaker [56, 57]. It should be noted that the Twitter real-time streamer is only able to handle a restricted set of keywords, therefore we rotated 50 distinct keywords every 3 hours to sample tweets for 6 months. As we will see now, sampling during this large period is required, since most of the collected tweets are highly contaminated with Spanish terms, and they would not be useful for the purpose of this work.

Cleaning the data

With 7.4M tweets collected, we conducted a language detection stage with three different tools that can detect Guarani texts, as a proxy to identify tweets that are Guarani-dominant: (i) `textcat`,[5] (ii) `fastText` [58] and (iii) `polyglot`.[6] Preliminarily, the text was assumed to be Guarani if it was classified as such by at least one of them. Furthermore, according to the Levenshtein distance, we removed tweets with a similarity higher than 70% to exclude similar tweets that might contaminate our dataset. This second cleaning step yielded 2.8K tweets with a predominance of Guarani (see also **S** 3.1 to see how we ensure that these candidate tweets are Guarani-dominant).

Limitations

Such a reduction in the number of collected tweets after the filtering process is a clear indicator of how difficult it is to collect data for truly low-resource languages. This issue is well-known in Guarani [43]. To counteract it, we needed

[4] Some examples of keywords used to download the tweets are: 'chera'a' ('friend', 'bro'), 'ojoaju' ('joins', 'unites'), 'rejapo' ('do', 'make'), 'ningo' (a link of the emphasizing type, which joins the subject and predicate of a sentence, which serves to remark that is something that the speaker knows to be true), 'aguyje' ('thanks', 'thank you'), 'haguére' ('because', 'due to'), 'irundy' ('four'), 'oñembo' ('pose as', 'pretend to be', 'impersonate', 'get hold of *sth.*'), 'pire' ('skin', 'leather', 'bark'), 'epyta' ('stop', 'keep/stand/stay still').

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

[6] https://polyglot.readthedocs.io/en/latest/Detection.html

to monitor tweets for a period of 6 months: a large and representative amount of time to collect tweets for academic purposes. With this strategy, and after an exhaustive filtering process to ensure the quality of the dataset, only 0.037% of the tweets were Guarani-dominant and therefore of interest for this task. This shows the challenges of working with low-resource domain setups, and the relevance of studying and designing strategies to work with data scarcity, both at training and inference time. Yet, note that the size is not an outlier in low-resource research, and it is consistent with the size of other sentiment-related datasets [59], and also with the size of low-resource datasets that are widely used by the NLP community in other areas, such as for instance named-entity recognition corpus for African languages [60] (datasets in the range of 2K samples) or low-resource dependency parsing [61] (a significant number of languages with less than 1,000 annotated sentences).

3.1 Annotation of the dataset

Two native Guarani and Spanish speakers, one female and one male, both aged 27-35, annotated the dataset. As in [43], annotators were also asked to determine the language of the tweet, with three tags to choose from: Guarani, Jopara, or another language. We then kept only the tweets labeled as Guarani or Jopara, finally resulting in 2,364 tweets, which we proceeded to annotate as described below.

Using these tweets, we created three new datasets for Jopara affective detection for the following tasks: (i) emotion recognition, (ii) humor detection, and (iii) toxic and offensive language identification. Note that not all tweets might be appropriate for all tasks. Figure 3 shows the overlap of the tweets between the corpora (that are annotated for different tasks). Even if it is out of the scope of this work, note that given the significant overlap between different tasks, an interesting future research line could be to apply multi-task learning techniques to see how one task might help improve the performance of the others.

It is important to underline that the resulting dataset is unbalanced. However, at the same time, it is more representative of the expected real-world distributions for the tasks addressed. Table 1 shows, in a detailed way, the number of tweets by classes and splits for each corpus.

The Jopara emotion recognition corpus

For emotion annotation, annotators were asked to identify the predominant emotion of the tweet into four classes: `happy`, `sad`, `angry`, and `other` (tweets that show none of the above emotions, or neutral emotion). For this, the guidelines of SemEval-2019 Task 3 [62] were followed. Then, we used Cohen's kappa metric to measure inter-annotator agreement, obtaining a moderate agreement (0.55) according to [Figure 1; 63, p. 576].

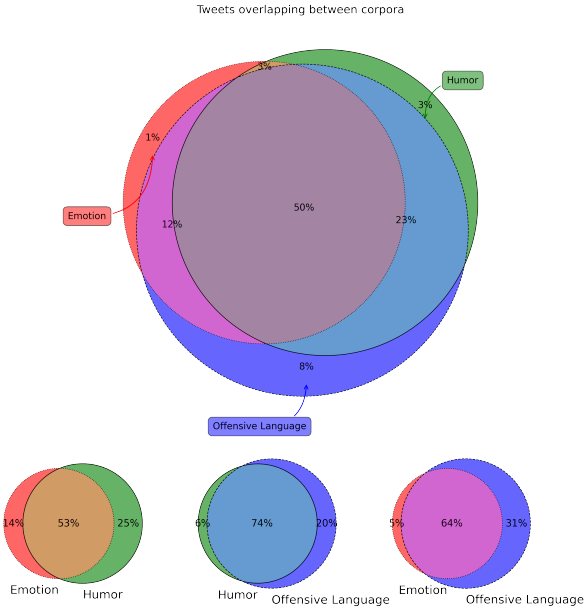

Fig. 3 Venn diagram showing the overlap between our corpora (adapted from [17, p. 57, Figure 3.1]). The Venn diagram at the top illustrates the overlap between our three corpora. At the bottom, (i) the diagram at the left shows the relationship between the emotion recognition and humor detection corpora; (ii) at the middle, between the former and the offensive language identification corpus; and finally (iii) at the right, between the latter and the emotion recognition corpus.

The Jopara humor detection corpus

In this corpus, annotators were asked to specify whether the tweet could be considered as `not humorous` or `humorous` (i.e., whether the post was intended to be funny, or contained humorous parts or jokes). In this case, for humor detection we followed the shared-tasks annotation guidelines by [64, 65]. We again obtained a moderate agreement (0.52) using Cohen's kappa.

The Jopara toxic and offensive language identification corpus

For this dataset, we asked annotators to label the content of the tweet as `not offensive` or `offensive` (i.e., hate speech, profane or inappropriate language, insults, or threats). About the guidelines to identify offensive language, SemEval-2020 Task 12 guidelines [48] were followed. We obtained a Cohen's kappa inter-annotator agreement of 0.73.

4 Models

Next, we describe sequential classifiers without pre-training as well as those with pre-training (i.e., fine-tuned large language models). For the former group, we considered recurrent neural networks (RNN), more specifically, long short-term memory networks [66, LSTM], together with convolutional neural

Table 1 The statistics for the Jopara affective analysis datasets and their splits for each proposed task (adapted from [17, p. 57, Table 3.3]).

