

HULAT-UC3M @ ADoBo 2025: A RoBERTa-based Pipeline for Anglicisms Detection in Spanish Texts

Javier Madrid¹, Paloma Martínez¹ and Lourdes Moreno¹

¹Computer Science and Engineering Department, Universidad Carlos III de Madrid, Leganés, Madrid, Spain

Abstract

This paper details the participation of HULAT-UC3M research group at ADoBo 2025: Automatic Detection of Borrowings at IberLEF 2025. The architecture proposed tested several encoder-based transformers to classify anglicisms, being `xlm-roberta-large-ner-spanish` model the most effective. A pipeline designed to refine predictions by addressing specific error patterns was integrated to improve classification. Experiments were conducted using the COALAS dataset, demonstrating the model's capabilities and the pipeline's impact on performance, achieving a F1-score of 0.9182. The paper also includes some challenges encountered, such as handling words registered by the Spanish Royal Academy (RAE), proper nouns and tokenization artifacts. Finally, potential avenues for future research are outlined, including synthetic data generation and exploration of newer NLP models to further enhance anglicism detection accuracy.

Keywords

Natural Language Processing, anglicisms, Named Entity Recognition, Spanish,

1. Introduction

Language is constantly evolving: including, deleting, and modifying new words and expressions each day, that usually are not registered in a dictionary. In the contemporary globalized world, each country has been influenced by other cultures and languages, leading to a significant influx of borrowings among languages. In the case of anglicisms we are talking about English words or expressions used in another language. The identification and analysis of these borrowed words are crucial for various linguistic and computational applications, such as lexicography, language teaching, machine translation, sentiment analysis and simplification.

This work describes the participation of HULAT-UC3M in ADoBo task [1] at IberLEF 2025 [2] developing an approach based on encoded-based transformers for the automatic detection of anglicisms in Spanish texts. This involves a comprehensive process encompassing dataset creation and preprocessing tasks, model selection and fine-tuning, and the development of a post-processing pipeline to refine the predictions and address specific error patterns.

The following sections will detail the methodology employed, starting with the data preparation and the initial model training. Subsequently, the selection process for the optimal model will be discussed, followed by an in-depth description of the pipeline designed to enhance the model's performance. Finally, the experimentation phase, including the evaluation of the system on different datasets and the iterative improvements made, will be presented, culminating in a discussion of the results and potential avenues for future research.

2. Description of the solution

2.1. Overall architecture and model selection

Figure 1 describes the modules of our solution.

IberLEF 2025, September 2025, Zaragoza, Spain

✉ 100472291@alumnos.uc3m.es (J. Madrid); pmf@inf.uc3m.es (P. Martínez); lmoreno@inf.uc3m.es (L. Moreno)

🆔 0000-0003-3013-3771 (P. Martínez); 0000-0002-9021-2546 (L. Moreno)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

