Automatic Contradiction Detection in Spanish

Robiert Sepúlveda-Torres

Department of Software and Computing Systems, University of Alicante, Apdo. de Correos 99, E-03080 Alicante, Spain

Abstract

This paper addresses the lack of automated contradiction detection systems for the Spanish language. The ES-Contradiction dataset was created and contains examples with two pieces of information classified as **Compatible**, **Contradiction**, or **Unrelated**. To the author's knowledge, a Spanish-language contradiction dataset is non-existent and therefore, the ES-Contradiction dataset fills an important research gap, given Spanish being one of the most widely spoken languages. Moreover, the dataset built includes a fine-grained annotation of the different types of contradictions in the dataset. A baseline system was designed to validate the effectiveness of the dataset. The BETO transformer model was used to build this baseline system, which obtained a good result to detect the three class labels **Compatible**, **Contradiction**, or **Unrelated**.

Keywords

contradiction dataset, contradiction detection, natural language processing, deep learning, fake news

1. Introduction

A huge amount of fake news is generated and distributed by digital media every day. Hence, the manual evaluation of its veracity is practically impossible in a reasonable time frame [1].

Fake news has existed for a long time [2], but the term "fake news" is relatively new, and it was defined by The New York Times as a "made up story with the intention to deceive, often with monetary gain as a motive" [3]. Artificial Intelligence techniques have been used in recent years to tackle the fake news problem [4, 5]. When a news item is misleading, it introduces contradictory information to a true news item and therefore, the detection of contradictions is a fundamental task to identify fake news [6]. One of the open problems raised by previous research [6] related to contradiction detection was to expand the search for resources (dataset, models, systems) that would allow the creation of contradiction detection systems in other languages or with cross-lingual approaches. This paper is a summary of a key section of my PhD whose aim is to design a generic architecture for fake news detection.

For this purpose, a deep search for resources was carried out on contradiction detection from which we concluded that most of the resources and systems for contradiction detection are developed in English [7, 8, 9, 10]. Moreover, despite Spanish being one of the most widely spoken languages in the world, there are no powerful resources in the Spanish language to carry out the task of detecting contradictions. To fill this research gap and address the lack of

resources in the Spanish language, the main contributions of this research are the following:

- First, a new Spanish dataset is built with different types of compatible, contradictory, and unrelated information for the purpose of creating a language model that is capable of automatically detecting contradictions between two pieces of information in this language.
- Second, in a new Spanish dataset each contradiction is annotated with a fine-grained annotation, differentiating between different types. Specifically, four of the types of contradictions defined in [11] are covered: antonymy, negation, date/number mismatch, and structural.
- Third, a set of experiments using the BETO [12] pretrained model has been applied to build the language model and validate its effectiveness.

2. Related Work

The research presented in this paper is focused on contradiction detection by using computational models.

The most common approaches to contradiction detection in texts use the linguistic features extracted from texts to build a classifier by training from the annotated examples, such as the works in [13, 11, 14]. Linguistic evidences such as polarity, numbers, dates and times, antonymy, structure, factivity, and modality features were used by the authors of [11] to detect contradiction. Simple text similarity metrics (cosine similarity, f1 score, and local alignment) were used as baseline in [14], obtaining good results for contradiction classification. [13] tackled contradictions by means of three types of linguistic information: negation, antonymy, and semantic and pragmatic information associated with discourse relations.

Currently, the availability of large annotated datasets for contradiction detection are mainly present in English [15], such as SNLI [7], MultiNLI (including multiple genres) [8], or even the cross-lingual dataset XNLI [16]. These datasets have allowed the training of complex deep learning systems, which require very large corpora to obtain successful results.

There are a few studies that address the detection of contradictions in languages other than English, such as:

1. Machine translation of SNLI from English into German. A model was built using the German version of SNLI and the results of the predictions are very similar to the same model trained on the original SNLI version in English [15].
2. A large-scale database of contradictory event pairs in the Japanese language has been created. This database is used to generate coherent statements for a dialogue system [17].
3. Baseline systems are introduced to detect contradictions in the Persian language. Automatic translation into the Persian language of SNLI and MNLI dataset samples was performed [18].