Class	grn	jopara	Total	Train	Development	Test
			Emotion Recognition			
e_angry	166	315	481	315	75	91
e_happy	163	194	357	216	52	89
e_other	305	336	641	411	111	119
e_sad	50	42	92	63	13	16
Total	684	887	**1,571**	1,005	251	315
			Humor Detection			
fun	145	359	504	323	75	106
no_fun	668	670	1,338	855	220	263
Total	813	1,029	**1,842**	1,178	295	369
			Offensive Language Identification			
off	110	239	349	232	47	70
no_off	833	988	1,821	1,156	301	364
Total	943	1,227	**2,170**	1,388	348	434

networks [67, CNN], whose use might be relevant, compared to language models, in low-resource setups where the lack of large datasets might be an issue. For the latter group, we considered the multilingual version of BERT [68, mBERT] (with the aim of exploiting the cross-lingual capabilities of BERT, regardless of the amount of word overlapping [69]) and a BERT model for Spanish [70, BETO] (since it is the most intertwined language with Guarani, and despite being a Guarani-dominant dataset, Jopara tweets contain words and subwords coming from Spanish). Additionally, we also introduce specific BERT-based models for Guarani. Finally, to keep a homogeneous evaluation, the models do not include handcrafted features or any additional information apart from words.

4.1 Sequential classifiers without pre-training

We build bidirectional LSTMs (biLSTMs) and CNNs as implemented in NCRF++ [71]. It should be noted that CNNs have traditionally been shown to perform well in sentiment analysis [72] and that LSTMs are widely used in numerous NLP tasks. We use randomly initialized word vectors for the input word embeddings. We also concatenate another word embedding, which can be either computed by character CNNs or biLSTMs. We report the details about the configurations of these models in Appendix A.

4.2 Sequential classifiers that use existing pre-trained language models

We include the following BERT [68] models in our study: (i) BETO-base-cased (pre-trained on Spanish, 12 layers) [70], and (ii) multilingual-bert-base-cased (mBERT-base-cased, a 104-languages version of BERT with 12 layers). Guarani was not included during the pre-training of the multilingual model, which implies that the models lack knowledge to recognize and understand

Guarani language patterns and nuances. However, BERT's pre-trained models use a word-piece tokenizer [73] to build up a vocabulary of the most frequently occurring sub-word pieces, and multilingual language models have shown nice cross-lingual abilities [69, 74–76], even on few-shot learning setups [77, 78].

4.2.1 Specific language models for Guarani

We take the approach presented in the Bertinho paper [79] for training monolingual language models for low-resourced languages for the pre-training phase, but in our case, there are extra challenges such as the much smaller training data size. Also, following the RoBERTa paper [80], we pre-trained only on the masked language objective, i.e., we have not applied the next-sentence prediction objective, since it was shown in [80] that such objective does not come with significant improvements.

Guarani unlabeled data

To train our models, we relied on the Guarani version of Wikipedia. Specifically, we used Wikipedia[7] and Wiktionary[8] dumps, both cleaned with `wikiextractor` [81]. We obtained 800K tokens, far less than, for example, the amount of data used for either Spanish [70] or English [68] with around 3B tokens, even less than that of other languages with low resources such as Basque [82, about 224M tokens], or Galician [79, about 45M tokens]. It is worth noting that the available BERT models of truly low-resource languages are not even comparable to the small amount of data used in our BERT models. For example, [25, AfriBERta] (BERT model for African languages) was trained on a multilingual corpus, containing approximately 108M tokens. More particularly, we use 760K tokens, i.e., 95% of these wiki articles, for the training set, and for the development set, the remaining 40K tokens (5% of the articles). We use the latter to ensure that the training converges by tracking the loss at different checkpoints. More specifically, we considered two types of models: (i) those that are initialized with the weights of an existing model, and (ii) those trained from scratch. We will explain these two strategies shortly, but first, we will comment on a few general aspects.

General comments on the proposed Guarani language models

We focus on masked language models since they are the most popular choice as a backbone for downstream natural language processing and language and vision models [83, 84]. Note that although generative models are also powerful, they are originally designed for language generation, and do not benefit from the advantages of bidirectionality. We will rely on the training hyperparameters introduced in the BERT [68] paper. We introduce some minor variations, such as training some additional smaller models in terms of the number of layers and attention heads (in particular, these are the tiny and small models, which we will detail in the following sections). See Table 3 for details. Also, due to

[7] From https://dumps.wikimedia.org/gnwiki/, keeping both main texts and headers.
[8] From https://dumps.wikimedia.org/gnwiktionary/.

limitations of physical resources, we restrict the size of the input sequences as specified in Table 3, to control the impact of the attention mechanism in the amount of memory that we need to allocate.

Language models for Guarani with pre-initialized weights

To counteract the truly small amount of available, raw data for Guarani, we are considering a second pre-training phase, starting with existing models trained on related languages. The motivation behind this is that this could be a better strategy than starting training an architecture on randomly initialized weights. Note that although these models were not initially trained on Guarani text, their representational space might be encoding grammar and word-piece vocabulary information from related languages [75]. In this context, there is some empirical evidence that shows that this is a strategy worth exploring. For instance, authors such as [85] and [86] used mBERT models as the starting point to train language models for Russian and Portuguese, instead of training them from just random weights. They compared this approach against the corresponding monolingual model trained from scratch with randomly initialized weights and showed that the former obtained better results across different downstream tasks, while requiring less training time than the randomly initialized counterparts. Still, that the application of this strategy for a language that is not included in multilingual models and that has very few available resources, such as the case of Guarani, has been explored little to date.

We are using two models: BETO[9] [70] (a Spanish BERT model, a language with which Guarani and in particular Jopara is widely intertwined), and mBERT[10] [68]; as a basis for training on top of their representations a model with Guarani data. To train these models, we relied on the Hugging Face library [87] and a standalone GPU (NVIDIA GeForce RTX 3090 with 24GB) was used to train all models. We will explain these models and the training strategies next.

Language models for Guarani trained from scratch

For comparison against the previous models, we also consider training a monolingual model with randomly initialized weights. We considered a vocabulary of 30K sub-word pieces for the tokenizer, which is a standard size [68, 70, 79, 86], and considered words that occurred 2 or more times. A sample sentence tokenized by BETO, mBERT, and the proposed tokenizer are shown in Table 2. We trained four models and varied their batch size based on the model size, to exploit most of the GPU[11] (more details can be seen in Table 3):

1. *gnBERT-tiny*:[12] A BERT-tiny-cased model of only 2 layers of Transformers. The model was trained up to 1.5M steps (which took us days). Using

[9] https://huggingface.co/mmaguero/beto-gn-base-cased
[10] https://huggingface.co/mmaguero/multilingual-bert-gn-base-cased
[11] A NVIDIA GeForce RTX 3090 with 24GB.
[12] https://huggingface.co/mmaguero/gn-bert-tiny-cased

training batches of 192 in size and considering sequences of 144 tokens (due to our hardware limitations).
2. *gnBERT-small*:[13] A BERT-small-cased model with 6 Transformer layers. The training batches adopted for the model were of size 96 with sequences of 128 tokens. Similarly to `gnBERT-tiny`, the model was trained up to 1.5M steps (which took days).
3. *gnBERT-base*:[14] This is a BERT-base-cased model of 12 Transformer layers. The model was trained up to 2M steps (taking 22 days to be completed). Using training batches of size 96 and sequences of 128 tokens.
4. *gnBERT-large*:[15] A BERT-large-cased model using 24 Transformer layers. Sequences of 128 tokens are used and training batches of size 32 are considered. This model was trained up to 3M steps (the training was finished in 32 days).