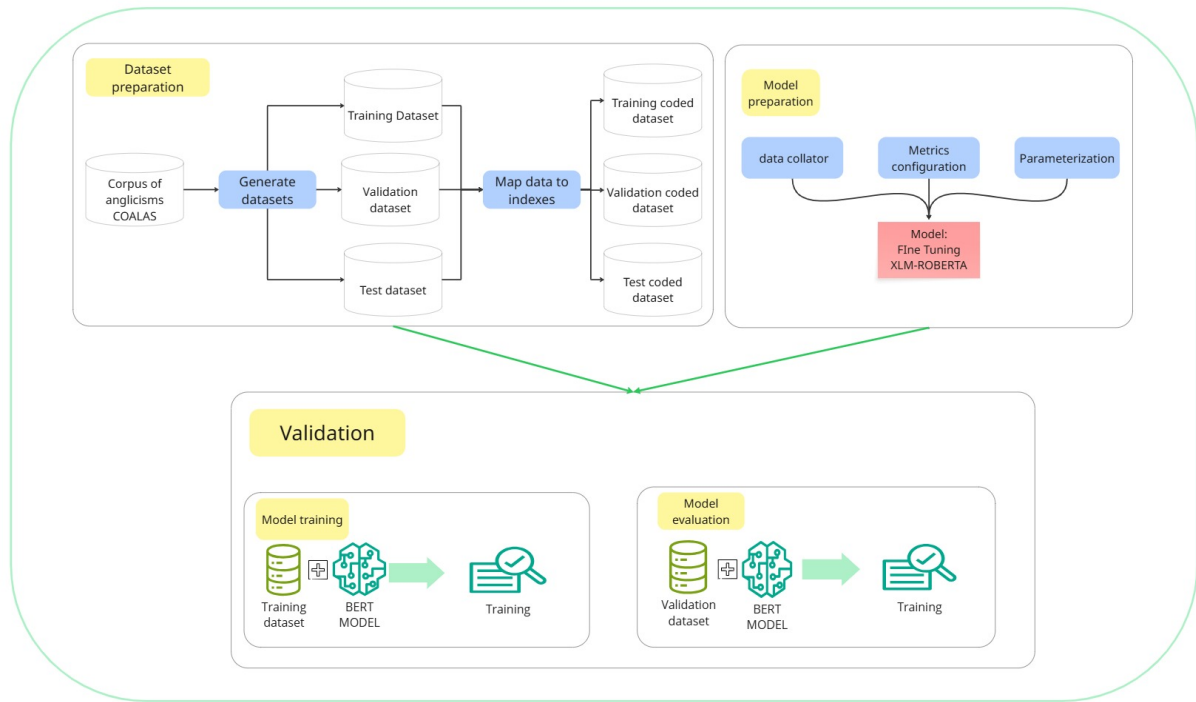


Figure 1: Architecture Overview of the process

- **Dataset preparation:** Using the `datasets` library, a `DatasetDict` is created with the partitions `train`, `validation`, and `test`, each pointing to a GitHub URL with the raw files from the COALAS corpus [3]. Then, each sentence in the dataset is structured as a dictionary with two lists: `tokens` and `tags`. The `tokens` list contains the words or parts of a word that form the corpus, while the `tags` list contains the corresponding labels for each word. Additionally, labels are mapped to indices `{‘O’: 0, ‘B-ENG’: 1, ‘B-OTHER’: 2, ‘I-ENG’: 3, ‘I-OTHER’: 4}`, along with their inverse. A `ner_tags` list is created for each sample with the indexes corresponding to the labels present in `tags`.
- **Model preparation:** The `AutoTokenizer` class from `transformers` is used to load the model to be used. For this task, it was decided to test several discriminative models:
 - `xlm-roberta-large-ner-spanish`: BERT model trained for NER tasks in the Spanish language using the Spanish portion of the CoNLL-2002 dataset [4].¹
 - `roberta-base-bne-capitel-ner-plus`: This is one of the NER task models created by the Barcelona Supercomputing Center by fine-tuning a RoBERTa base model pre-trained using the largest Spanish corpus known to date, compiled from the web crawlings performed by the National Library of Spain (Biblioteca Nacional de España) from 2009 to 2019.²
 - `bert-spanish-cased-finetuned-ner`: This model is a fine-tuned model using the Spanish portion of the CoNLL datasets [4]. The model is based on the Spanish BERT case model (BETO) for NER tasks.³
 - `spanish_bert_based_ner`: A BERT model for NER tasks in Spanish, one of the most famous models for NLP tasks developed by Google, fine-tuned a Spanish dataset from Wikitext [5].⁴
 - `bert-base-spanish-wwm-cased-meddocan`: one of the most recent ones published on Huggingface for the Spanish language, a fine-tuned version of `bert-base-spanish-wwm-cased`