3. ES-Contradiction: A New Spanish Contradiction Dataset

Our dataset (ES-Contradiction) is focused on contradictions that are likely to appear in traditional news items written in the Spanish language. Unlike other datasets, for the dataset proposed in

this work, contradictions are annotated by distinguishing the type of contradiction according to its specific characteristics [19].

In order to create the ES-Contradiction dataset, news articles from a renowned Spanish source were automatically collected, including the headline and body text. According to the journalistic structure of a news item, the headline is the title of the news article, and it provides the main idea of the story. A headline is expected to be as effective as possible, without losing accuracy or becoming misleading [20]. Therefore, finding contradictions between headlines and body texts is a crucial task in the fight against the spread of disinformation.

The ES-Contradiction dataset is annotated according to the following classes:

- **Compatible** (two pieces of texts that address the same fact contain compatible information at the same time frame)
- **Contradiction** (two texts that address the same fact contain incompatible information or two texts that address antonym facts contain compatible information at the same time)
- **Unrelated** (two texts that address different facts)

3.1. Dataset Annotation Stages

The dataset was built in four stages, subsequently outlined and detailed:

1. **Extracting information from data source:** The headline, body text, and date of the news item are extracted from a reliable data source. In this case, the news agency EFE
2. [1] . The news are extracted assuming that the headlines and body texts are compatible, although in the third stage this relationship is verified.
3. **Modifying news headlines**: The aim of this stage is to make the news headline contradictory to the body text by including simple alterations to the headline structure. These alterations will change the semantic content of the sentence, making it contradictory to the previous headline and body text. The changes to the headline together with some examples are given as follows:
   - NEGATION (Con_Neg): This alteration consists of negating the headline of the news item.
     a) Original headline: *"El Gobierno <u>pide</u> permanecer en los domicilios porque la situación <u>es</u> grave"* ("The government <u>asks</u> people to stay home as the situation <u>is</u> serious")
     b) Modified headline: *"El Gobierno <u>no pide</u> permanecer en los domicilios porque la situación <u>no es</u> grave"* ("The government <u>doesn't ask</u> people to stay home as the situation <u>isn't</u> serious")
   - ANTONYM (Con_Ant): This transformation consists of replacing the verb denoting the main event of the headline with an antonym.
     a) Original headline: *"Las licencias VTC <u>aumentan</u> un 19% en lo que va de año"* ("VTC licences <u>increase</u> by 19% this year")
     b) Modified headline: *"Las licencias VTC <u>caen</u> un 19% en lo que va de año"* ("VTC licences <u>fall</u> by 19% this year")

---

[1]https://www.efe.com/efe/espana/1, accessed on 15 June 2021

- NUMERIC (Con_Num): This amendment consists of changing numbers, dates, or times appearing in the headline.
  a) Original headline: *"Accionistas de Nissan aprueban la incorporación de <u>cuatro</u> nuevos consejeros"* ("Nissan shareholders approve the appointment of <u>four</u> new board directors")
  b) Modified headline: *"Accionistas de Nissan aprueban la incorporación de <u>diez</u> nuevos consejeros"* ("Nissan shareholders approve the appointment of <u>ten</u> new board directors"')
- STRUCTURE (Con_Str): This modification consists of changing the position of one word for another or substituting words in the sentence.
  a) Original headline: *"Tokio avanza optimista por el diálogo entre <u>EEUU</u> y <u>China</u>"* ("Tokyo is optimistic about the dialogue between <u>USA</u> and <u>China</u>")
  b) Modified headline: *"Tokio avanza optimista por el diálogo entre <u>EEUU</u> y <u>UE</u>"* ("Tokyo is optimistic about the dialogue between <u>USA</u> and <u>EU</u>")

4. **Classifying the relationship between the headline and the body text**: The semantic relationship between the headline and the body text was annotated in two phases: The first phase consisted of classifying the information into **Compatible** (compatible information) or **Contradiction** (contradictory information). In the second phase, in the case of Contradiction, the type of contradiction was also annotated (Negation, Antonym, Numeric, Structure).