Table 2 Guarani and Jopara sentence tokenization using the tokenizers of mBERT, BETO and our model (adapted from [17, p. 81, Table 6.2]). *Ahata pohanohárape ha upéi cherógape.* (`grn`). *Ahata doctorpe ha upéi cherógape.* (`jopara`). *I am going to the doctor's and then to my house.* (`eng`). The word pieces of BETO are shorter and more difficult to interpret, as are those of mBERT. Thanks to the correspondence with words of the Guarani dictionary, our Guarani BERT's subword tokenization provides a better understanding of word formation. This is clear with *pohanohára* (Doctor of Medicine), *róga* (house), *pe* (to the / to / in), and *upéi* (then / after).

Model	Tokenization
	grn
BETO	Ah ##ata po ##han ##oh ##ár ##ap ##e ha up ##é ##i che ##ró ##ga ##pe .
mBERT	Ah ##ata po ##han ##oh ##ára ##pe ha up ##éi che ##ró ##gap ##e .
Ours	Aha ##ta pohanohára ##pe ha upéi che ##róga ##pe .
	jopara
BETO	Ah ##ata doctor ##pe ha up ##é ##i che ##ró ##ga ##pe .
mBERT	Ah ##ata doctor ##pe ha up ##éi che ##ró ##gap ##e .
Ours	Aha ##ta doctor ##pe ha upéi che ##róga ##pe .

Hyperparameters. At the pre-training time, a learning rate of 1×10^{-4} is used with a linear weight decay of 0.01, using an Adam network weight optimizer set to $\epsilon = 1 \times 10^{-8}$. In addition, from an input sentence, at random 15% of the tokens were masked, of which 80 % were changed to the wildcard symbol `[MASK]`, and the remaining 10 % were replaced by a random word from the input vocabulary and the last 10 % were left unchanged; similar to what was proposed in the original BERT paper.

5 Experiments

We now proceed to describe our experimental setup. For reproducibility, we release the IDs of the tweets of the datasets and the Guarani BERT language

[13] https://huggingface.co/mmaguero/gn-bert-small-cased
[14] https://huggingface.co/mmaguero/gn-bert-base-cased
[15] https://huggingface.co/mmaguero/gn-bert-large-cased

Table 3 Models pre-trained on language modeling for Guarani.

		Monolingual			Pre-fine-tuning	
Parameter	**tiny**	**small**	**base**	**large**	**BETO**	**mBERT**
Data[1]	800K (Guarani Wikipedia and Wiktionary)					
Transformer layers	2	6	12	24	12	12
Hidden size	256	768	768	1024	768	768
Self-attention heads	4	12	12	16	12	12
Filter intermediate	768	3072	3072	4096	3072	3072
Tokens sequences	144	128	128	128	128	128
Batch[2]	192	96	96	32	32	32
Steps[3]	1.5M	1.5M	2M	3M	1M	1M
Days[4]	10	12	22	32	5	5

[1] 30K vocabulary size for monolingual models.

[2] It should be noted that, given the limitations of our hardware, we used small batches at training time.

[3] To compensate for the small training batch, the models used more steps during training.

[4] 2.95+e18 floating point operations (FLOP) equals about 1 day of training.

models (see **S** 6). We ran experiments on the three proposed datasets (emotion recognition, humor detection, and offensive language identification), and evaluated the models outlined above. To fine-tune each model, we performed a small hyperparameter search on the development set (for detailed information, see Appendix A). Also, note that performance across tasks can vary since, some tasks can be more challenging than others for the models, even if they are performed over overlapping texts.

Unbalanced datasets

As we highlighted in **S** 3.1, we are interested in identifying affective computation in Guarani/Jopara tweets as they occur in the original distribution of the dataset. We discard applying oversampling and downsampling techniques because we are especially interested in setups that can be of interest for the area of cognitive computation, a setup where such techniques would not fit well (for instance, humans are able to learn from extremely unbalanced distributions and do not apply this time of techniques). Also, note that, as previous work on low-resource language has suggested [88], it is recommendable to stick to realistic practices and not to overestimate the performance of language technologies in this type of language. In this line, we have preferred to respect and keep the data and its distribution as when it was collected.

Metrics

For evaluation, we used macro-accuracy instead of micro-accuracy for mitigating the impact of the imbalanced classes in our corpora. Still, for a thorough picture, we also report micro-accuracy, and macro- and micro-F1[16] scores.

[16] For the emotion recognition corpus, we calculated the micro-F1 score for the emotion classes (`angry`, `happy`, and `sad`), excluding the `other` class, according to [62, **S** 6]. Similarly, for the binary humor detection and offensive language identification corpora, we computed the micro-F1 score on the positive class (`fun` and `off`, respectively).

5.1 Results

In this section, we show the results obtained for each task. All the results correspond to pre-trained language models that are first fine-tuned on the corresponding training split of the affective dataset.

5.1.1 Emotion recognition task

Table 4 shows the results for all models and the emotion recognition task. Among the models that do not use pre-training, we find that a biLSTM model using character- and word-level embeddings performed the best. With respect to the classifiers that rely on pre-trained language models, they performed better than the classifiers that did not have pre-training. Overall, the results for the emotion recognition task are lower than for the other proposed tasks. This can be associated with different factors such as this task being a multi-class classification problem (while the humor detection and offensive language identification tasks are binary classification) and being a harder and more subjective problem. In addition, the size of the emotion recognition dataset is slightly smaller than the others, leading to extra challenges in capturing the diversity and nuances of emotional states accurately.

Table 4 Results for the emotion recognition task (adapted from [17, p. 82, Table 6.3], adding the `gnBERT-small` model to compare the performance with the others).

Model	Corpus Emotion Recognition			
	macro-Acc.	**Acc.**	**macro-F1**	**F1**
Models without pre-training				
^{C}CNN-^{W}biLSTM	0.4211	0.5015	0.4211	0.4908
^{C}CNN-^{W}CNN	0.5491	0.5428	0.5491	0.5837
^{C}biLSTM-^{W}CNN	0.5567	0.5809	0.5567	0.5980
^{C}biLSTM-^{W}biLSTM	0.5997	0.5841	0.5997	0.5952
Models without pre-fine-tuning				
BETO$_{base,cased}$	0.609	0.654	0.606	0.648
mBERT$_{base,cased}$	0.5684	0.6317	0.5644	0.6463
Models with pre-fine-tuning				
BETO+gn$_{base,cased}$	0.5761	0.6349	0.5673	0.6183
mBERT+gn$_{base,cased}$	0.5599	0.6317	0.5637	0.6146
Models pre-trained from scratch				
gnBERT$_{tiny,cased}$	0.5766	0.658	0.5763	0.6718
gnBERT$_{small,cased}$	0.5528	0.5937	0.5598	0.607
gnBERT$_{base,cased}$	0.5265	0.6032	0.5409	0.6038
gnBERT$_{large,cased}$	0.25	0.3778	0.1371	0.0

^{C}Uses character embeddings. ^{W}Uses word embeddings.