¹<https://huggingface.co/MMG/xlm-roberta-large-ner-spanish>

²<https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne-capitel-ner-plus>

³<https://huggingface.co/mrm8488/bert-spanish-cased-finetuned-ner>

⁴https://huggingface.co/syubraj/spanish_bert_based_ner

for NER tasks using the Meddocan dataset [6] with a F1 score of 0.957.⁵

- **anglicisms-spanish-mbert**: Finally, this model is one of the two already trained with the COALAS dataset [3] used for the same task, so it is necessary to check if this is already the best option.⁶

Each model was trained and tested with the COALAS corpus (see [3]) obtaining precision, recall, and F1-score for tokens labeled with ENG, the anglicisms, and finally again the same three metrics but considering only the tokens that are not anglicisms, labeled with an O for others. The models were optimized using the train and development COALAS dataset, and they were tested using the test dataset. Table 1 shows the distribution of anglicisms in the datasets.

The models were trained using the same hyperparameters: a weight decay of 0.01, a batch size of 16, a learning rate of 2e-5, and 5 epochs.

Table 1

Summary of the .conll files of COALAS datasets

File	Tokens	Sentences	Borrowings	% Borrowings
training	231126	8216	1701	0.74%
evaluation	82578	2025	424	0.51%
test	57997	1811	1671	2.88%

Table 2 displays the results obtained after training and comparing the models.

Table 2

Results of comparing BERT models in anglicisms detection over COALAS dataset.

Model	Total Prec.	Total Rec.	Total F1	Eng. Prec.	Eng. Rec.	Eng. F1	Other Prec.	Other Rec.	Other F1
xlm-roberta-large-ner-spanish (xlm-roberta)	99.50	99.51	99.50	96.20	89.57	92.77	99.66	99.89	99.78
roberta-base-bne-capitel-ner-plus (BSC-LT)	99.12	99.19	99.13	95.75	79.52	86.88	99.34	99.90	99.62
bert-spanish-cased-finetuned-ner (BERT)	98.91	99.01	98.94	91.79	76.30	83.33	99.27	99.82	99.55
spanish-bert-based-ner (BERT)	99.32	99.38	99.34	93.57	88.18	90.80	99.59	99.83	99.71
bert-base-spanish-wwm-cased-meddocan (BERT)	99.21	99.26	99.22	93.62	84.26	88.69	99.50	99.84	99.67
anglicisms-spanish-mbert (ADoBo 2021)	98.45	98.58	98.43	90.80	61.01	72.98	98.78	99.81	99.29

As can be seen in the results, the **xlm-roberta-large-ner-spanish** model stands out as the most effective overall, achieving the best performance in three of the evaluated metrics, in addition to obtaining excellent results in the others.

- **Best F1**: **xlm-roberta-large-ner-spanish** with a value of **0.9951**.
- **Best overall F1**: **xlm-roberta-large-ner-spanish** with a value of **0.9950**.
- **Best F1 for anglicisms**: **xlm-roberta-large-ner-spanish** with a value of **0.9277**.

2.2. Postprocessing

Although the model is capable of correctly detecting a large number of anglicisms, there are cases that it does not cover correctly, so a pipeline has been developed to amend these errors.

The goal of postprocessing is updating tags based on whether a given word or phrase is already registered in the dictionary of the Spanish Royal Academy (RAE). This is because even if a word comes from a language other than Spanish, the moment it is registered in the RAE, the word is not considered an anglicism anymore. This can happen for reasons such as the word having been used in our language for decades, as is the case with *whiskey* or *jazz*. In this way, various false positives that may be obtained from the model are resolved.

As next step *spaCy model* [7], trained for NER tasks in the Spanish language is used to discard false positive cases due to more specific cases not contemplated by the model, such as proper nouns, names referring to political institutions or languages, digits, etc., not being counted as anglicisms.