The experiments showed that the `BETO-base` model, which did not include Guarani during pre-training or pre-fine-tuning, achieved the best results on the macro-accuracy and macro-F1 scores, i.e., it is the best-performing model across classes if we ignore their frequency. Still, among the Guarani pre-trained language models, the `gnBERT-tiny` (2 layers, pre-trained only on Guarani) did equally well on predicting the micro-F1 score while saving significant computational power. On the contrary, the large Guarani models, such as

`gnBERT-large` did not perform well. We hypothesize this is due to architectures being too complex with a high number of parameters which cannot properly exploit the lack of data for Guarani. This model always predicted the same class, i.e., the (majority) `other` class, despite our efforts for tuning hyperparameters, reducing learning rates, and changing the optimizers. This `gnBERT-large` model showed these same instability problems for all the three evaluated corpora. Comparing these results with Table 3, which shows the amount of time that it took to pre-train on the Guarani Wikipedia the tiny (10 days), small (12 days), based (22 days), and large (32 days) Guarani models, and the multilingual language models (5 days), we can state that, to carry out emotion analysis in Guarani - either smaller models or models initialized with weights coming from multilingual models - are a more efficient choice in terms of pre-training time while obtaining at the same time the best performance.

5.1.2 Humor detection task

Table 5 shows the results for the humor detection task. In this case, for the sequential classifiers without pre-training - the one using a CNN at the character level and a biLSTM encoder at the word level - achieved the strongest results.

Table 5 Results for the humor detection task (adapted from [17, p. 83, Table 6.4], adding the `gnBERT-small` model to compare the performance with the others).

Model	Corpus Humor Detection			
	macro-Acc.	**Acc.**	**macro-F1**	**F1**
Models without pre-training				
^{C}CNN-^{W}biLSTM	0.6773	0.7127	0.6773	0.5431
^{C}CNN-^{W}CNN	0.6323	0.7127	0.6323	0.47
^{C}biLSTM-^{W}CNN	0.6407	0.7046	0.6407	0.4882
^{C}biLSTM-^{W}biLSTM	0.6637	0.7615	0.6637	0.5111
Models without pre-fine-tuning				
BETO$_{base,cased}$	0.6841	0.7263	0.6771	0.5511
mBERT$_{base,cased}$	0.7054	0.6965	0.6708	0.5627
Models with pre-fine-tuning				
BETO+gn$_{base,cased}$	0.6823	0.7317	0.6786	0.5092
mBERT+gn$_{base,cased}$	0.691	0.7642	0.6988	0.5584
Models pre-trained from scratch				
gnBERT $_{tiny,cased}$	0.7208	0.7344	0.7	0.5984
gnBERT $_{small,cased}$	0.7076	0.7317	0.6923	0.5823
gnBERT $_{base,cased}$	0.6963	0.7317	0.6886	0.5677
gnBERT $_{large,cased}$	0.5	0.7458	0.4272	0.0

^{C}Uses character embeddings. ^{W}Uses word embeddings.

Among the fine-tuned language models, the best-performing models were those that were pre-trained or pre-fine-tuned on Guarani. More particularly, our `gnBERT-tiny` (2 layers, exclusively pre-trained on Guarani) achieved the top rank in almost all metrics, and the mBERT model pre-fine-tuned on Guarani data, i.e., `mBERT+gn` (12 layers) also obtained strong results. Overall, the results follow the trend obtained for the emotion recognition task and the conclusions are similar: smaller Guarani models pre-trained from scratch

or models initialized with weights already initialized from multilingual models are a more robust choice. More specifically, the results across the board for this task show that improvements between non-pretrained and pre-trained models range between 1 and 3 percentage points. This demonstrates a consistently better performance, with the exception of a biLSTM model (at character and word level) that performs very closely to language models in terms of the micro-accuracy metric.

5.1.3 Offensive language identification task

Table 6 Results for the offensive language identification task (adapted from [17, p. 83, Table 6.5], adding the `gnBERT-small` model to compare the performance with the others).

Model	Corpus Offensive Language Identification			
	macro-Acc.	**Acc.**	**macro-F1**	**F1**
Models without pre-training				
^{C}CNN-^{W}biLSTM	0.7527	0.8755	0.7527	0.5970
^{C}CNN-^{W}CNN	0.7725	0.841	0.7725	0.5766
^{C}biLSTM-^{W}CNN	0.7593	0.8479	0.7593	0.5714
^{C}biLSTM-^{W}biLSTM	0.7373	0.8594	0.7373	0.5611
Models without pre-fine-tuning				
BETO$_{base,cased}$	0.8058	0.8871	0.7971	0.6621
mBERT$_{base,cased}$	0.8269	0.8548	0.7726	0.6358
Models with pre-fine-tuning				
BETO+gn$_{base,cased}$	0.7538	0.8871	0.7774	0.6142
mBERT+gn$_{base,cased}$	0.7791	0.9101	0.8127	0.6777
Models pre-trained from scratch				
gnBERT $_{tiny,cased}$	0.7626	0.8825	0.7736	0.6165
gnBERT $_{small,cased}$	0.7901	0.8802	0.7835	0.6389
gnBERT $_{base,cased}$	0.7959	0.8802	0.7859	0.6438
gnBERT $_{large,cased}$	0.5	0.8337	0.4561	0.0

^{C}Uses character embeddings. ^{W}Uses word embeddings.

For our third corpus on offensive language identification, we show the results in Table 6. Once again, we observed the same trends for the models without pre-training as for emotion recognition (Table 4) and humor detection tasks (Table 5), and again models that use character-level representations and biLSTMs to contextualize the input sentences performed better. Overall, the experiments across the three corpora support our findings that, to process Guarani, the use of character-level representations is the optimal choice.

Among pre-trained language models, the main trends remain and once again those models that consider Guarani data during pre-training obtained the best results, especially `mBERT+gn` while the tiny, small, and base Guarani models also obtain a competitive performance.

6 Conclusion

In this study, we presented a text-based affective classification framework for Guarani and Jopara, a code-switching language combining Guarani and Spanish. Our approach involved the collection of a substantial number of

Guarani-dominant tweets, which were then annotated for three separate tasks: emotion recognition, humor detection, and offensive language identification. Then, we trained various neural models, considering the specific characteristics of Jopara and the challenges posed by low-resource settings, such as limited pre-training data and hardware resources for training language models.

Overall, despite these challenges, we observed that models with pre-training generally showed a better performance than models without pre-training. Due to the low-resource nature of the setup and the different nature of the proposed tasks, the magnitude of the improvements can change from task to task. In addition, we found that models incorporating language pre-training on Guarani, even with limited unlabeled data, yielded the best results. Additionally, it was intriguing to note that a two-layer Guarani monolingual BERT model demonstrated robust performance while being computationally efficient. However, we were not able to obtain clear improvements with larger language models, despite requiring more memory and training time. This finding suggests that, in certain resource-constrained scenarios, a simpler architecture could be the optimal choice. Thus, we conclude that pre-trained multilingual models adaptable to Guarani or smaller monolingual language models trained from scratch - specifically for Guarani - appear to be the most reliable options for addressing the problem investigated in this study.

As future work, we plan the exploration of multi-task learning approaches on our datasets, that have already demonstrated benefits for affective detection [89] and other low-resource language setups [90–92].

Data and materials availability. The tweet IDs of the datasets for text-based affective detection in Guarani/Jopara and Guarani BERT language models can be obtained at https://github.com/mmaguero/guarani-multi-affective-analysis and https://huggingface.co/mmaguero, respectively. For further details, please contact the authors.

Acknowledgements. We thank the annotators for their work. We are also grateful to *ExplosionAI* for giving us access to their `Prodigy` annotation tool[17] under the research license. We also thank (i) the Visual Information Processing Group of the University of Granada (especially Javier Mateos Delgado) and (ii) the Generalitat Valenciana and the University of Alicante through the DGX computing platform for giving us access to the GPU hardware necessary to carry out the training of the language models. Finally, we thank Olga Zamaraeva and Ana Vilar for their valuable assistance with English writing.

Compliance with ethical standards

Funding. This work is supported by a 2020 Leonardo Grant for Researchers and Cultural Creators from the FBBVA.[18] This paper has also received funding from grant SCANNER-UDC (PID2020-113230RB-C21) funded by

[17] https://prodi.gy/

[18] FBBVA accepts no responsibility for the opinions, statements, and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors.

MCIN/AEI/10.13039/501100011033, the European Research Council (ERC), which has supported this research under the European Union's Horizon Europe research and innovation programme (SALSA, grant agreement No 101100615), Xunta de Galicia (ED431C 2020/11), and Centro de Investigación de Galicia "CITIC", funded by Xunta de Galicia and the European Union (ERDF - Galicia 2014-2020 Program), by grant ED431G 2019/01. Additionally, the research leading to these results received funding from the University of Granada, Generalitat Valenciana and the University of Alicante (IDIFEDER/2020/003).

Conflict of interest. The authors declare no competing interests.

Ethical approval. This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix A Hyperparameters search and implementation details

Table A1 Hyperparameters for model training (adapted from [17, p. 228, Table C.2]).

Parameter	Options
NCRF++ [71]	
Optimizer	[Adam, AdaGrad, SGD]
Avg. batch loss	[True, False]
Learning rate	[2e-5, 5e-4, 0.015, 0.16, 0.1, 0.2]
Char hidden dim.	[50, 100, 200, 400]
Word hidden dim.	[50, 100, 200, 400]
Momentum	[0.0, 0.9, 0.95, 0.99]
LSTM Layers	[1, 2]
Dropout	[0.0, 0.5]
Iteration	[20,30,40,50]
Hugging Face [87]	
Eval. steps	[200]
Eval. strategy	[steps]
Disable tqdm	[False]
Eval. batch size	[8, 16, 32]
Train batch size	[8, 16, 32]
Learning rate[1]	[2e-5 - 5e-5]
Dropout	[0.1 - 0.6]
Epoch	[10 - 50]
Weight decay	[0.0 - 0.3]

[1]Except for the original BETO model, where 3e-5 was enough to converge.

Table A1 shows the hyperparameters used to train all proposed models, using (i) the `NCRF++` [71] for training our sequential classifiers without pre-training and (ii) for the sequential classifiers that use pre-trained language models, the *Hugging Face* package [87].

For models with Transformers, hyperparameter selection was performed with a Bayesian hyperparameter search method using the platform *W&B* [19] [93] on the dev set. It chooses the parameters to optimize the probability of improvement based on the relation between these parameters and the model metric (macro-accuracy in our case). On the other hand, a batch size of 10 was

[19]https://docs.wandb.ai/guides/sweeps

used to train the CNN and biLSTM models, and the hyperparameter selection was performed with a random hyperparameter search method. In addition, we train these models for a maximum of 50 epochs. Early stopping criteria (set to 3) were used to train the Transformer-based models. Finally, an NVIDIA Tesla T4 (16GB GPU) is used to train all models.

References

[1] Mager M, Gutierrez-Vasques X, Sierra G, Meza-Ruiz I. Challenges of language technologies for the indigenous languages of the Americas. In: Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico, USA: Association for Computational Linguistics; 2018. p. 55–69. Available from: https://aclanthology.org/C18-1006.

[2] Mager M, Oncevay A, Ebrahimi A, Ortega J, Rios A, Fan A, et al. 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. Online: Association for Computational Linguistics; 2021. p. 202–217. https://doi.org/10.18653/v1/2021.americasnlp-1.23. Available from: https://aclanthology.org/2021.americasnlp-1.23.

[3] García Trillo MA, Estrella Gutiérrez A, Gelbukh A, Peña Ortega AP, Reyes Pérez A, Maldonado Sifuentes CE, et al. Procesamiento de lenguaje natural para las lenguas indígenas. 1. Universidad Michoacana de San Nicolás de Hidalgo; 2021. https://isbnmexico.indautor.cerlalc.org/catalogo.php?mode=detalle&nt=334970.

[4] Estigarribia B. Guarani-Spanish Jopara Mixing in a Paraguayan Novel: Does it Reflect a Third Language, a Language Variety, or True Codeswitching? Journal of Language Contact. 2015;8(2):183–222. https://doi.org/https://doi.org/10.1163/19552629-00802002.

[5] Chiruzzo L, Góngora S, Alvarez A, Giménez-Lugo G, Agüero-Torales M, Rodríguez Y. Jojajovai: A Parallel Guarani-Spanish Corpus for MT Benchmarking. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2022. p. 2098–2107. Available from: https://aclanthology.org/2022.lrec-1.226.

[6] Boidin C. "Jopara : una vertiente sol y sombra del mestizaje". In: et Haralambos Symeonidis WD, editor. Tupí y Guaraní. Estructuras, contactos y desarrollos. vol. 11 of Regionalwissenschaften Lateinamerika. Munster, Germany: LIT-Verlag; 2005. p. 303–331. Available from: https://halshs.archives-ouvertes.fr/halshs-00257767.

[7] Bittar Prieto J. A variationist perspective on Spanish-origin verbs in Paraguayan Guarani [Master's Thesis]. The University of New Mexico. New Mexico; 2016. Available from: https://digitalrepository.unm.edu/ling_etds/4.

[8] Bittar Prieto J.: A Constructionist Approach to Verbal Borrowing: The Case of Paraguayan Guarani. The University of New Mexico's Latin American & Iberian Institute 2020 PhD Fellows. URL: https://www.youtube.com/watch?v=C5XiLqR4onA.

[9] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Association for Computational Linguistics; 2002. p. 79–86. https://doi.org/10.3115/1118693.1118704. Available from: https://aclanthology.org/W02-1011.

[10] Cambria E, Hussain A. Sentic computing. Cognitive Computation. 2015;7(2):183–185. https://doi.org/10.1007/s12559-015-9325-0. Available from: https://doi.org/10.1007/s12559-015-9325-0.

[11] Ghosh S, Ekbal A, Bhattacharyya P. A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. Cognitive Computation. 2022;14(1):110–129. https://doi.org/10.1007/s12559-021-09828-7. Available from: https://doi.org/10.1007/s12559-021-09828-7.

[12] Lieberman MD. Affect labeling in the age of social media. Nature Human Behaviour. 2019;3(1):20 – 21. https://doi.org/10.1038/s41562-018-0487-0. Available from: https://doi.org/10.1038/s41562-018-0487-0.

[13] Adwan OY, Al-Tawil M, Huneiti A, Shahin R, Abu Zayed A, Al-Dibsi R. Twitter Sentiment Analysis Approaches: A Survey. International Journal of Emerging Technologies in Learning (iJET). 2020 Aug;15(15):pp. 79–93. https://doi.org/10.3991/ijet.v15i15.14467. Available from: https://online-journals.org/index.php/i-jet/article/view/14467.