⁵<https://huggingface.co/IIC/bert-base-spanish-wwm-cased-meddocan>

⁶<https://huggingface.co/lirondos/anglicisms-spanish-mbert>

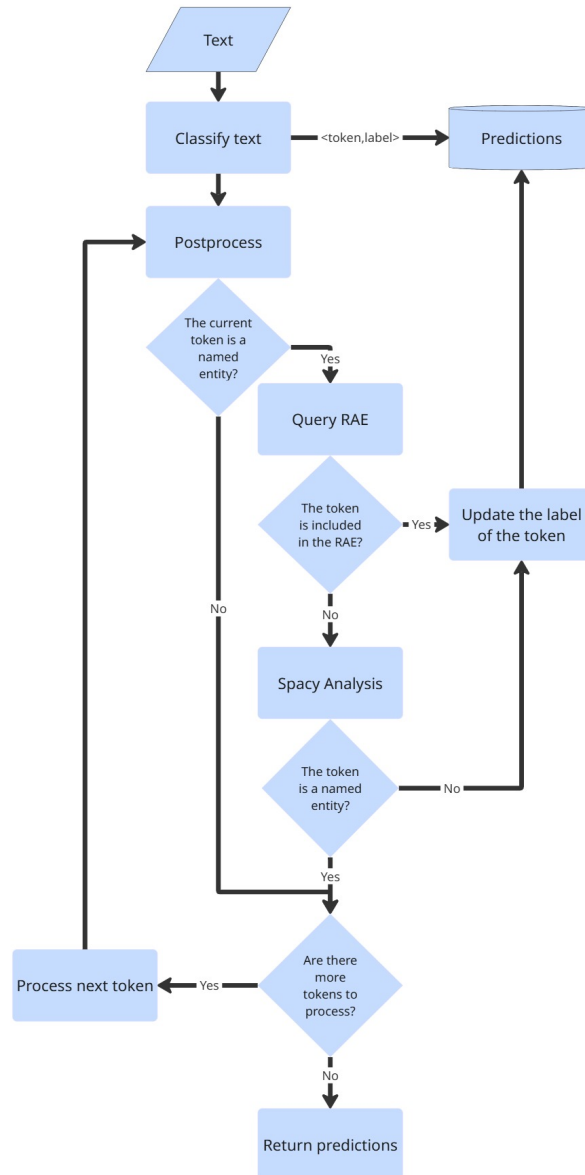


Figure 2: Flow diagram of postprocessing

3. Experimentation

After adapting the RoBERTa model and developing the postprocessing, a series of tests were carried out to analyze the results. For these tests, the COALAS test dataset was used, which had already been used to obtain the model's metrics. Table 3 shows the results just using the `xlm-roberta-large-ner-spanish` model.

As can be observed, these are good results, with an F1 score above 90%. Even so, it can be improved. Among the false positives obtained, there are different types of errors: There are words that the model simply interprets as anglicisms when they turn out to be Spanish, because they are anglicisms that have been adapted to Spanish. Some examples are *cripto*, derived from the anglicism *crypto*, or *whiskazo*, a verb formed from the word *whiskey*. There are also cases of words that, although they technically come from a foreign language, are listed in the RAE, so they do not count as anglicisms, such as *videoclips* or *sake*.

Other noteworthy cases not labeled as anglicisms following annotation rules [8] are *melting* in the sentence "Investigadores del grupo LaserON de la Escuela de Ingeniería Industrial de la Universidade de

Table 3

xlm-roberta-large-ner-spanish model performance on COALAS test set

Metric	Value
Accuracy	99.58
Precision	96.60
Recall	88.97
F1 Score	92.63
True Positives	1533
False Positives	54
True Negatives	56220
False Negatives	190

Vigo han desarrollado un proceso de fibrado continuo por fusión láser (Cofiblas por sus siglas en inglés: Continuous fiberizing by laser melting) que combina un láser de alta potencia y un chorro supersónico de aire para , calentando y moldeando el material de partida , conseguir fibras de vidrio continuas" since words or sequences in languages other than Spanish that literally reflect what someone said or wrote (as in a quote, a statement, or a slogan) will not be considered a borrowing, as is this case. Another sentence to highlight in which there are 9 errors is "Ahora la producción de almacenamiento masivo ya se está produciendo en Samsung o IBM – que llevó al mercado la magnetorresistencia gigante en 1997 – con los STT - RAM (spin - transfer torque magnetic random - access memory) que están basados en principios de la espintrónica, en particular sobre fenómenos cuánticos como es el efecto túnel cuántico que depende del espín, una propiedad cuántica de los electrones" since all tokens composing the words *spin - transfer torque magnetic random - access memory* are detected as anglicisms, but they should not be.

Finally, the case of the word *sms* detected as an anglicism, and the digits 3, 25, and 10 are also incorrectly highlighted as anglicisms, although these cases should be resolved after using the pipeline. In the case of false negatives, we have anglicisms that the model has not recognized, such as *webcams*, *SPOILERS*, or *workaholic*. Table 4 shows the results integrating the postprocessing.

Table 4

xlm-roberta-large-ner-spanish model performance and postprocessing on COALAS test set

Metric	Value
Accuracy	98.46
Precision	97.69
Recall	49.16
F1 Score	65.41
True Positives	847
False Positives	20
True Negatives	56254
False Negatives	876

Comparing these metrics with those obtained using only xlm-roberta-large-ner-spanish model, a clear decrease in both recall and F1 score is observed. This is mainly because, although the number of false positives has more than halved, the number of false negatives has almost quadrupled.

Regarding false positives, the pipeline has correctly eliminated some cases, such as the digits 3 and 10, or words that do not count as anglicisms because they are recognized by the RAE, such as *videoclips* or *sake*. However, new false negatives have appeared, which can be divided into two types: those originating from spaCy, and those originating from querying RAE dictionary. To better appreciate which errors belonged to each stage of the pipeline, two new tests were conducted in which the model and only querying RAE dictionary or the spaCy model were used.

After an analysis of the results obtained, on the one hand, it was observed that although spaCy has a category for proper nouns, this category includes both proper and common nouns. Because of this, numerous nouns that were initially recognized as anglicisms cease to be so after being processed by

spaCy, such as *ecommerce* or *street*. To solve this problem, an effective method to distinguish between common and proper nouns must clearly be sought.

Regarding RAE dictionary there are numerous anglicisms that appear in the dictionary although they have a different meaning. Some cases that can be highlighted from the errors obtained are, for example, *late* is an anglicism, which translated means *tarde*, but can also be understood in Spanish as a conjugation of the verb *latir*.

A possible solution that was considered was to search based on the lemma instead of the word itself but some are not recognized in RAE dictionary. Additionally, spaCy model did not correctly recognize the lemma in the case of anglicisms. This problem was solved by changing to the spaCy model trained for English, which obtained the lemma correctly and solved the problem. Although some cases were resolved, the number of errors was still higher than using the model alone, as there were anglicisms whose lemma was equal to its Spanish counterpart, such as the word *me*, a pronoun in English and Spanish, so it is only detected based on context. In the end, it was decided to discard querying RAE dictionary.

Finally, we decided to test the pipeline using only spaCy's digit and email analysis, which resulted in a reduction of false positives. Even so, it was decided to analyze the results obtained more deeply, but this time, obtaining the model errors based on the number of tokens that made up the anglicism. Table 5 shows the analysis.