[14] Jakobsen AL, Mesa-Lao B. Translation in transition: Between cognition, computing and technology. vol. 133. John Benjamins Publishing Company; 2017. Available from: https://www.jbe-platform.com/content/books/9789027265371.

[15] Jain DK, Boyapati P, Venkatesh J, Prakash M. An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification. Information Processing & Management. 2022;59(1):102758. https://doi.org/https://doi.org/10.1016/j.ipm.2021.

102758. Available from: https://www.sciencedirect.com/science/article/pii/S0306457321002399.

[16] Green D. Language Control in Different Contexts: The Behavioral Ecology of Bilingual Speakers. Frontiers in Psychology. 2011;2. https://doi.org/10.3389/fpsyg.2011.00103. Available from: https://www.Frontiersin.org/articles/10.3389/fpsyg.2011.00103.

[17] Agüero-Torales MM. Machine Learning approaches for Topic and Sentiment Analysis in multilingual opinions and low-resource languages: From English to Guarani [Ph.D. thesis]. University of Granada. Granada; 2022. Available from: http://hdl.handle.net/10481/72863.

[18] Hedderich MA, Lange L, Adel H, Strötgen J, Klakow D. A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: Association for Computational Linguistics; 2021. p. 2545–2568. https://doi.org/10.18653/v1/2021.naacl-main.201. Available from: https://aclanthology.org/2021.naacl-main.201.

[19] Pajupuu H, Altrov R, Pajupuu J. Identifying Polarity In Different Text Types. Folklore (14060957). 2016;64. https://doi.org/10.7592/FEJF2016.64.polarity.

[20] Afli H, McGuire S, Way A. Sentiment translation for low resourced languages: Experiments on irish general election tweets. In: 18th International Conference on Computational Linguistics and Intelligent Text Processing; 2017. p. 1–10. Available from: https://doras.dcu.ie/23370/.

[21] Batra R, Kastrati Z, Imran AS, Daudpota SM, Ghafoor A.: A large-scale tweet dataset for urdu text sentiment analysis. Available from: https://www.preprints.org/manuscript/202103.0572/v1.

[22] Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLOS ONE. 2015 12;10(12):1–22. https://doi.org/10.1371/journal.pone.0144296. Available from: https://doi.org/10.1371/journal.pone.0144296.

[23] Khan MY, Nizami MS. Urdu Sentiment Corpus (v1.0): Linguistic Exploration and Visualization of Labeled Dataset for Urdu Sentiment Analysis. In: 2020 International Conference on Information Science and Communication Technology (ICISCT). IEEE; 2020. p. 1–15. https://doi.org/10.1109/ICISCT49550.2020.9080043.

[24] Muhammad SH, Adelani DI, Ruder S, Ahmad IS, Abdulmumin I, Bello BS, et al.: NaijaSenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis. Marseille, France: European Language

Resources Association. Available from: https://aclanthology.org/2022.lrec-1.63.

[25] Ogueji K, Zhu Y, Lin J. Small Data? No Problem! Exploring the Viability of Pretrained Multilingual Language Models for Low-resourced Languages. In: Proceedings of the 1st Workshop on Multilingual Representation Learning. Punta Cana, Dominican Republic: Association for Computational Linguistics; 2021. p. 116–126. Available from: https://aclanthology.org/2021.mrl-1.11.

[26] Devi MD, Saharia N. Exploiting Topic Modelling to Classify Sentiment from Lyrics. In: Bhattacharjee A, Borgohain SK, Soni B, Verma G, Gao XZ, editors. Machine Learning, Image Processing, Network Security and Data Sciences. Singapore: Springer Singapore; 2020. p. 411–423. https://doi.org/10.1007/978-981-15-6318-8_34.

[27] Chen Y, Skiena S. Building Sentiment Lexicons for All Major Languages. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Baltimore, Maryland: Association for Computational Linguistics; 2014. p. 383–389. https://doi.org/10.3115/v1/P14-2063. Available from: https://aclanthology.org/P14-2063.

[28] Asgari E, Braune F, Roth B, Ringlstetter C, Mofrad M. UniSent: Universal Adaptable Sentiment Lexica for 1000+ Languages. In: Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2020. p. 4113–4120. Available from: https://aclanthology.org/2020.lrec-1.506.

[29] Duran M. Transformations and Paraphrases for Quechua Sentiment Predicates. In: Bekavac B, Kocijan K, Silberztein M, Šojat K, editors. Formalising Natural Languages: Applications to Natural Language Processing and Digital Humanities. Cham: Springer International Publishing; 2021. p. 61–73. https://doi.org/10.1007/978-3-030-70629-6_6.

[30] Ríos AA, Amarilla PJ, Lugo GAG. Sentiment categorization on a creole language with lexicon-based and machine learning techniques. In: 2014 Brazilian Conference on Intelligent Systems. IEEE; 2014. p. 37–43. https://doi.org/10.1109/BRACIS.2014.18.

[31] Borges Y, Mercant F, Chiruzzo L. Using Guarani Verbal Morphology on Guarani-Spanish Machine Translation Experiments. Procesamiento del Lenguaje Natural. 2021;66(0):89–98. Available from: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6325.

[32] Giossa N, Góngora S. Construcción de recursos para traducción automática guaraní-español [Bachelor's Thesis]. Universidad de la

República (Uruguay). Facultad de Ingeniería; 2021. (Bachelor's Thesis). Available from: https://hdl.handle.net/20.500.12008/30019.

[33] Kann K, Ebrahimi A, Mager M, Oncevay A, Ortega JE, Rios A, et al. AmericasNLI: Machine translation and natural language inference systems for Indigenous languages of the Americas. Frontiers in Artificial Intelligence. 2022;5. https://doi.org/10.3389/frai.2022.995667. Available from: https://www.frontiersin.org/articles/10.3389/frai.2022.995667.

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

[35] Cordova J, Boidin C, Itier C, Moreaux MA, Nouvel D. Processing Quechua and Guarani Historical Texts Query Expansion at Character and Word Level for Information Retrieval. In: Lossio-Ventura JA, Muñante D, Alatrista-Salas H, editors. Information Management and Big Data. Cham: Springer International Publishing; 2019. p. 198–211. https://doi.org/10.1007/978-3-030-11680-4_20. Available from: https://doi.org/10.1007/978-3-030-11680-4_20.

[36] Chiruzzo L, Agüero-Torales MM, Alvarez A, Rodríguez Y. Initial Experiments for Building a Guarani WordNet. In: Proceedings of the 12th International Global Wordnet Conference. Donostia/San Sebastian, Basque Country, Spain; 2023. Available from: https://www.hitz.eus/gwc2023/sites/default/files/aurkezpenak/GWC2023_paper_9051.pdf.

[37] Mazumder M, Chitlangia S, Banbury C, Kang Y, Ciro JM, Achorn K, et al. Multilingual Spoken Words Corpus. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2); 2021. Available from: https://openreview.net/forum?id=c20jiJ5K2H.