Table 5
Summary of errors in anglicism detection

Type of anglicism	False Positives	False Negatives
1-word anglicisms	35	150
2-word anglicisms	14	22
3-word anglicisms	2	13
4-or-more-word anglicisms	3	0

We performed an analysis of the distribution of anglicisms in the dataset:

- There are 917 single-word anglicisms, so almost a quarter are not detected as anglicisms. It is also noteworthy that there is a higher number of false positives in single-word anglicisms compared to other cases.
- In the case of two-word anglicisms, there are 225 in the dataset, so proportionally, the number of false negatives is approximately 10%, while in the case of false positives, there are more than half the errors compared to the number of false negatives. One of the reasons discovered to explain this result is that there are anglicisms composed of multiple tokens, which are only partially detected. For example, for the anglicism *endless - runner*, composed of three tokens, *endless* and *-* are detected as parts of the anglicism, but *runner* is not, which causes not only a false positive but also a false negative, since an error has been detected for the same anglicism because it is incomplete, and another for not detecting the entire anglicism. Some of the false negatives are anglicisms that are partially detected and other false positives such as *Future - ready* and *look working girl*. The first error is a company name in the sentence: *La nueva colección Future - ready de Bershka hace aún más apetecible el streetwear*, so although the words that make up the name are not from our language, it does not count as an anglicism. The second case is a bit more complicated. *Look working girl* are two anglicisms, *Look* and *working girl*, as they come from the sentence *Tamara Falcó apuesta por un look working girl con chaleco en El Hormiguero*, so the error is that the model detects these three words as if they belong to the same anglicism, although the reality is that they are two different anglicisms, causing a false positive and two false negatives for the two undetected anglicisms.
- Although there are 88 3-token anglicisms in the dataset, the number of errors is drastically reduced. Although proportionally, there is a higher number of 3-token false negatives than 2-token false negatives, we only find two false positives.

- Finally, in the case of 4-or-more-token anglicisms, there are only 3 false positives and no false negatives. Two of the false positives are because, although the errors are composed of foreign words, they are not anglicisms due to grammatical rules, and the other because two anglicisms together have been detected as a single anglicism.

After this exploration of false positives and false negatives, we could observe how the highest percentage of false positives and false negatives are found in single-token anglicisms, while the percentage of errors is reduced in the other cases. Also noteworthy are those errors due to the partial detection of some anglicisms, although for the moment, no possible solution is contemplated by modifying the postprocessing to solve this problem; rather, the model itself would have to be improved.

Once the experiments were finished with the COALAS datasets, the datasets given for the task in the development phase were used. Table 6 shows the results using the roBERTa model and postprocessing phase.

Table 6

Results of the configurations using xlm-roberta-large-ner-spanish on development and test datasets

Dataset Phase	Precision	Recall	F1-score
Development dataset results	28.02	19.66	23.11
Development dataset results (with pre-processing)	88.17	78.31	82.94
Development dataset results (with post-processing)	91.57	81.02	85.97
Final test dataset results	83.05	33.04	47.28
Final Test dataset results (with normalization)	96.20	87.86	91.84

As can be observed, if we compare these results with those obtained using the COALAS test dataset, they are considerably worse. After analyzing the results, it was discovered that this was mainly due to two reasons:

- There were numerous false positives and false negatives due to the difference in formats. The developed pipeline had been created to work with the CoNLL format, so when creating the CSV, the anglicisms were copied into the CSV separated by tokens, and not based on the entire anglicism. This problem was solved
- Problems detecting anglicisms composed of multiple words, but for those of only one word, they had been evaluated correctly. Something that caught attention with these errors was that a large number of the false negatives obtained were in quotation marks. For example, in the sentence: "Somos un país en el que 'youtubers' y 'gamers' millonarios deciden irse al paraíso fiscal de Andorra porque tributan menos sin preocuparse del bienestar de sus vecinos y de quienes les han hecho ricos, ni acordarse de la Educación, Sanidad, infraestructuras de las que han disfrutado durante años gracias a la solidaridad de todos", the anglicisms *youtubers* and *gamers* were not detected, unless the quotation marks were removed from the sentence. Because of this, a preprocessing step was added to the pipeline which would remove all single and double quotation marks from the text before passing it to the model.

After all these changes, the pipeline was used again and Table 6 shows the new results. A great improvement can be observed compared to previous tests, although multiple false negatives which contained some punctuation marks were found. Some examples are *poetry slam?*, *dance cover*), and *deepfakes*:. As can be seen, although the anglicisms were technically detected correctly, a tokenizer error led to the mistakes. To solve this, a post-processing stage was added to the pipeline, in which, for each token detected as an anglicism, punctuation marks would be removed if present in the token.

In the rest of the errors, no common characteristics or mistakes were detected that could be solved by modifying the postprocessing phase. Table 6 shows the best results.

Table 6 shows the results obtained with the final test dataset provided by the task organizers.

These results are likely due to the format of the sentences in the dataset, as there are multiple repeated sentences in different forms, for example:

- ""Eyeliner"" a todo color para para un verano fantástico"
- Eyeliner a todo color para para un verano fantástico
- ""eyeliner"" a todo color para para un verano fantástico"
- ""Eyeliner"" A Todo Color Para Para Un Verano Fantástico"
- ""EYELINER"" A TODO COLOR PARA PARA UN VERANO FANTÁSTICO"
- ""Eyeliner"" a todo color para para un verano fantástico"
- ""EYELINER"" a todo color para para un verano fantástico"
- eyeliner a todo color para para un verano fantástico
- Eyeliner A Todo Color Para Para Un Verano Fantástico
- EYELINER A TODO COLOR PARA PARA UN VERANO FANTÁSTICO
- Eyeliner a todo color para para un verano fantástico
- EYELINER a todo color para para un verano fantástico

To solve this problem, it was decided to normalize the text before being used by the model, as the differences in the sentences seemed to affect the results obtained. The normalization consisted of eliminating the single and double marks from the text, and converting uppercase letters to lowercase. Finally, the table 6 shows the best metrics obtained.

4. Conclusions

After multiple tests, a model with an F1 score greater than 90% for predicting anglicisms has finally been obtained. Although it is a good score, it does not mean that a better model cannot be obtained. Firstly, as we have tested, increasing the training data improves the model's effectiveness. For this project, data sources provided by the 2021 and 2025 ADoBo editions tasks have been used, but other options were considered to increase the available data.

One option that was contemplated was to generate synthetic data. The idea was to use a generative artificial intelligence which, through various prompts, would generate sentences containing an anglicism. The sentences would not need to be true, only syntactically meaningful, which AI does excellently, so we could pass it a series of anglicisms to construct new sentences. These anglicisms could be obtained from various data sources. For this project, multiple alternative data sources were investigated in case they were needed:

- Anglicor, a database of anglicisms in the Spanish language [9].
- Study on the compilation of a corpus of anglicisms present in the Spanish press in the 21st century [10].
- Analysis of new lexical anglicisms in the Spanish language in the context of digital academic works and corpora [11]
- Reference Corpus of the Spanish Language (CREA) [12]
- Dictionaries of anglicisms [13][14].

Another possible option would be to change the model used. In the ADoBo 2021 task, the best model obtained an F1 score of **84.8**, while the model chosen for this project, **xlm-roberta-large-ner-spanish**, has surpassed it. This project was last modified in 2023, two years after the ADoBo 2021 task. New, larger, and better-optimized models appear every day, so it would not be surprising if, in the future, a model appears that surpasses the one used.

Finally, one could also continue iterating on the pipeline. Although several ideas for improving it have been discarded, there are surely other options that would allow the model to improve its results, such as the ability to discard anglicisms that are proper nouns, or to discard those anglicisms that have been adapted to Spanish, such as *cripto* or *craftear*.

In the end, there are multiple options to continue exploring, and with sufficient research and work, it will surely be possible to continue obtaining better results in the task of anglicism detection.