[38] Babu A, Wang C, Tjandra A, Lakhotia K, Xu Q, Goyal N, et al. XLS-R: Self-supervised cross-lingual speech representation learning at scale. In: Proceedings of the 23rd InterSpeech Conference; 2022. p. 2278–2282. https://doi.org/10.21437/Interspeech.2022-14. Available from: https://www.isca-speech.org/archive/pdfs/interspeech_2022/babu22_interspeech.pdf.

[39] Baevski A, Zhou Y, Mohamed A, Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems. vol. 33. Curran Associates, Inc.; 2020.

p. 12449–12460. Available from: https://proceedings.neurips.cc/paper/2020/file/92d1e1eb1cd6f9fba3227870bb6d7f07-Paper.pdf.

[40] Xu Q, Baevski A, Likhomanenko T, Tomasello P, Conneau A, Collobert R, et al. Self-Training and Pre-Training are Complementary for Speech Recognition. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2021. p. 3030–3034. https://doi.org/10.1109/ICASSP39728.2021.9414641.

[41] NLLB Team, Costa-jussà MR, Cross J, Çelebi O, Elbayad M, Heafield K, et al.: No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv. Available from: https://arxiv.org/abs/2207.04672.

[42] Yong ZX, Schoelkopf H, Muennighoff N, Aji AF, Adelani DI, Almubarak K, et al.: BLOOM+1: Adding Language Support to BLOOM for Zero-Shot Prompting. arXiv. Available from: https://arxiv.org/abs/2212.09535.

[43] Agüero-Torales MM, Vilares D, López-Herrera A. On the logistical difficulties and findings of Jopara Sentiment Analysis. In: Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching. Online: Association for Computational Linguistics; 2021. p. 95–102. https://doi.org/10.18653/v1/2021.calcs-1.12. Available from: https://aclanthology.org/2021.calcs-1.12.

[44] Strapparava C, Mihalcea R. Affect Detection in Texts. In: The Oxford Handbook of Affective Computing. Oxford Library of Psychology; 2015. https://doi.org/10.1093/oxfordhb/9780199942237.013.022.

[45] Ekman P. An argument for basic emotions. Cognition and Emotion. 1992;6(3-4):169–200. https://doi.org/10.1080/02699939208411068. Available from: https://doi.org/10.1080/02699939208411068.

[46] Plutchik R. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist. 2001;89(4):344–350. Available from: http://www.jstor.org/stable/27857503.

[47] Mihalcea R, Strapparava C. Learning To Laugh (Automatically): Computational Models For Humor Recognition. Computational Intelligence. 2006;22(2):126–142. https://doi.org/https://doi.org/10.1111/j.1467-8640.2006.00278.x. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.2006.00278.x.

[48] Zampieri M, Nakov P, Rosenthal S, Atanasova P, Karadzhov G, Mubarak H, et al. SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020). In: Proceedings of the

Fourteenth Workshop on Semantic Evaluation. Barcelona (online): International Committee for Computational Linguistics; 2020. p. 1425–1447. https://doi.org/10.18653/v1/2020.semeval-1.188. Available from: https://aclanthology.org/2020.semeval-1.188.

[49] Ranasinghe T, Zampieri M. Multilingual Offensive Language Identification with Cross-lingual Embeddings. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics; 2020. p. 5838–5844. https://doi.org/10.18653/v1/2020.emnlp-main.470. Available from: https://aclanthology.org/2020.emnlp-main.470.

[50] Wang M, Yang H, Qin Y, Sun S, Deng Y. Unified Humor Detection Based on Sentence-pair Augmentation and Transfer Learning. In: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation. Lisboa, Portugal: European Association for Machine Translation; 2020. p. 53–59. Available from: https://aclanthology.org/2020.eamt-1.7.

[51] Lamprinidis S, Bianchi F, Hardt D, Hovy D. Universal Joy A Data Set and Results for Classifying Emotions Across Languages. In: Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Online: Association for Computational Linguistics; 2021. p. 62–75. Available from: https://aclanthology.org/2021.wassa-1.7.

[52] Pfeiffer J, Vulić I, Gurevych I, Ruder S. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics; 2020. p. 7654–7673. https://doi.org/10.18653/v1/2020.emnlp-main.617. Available from: https://aclanthology.org/2020.emnlp-main.617.

[53] Estigarribia B. A grammar of Paraguayan Guarani. London: UCL Press; 2020. Available from: https://library.oapen.org/handle/20.500.12657/51773.

[54] Abdellaoui H, Zrigui M. Using Tweets and Emojis to Build TEAD: an Arabic Dataset for Sentiment Analysis. Computación y Sistemas. 2018 09;22:777 – 786. https://doi.org/10.13053/cys-22-3-3031. Available from: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462018000300777&nrm=iso.

[55] Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowledge and Information Systems. 2019;60(2):617–663. https://doi.org/10.1007/s10115-018-1236-4.

[56] Tejwani R.: Two-dimensional Sentiment Analysis of text. arXiv. Available from: https://arxiv.org/abs/1406.2022.

[57] Yen MF, Huang YP, Yu LC, Chen YL. A two-dimensional sentiment analysis of online public opinion and future financial performance of publicly listed companies. Computational Economics. 2021;p. 1–22. https://doi.org/10.1007/s10614-021-10111-y.

[58] Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Valencia, Spain: Association for Computational Linguistics; 2017. p. 427–431. Available from: https://aclanthology.org/E17-2068.

[59] Mamta, Ekbal A, Bhattacharyya P. Exploring Multi-Lingual, Multi-Task, and Adversarial Learning for Low-Resource Sentiment Analysis. ACM Trans Asian Low-Resour Lang Inf Process. 2022 sep;21(5). https://doi.org/10.1145/3514498. Available from: https://doi.org/10.1145/3514498.

[60] Adelani DI, Abbott J, Neubig G, D'souza D, Kreutzer J, Lignos C, et al. MasakhaNER: Named Entity Recognition for African Languages. Transactions of the Association for Computational Linguistics. 2021;9:1116–1131. https://doi.org/10.1162/tacl_a_00416. Available from: https://aclanthology.org/2021.tacl-1.66.

[61] de Marneffe MC, Manning CD, Nivre J, Zeman D. Universal Dependencies. Computational Linguistics. 2021 07;47(2):255–308. https://doi.org/10.1162/coli_a_00402. https://direct.mit.edu/coli/article-pdf/47/2/255/1938138/coli_a_00402.pdf. Available from: https://doi.org/10.1162/coli_a_00402.

[62] Chatterjee A, Narahari KN, Joshi M, Agrawal P. SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text. In: Proceedings of the 13th International Workshop on Semantic Evaluation. Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019. p. 39–48. https://doi.org/10.18653/v1/S19-2005. Available from: https://aclanthology.org/S19-2005.

[63] Artstein R, Poesio M. Inter-Coder Agreement for Computational Linguistics. Computational Linguistics. 2008 12;34(4):555–596. https://doi.org/10.1162/coli.07-034-R2. Available from: https://doi.org/10.1162/coli.07-034-R2.

[64] Chiruzzo L, Castro S, Rosá A. HAHA 2019 Dataset: A Corpus for Humor Analysis in Spanish. In: Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2020. p. 5106–5112. Available from: https://aclanthology.

org/2020.lrec-1.628.