Acknowledgments

This work has been supported by grant PID2023-148577OB-C21 (Human-Centered AI: User-Driven Adapted Language Models-HUMAN_AI) by MICIU/AEI/ 10.13039/501100011033 and by FEDER/UE.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 in order to: grammar and spelling check and to translate correctly some sentences in the text. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] E. Álvarez-Mellado, J. Porta-Zamorano, C. Lignos, J. Gonzalo, Overview of ADoBo at IberLEF 2025: Automatic Detection of Anglicisms in Spanish, *Procesamiento del Lenguaje Natural* 75 (2025).
- [2] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025)*, co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.
- [3] E. Álvarez-Mellado, C. Lignos, Detecting unassimilated borrowings in Spanish: An annotated corpus and approaches to modeling, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 3868–3888. URL: <https://aclanthology.org/2022.acl-long.268/>. doi:10.18653/v1/2022.acl-long.268.
- [4] E. F. Tjong Kim Sang, Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition, in: *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*, Association for Computational Linguistics, Taipei, Taiwan, 2002, pp. 155–158. URL: <https://aclanthology.org/W02-2024>.
- [5] A. Rahimi, Y. Li, N. Dodevska, S. Ruder, Massively multilingual transfer for ner, *arXiv preprint arXiv:1902.00193* (2019).
- [6] M. Marimon, A. Gonzalez-Agirre, A. Intxaurre, H. Rodríguez, J. A. Lopez Martin, M. Villegas, M. Krallinger, Automatic de-identification of medical texts in spanish: the meddocan track, corpus, guidelines, methods and evaluation of results, in: *IberLEF@SEPLN, 2019*, pp. 618–638. URL: <https://temu.bsc.es/meddocan/>. doi:10.5281/zenodo.4279323.
- [7] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python (2020). doi:10.5281/zenodo.1212303.
- [8] E. Álvarez Mellado, Annotation Guidelines for Lexical Borrowings: With a Focus on Unassimilated Anglicisms in Spanish Newswire, ADoBo Shared Task at IberLEF 2021, 2021. URL: <https://adobo-task.github.io/docs/guidelines.pdf>, accessed: 2025-05-27.
- [9] I. d. I. Cruz Cabanillas, C. Tejedor Martínez, L. d. Marcos Ortega, Criterios para el diseño de una base de datos de préstamos ingleses en el español peninsular contemporáneo, in: *A survey of corpus-based research*, 1 ed., 2009, pp. 741–755. URL: <http://www.um.es/lacell/aelinco/contenido/pdf/50.pdf>.
- [10] E. E. Núñez Nogueroles, A corpus-based study of anglicisms in the 21st century spanish press, *Analecta Malacitana (AnMal electrónica)* (2018) 123–159. URL: <https://dialnet.unirioja.es/descarga/articulo/6633735.pdf>.
- [11] D. Folqués, Análisis de los nuevos anglicismos léxicos en la lengua española en el contexto de las obras y corpus académicos digitales, *Texto Livre: Linguagem e Tecnologia* 14 (2020) e24418. doi:10.35699/1983-3652.2021.24418.
- [12] R. A. Española, Corpus de referencia del español actual (crea), 2023. URL: <https://www.rae.es/banco-de-datos/crea>, accessed: 2025-05-27.

- [13] M. P. Castillo Bernal, Diccionario de anglicismos del español estadounidense, by francisco moreno-fernández, ed. instituto cervantes at the faculty of arts and sciences of harvard university, 2018. informes del observatorio / observatorio reports. 037-01/2018 sp, isbn: 978-0-692-04726-2 doi: 10.15427/or037-01/2018sp, Transletters. 1 (2018) 195–196. URL: <https://www.uco.es/ucopress/ojs/index.php/tl/article/view/11042>.
- [14] D. S. Iglesias, Diccionario de anglicismos y otros extranjerismos, Defensa del Idioma Español, España, 2018. URL: <https://defensadelidioma.com/wp-content/uploads/2018/09/diccionario-de-anglicismos-y-otros-extranjerismos.pdf>, accessed: 2025-05-27.