[65] Hossain N, Krumm J, Gamon M, Kautz H. SemEval-2020 Task 7: Assessing Humor in Edited News Headlines. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation. Barcelona (online): International Committee for Computational Linguistics; 2020. p. 746–758. https://doi.org/10.18653/v1/2020.semeval-1.98. Available from: https://aclanthology.org/2020.semeval-1.98.

[66] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[67] LeCun Y, Bengio Y, et al. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. 1995;3361(10):1995. Available from: http://yann.lecun.com/exdb/publis/pdf/lecun-bengio-95a.pdf.

[68] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 4171–4186. https://doi.org/10.18653/v1/N19-1423. Available from: https://aclanthology.org/N19-1423.

[69] K K, Wang Z, Mayhew S, Roth D. Cross-Lingual Ability of Multilingual BERT: An Empirical Study. In: International Conference on Learning Representations; 2020. Available from: https://openreview.net/forum?id=HJeT3yrtDr.

[70] Cañete J, Chaperon G, Fuentes R, Ho JH, Kang H, Pérez J. Spanish Pre-Trained BERT Model and Evaluation Data. In: PML4DC at ICLR 2020; 2020. Available from: https://pml4dc.github.io/iclr2020/program/pml4dc_10.html.

[71] Yang J, Zhang Y. NCRF++: An Open-source Neural Sequence Labeling Toolkit. In: Proceedings of ACL 2018, System Demonstrations. Melbourne, Australia: Association for Computational Linguistics; 2018. p. 74–79. https://doi.org/10.18653/v1/P18-4013. Available from: https://aclanthology.org/P18-4013.

[72] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2014. p. 655–665. https://doi.org/10.3115/v1/P14-1062.

[73] Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. CoRR. 2016;abs/1609.08144. Available from: https://research.google/pubs/pub45610/.

[74] Pires T, Schlinger E, Garrette D. How Multilingual is Multilingual BERT? In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 4996–5001. https://doi.org/10.18653/v1/P19-1493. Available from: https://aclanthology.org/P19-1493.

[75] Wu S, Dredze M. Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT. 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). Hong Kong, China: Association for Computational Linguistics; 2019. p. 833–844. https://doi.org/10.18653/v1/D19-1077. Available from: https://aclanthology.org/D19-1077.

[76] Conneau A, Wu S, Li H, Zettlemoyer L, Stoyanov V. Emerging Cross-lingual Structure in Pretrained Language Models. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics; 2020. p. 6022–6034. https://doi.org/10.18653/v1/2020.acl-main.536. Available from: https://aclanthology.org/2020.acl-main.536.

[77] Lauscher A, Ravishankar V, Vulić I, Glavaš G. From Zero to Hero: On the Limitations of Zero-Shot Language Transfer with Multilingual Transformers. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics; 2020. p. 4483–4499. https://doi.org/10.18653/v1/2020.emnlp-main.363. Available from: https://aclanthology.org/2020.emnlp-main.363.

[78] Winata GI, Madotto A, Lin Z, Liu R, Yosinski J, Fung P. Language Models are Few-shot Multilingual Learners. In: Proceedings of the 1st Workshop on Multilingual Representation Learning. Punta Cana, Dominican Republic: Association for Computational Linguistics; 2021. p. 1–15. https://doi.org/10.18653/v1/2021.mrl-1.1. Available from: https://aclanthology.org/2021.mrl-1.1.

[79] Vilares D, Garcia M, Gómez-Rodríguez C. Bertinho: Galician BERT Representations. Procesamiento del Lenguaje Natural. 2021;66(0):13–26. Available from: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6319.

[80] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al.: RoBERTa: A Robustly Optimized BERT Pretraining Approach. Available from: https://openreview.net/forum?id=SyxS0T4tvS.

[81] Attardi G.: WikiExtractor. GitHub. https://github.com/attardi/wikiextractor.

[82] Agerri R, San Vicente I, Campos JA, Barrena A, Saralegi X, Soroa A, et al. Give your Text Representation Models some Love: the Case for Basque. In: Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2020. Available from: https://aclanthology.org/2020.lrec-1.588.

[83] Naseem U, Razzak I, Khan SK, Prasad M. A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models. ACM Trans Asian Low-Resour Lang Inf Process. 2021 jun;20(5). https://doi.org/10.1145/3434237. Available from: https://doi.org/10.1145/3434237.

[84] Zhou K, Yang J, Loy CC, Liu Z. Learning to Prompt for Vision-Language Models. Int J Comput Vision. 2022 sep;130(9):2337–2348. https://doi.org/10.1007/s11263-022-01653-1. Available from: https://doi.org/10.1007/s11263-022-01653-1.

[85] Kuratov Y, Arkhipov M. Adaptation of deep bidirectional multilingual transformers for russian language. In: Proceedings of the International Conference "Dialogue 2019". Moscow, Russia: Computational Linguistics and Intellectual Technologies; 2019. p. 333–339. Available from: https://www.dialog-21.ru/media/4606/kuratovyplusarkhipovm-025.pdf.

[86] Souza F, Nogueira R, Lotufo R. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In: Cerri R, Prati RC, editors. Intelligent Systems. Cham: Springer International Publishing; 2020. p. 403–417. https://doi.org/10.1007/978-3-030-61377-8_28.

[87] Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Online: Association for Computational Linguistics; 2020. p. 38–45. Available from: https://www.aclweb.org/anthology/2020.emnlp-demos.6.

[88] Kann K, Cho K, Bowman SR. Towards Realistic Practices In Low-Resource Natural Language Processing: The Development Set. 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). Hong Kong, China: Association

for Computational Linguistics; 2019. p. 3342–3349. https://doi.org/10.18653/v1/D19-1329. Available from: https://aclanthology.org/D19-1329.

[89] Plaza-Del-Arco FM, Molina-González MD, Ureña-López LA, Martín-Valdivia MT. A Multi-Task Learning Approach to Hate Speech Detection Leveraging Sentiment Analysis. IEEE Access. 2021;9:112478–112489. https://doi.org/10.1109/ACCESS.2021.3103697.

[90] Schulz C, Eger S, Daxenberger J, Kahse T, Gurevych I. Multi-Task Learning for Argumentation Mining in Low-Resource Settings. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans, Louisiana: Association for Computational Linguistics; 2018. p. 35–41. https://doi.org/10.18653/v1/N18-2006. Available from: https://aclanthology.org/N18-2006.

[91] Hu Y, Huang H, Lan T, Wei X, Nie Y, Qi J, et al. Multi-task Learning for Low-Resource Second Language Acquisition Modeling. In: Wang X, Zhang R, Lee YK, Sun L, Moon YS, editors. Web and Big Data. Cham: Springer International Publishing; 2020. p. 603–611. https://doi.org/10.1007/978-3-030-60259-8_44.

[92] Magooda A, Litman D, Elaraby M. Exploring Multitask Learning for Low-Resource Abstractive Summarization. In: Findings of the Association for Computational Linguistics: EMNLP 2021. Punta Cana, Dominican Republic: Association for Computational Linguistics; 2021. p. 1652–1661. Available from: https://aclanthology.org/2021.findings-emnlp.142.

[93] Biewald L.: Experiment Tracking with Weights and Biases. Software available from wandb.com. Available from: https://www.wandb.com/.