5. **Aleatory mixing headline and body text**: The news items reserved in the first stage were used to generate unrelated examples (**Unrelated**). The headline was separated from the corresponding body text and all the headlines were randomly mixed with the body texts. In the mixing process, it was verified that the headline is not mixed with the corresponding body text. This step was done automatically without the intervention of the annotators.

3.2. Dataset Description

The dataset consists of 7403 news items, of which 2431 contain **Compatible** headline–body news items, 2473 contain **Contradiction** headline–body news items, and 2499 are **Unrelated** headline–body news items. This represents a balanced dataset with three main classification items. The dataset split sizes for each annotated class are presented in Table 1. We partitioned the annotated news items into training and test sets. The dataset is available at Zenodo[2].

As can be seen in Table 2, our dataset contains examples of each type of contradiction. However, it is important to clarify that there are few examples of structure contradiction, given the complexity of finding sentences that allow for this type of modification.

3.3. Dataset Validation

Due to the particularities of the dataset annotation process, it was necessary to validate the second and third stages of the process. For the second stage, a super-annotator validation

---

Table 1

Dataset split sizes for each class.

Split	Compatible	Contradiction	Unrelated
Training	1703	1733	1755
Test	728	740	744
Total items	2431	2473	2499

Table 2

Contradiction types in the dataset.

Split	Con-Neg	Con-Ant	Con-Num	Con-Str
Training	674	552	430	77
Test	287	236	184	33
Total items	961	788	614	110

was conducted, while for the third stage, an inter-annotator agreement was carried out. We randomly selected 4% of the **Compatible** and **Contradiction** pairs (n = 200) to carry out the dataset validations.

3.3.1. Super-Annotator Validation

For the second stage, it was not possible to make an inter-annotator agreement because this stage consists of headline modifications and the possible variations are infinite. In this case, a manual review of the modified headlines is performed by the Super-Annotator to detect inconsistencies with the indications in the annotation guide. Only 2% of the analyzed examples present inconsistencies with the annotation guide, corroborating the validity of this stage.

3.3.2. Inter-Annotator Agreement

In order to measure the quality of the third stage annotation, an inter-annotator agreement between two annotators was performed. In cases where there was no agreement, a consensus process was carried out among the annotators. Using Cohen's $kappa$ [21] a $k = 0.83$ was obtained, which validates the third-stage labeling.

4. Experiments and results

To test the validity of the newly created Spanish contradiction dataset in this task, a baseline system was created that is based on the BETO[3] model described in [12] that was previously pretrained in a Spanish dataset. The system was implemented using the Simple Transformer[4] and

---

[3]https://github.com/dccuchile/beto, accessed on 22 March 2021
[4]https://simpletransformers.ai/, accessed on 15 June 2021

PyTorch[5] libraries. In our experiments, the hyperparameter values of the model are maximum sequence length of 512, batch size of 4, training rate of 2e-5, and training performed for 3 epochs.

The main objective of the experimentation proposed in this research is to demonstrate that a model is able to learn how to automatically detect contradiction types and contradictions with high accuracy from the ES-Contradiction dataset.

4.1. Predicting all classes

This experiment is performed on the entire dataset to predict **Compatible**, **Contradiction**, or **Unrelated** for each example. The system created is capable of detecting the Unrelated class with a high level of precision and achieves significantly good results in the **Compatible** and **Contradiction** classes. Table 3 presents the results.

Table 3
Results obtained from Experiment 1: Predicting **Compatible**, **contradiction**, or **Unrelated** information.

	F_1 **Score (%)**			F_1m **(%)**	**Acc (%)**
System	Compatible	Contradiction	Unrelated		
BETO-All classes	88.70	89.12	99.59	92.47	92.49

The results obtained in the Unrelated class indicate that the system is capable of detecting with excellent F_1m these types of examples, corroborating the results obtained in the literature on this type of semantic relation between texts [22]. The other two classes have room for improvement, by using, for instance, external knowledge.

4.2. Detecting Contradiction vs. Compatible Information

In this experiment, the Unrelated class is removed from the ES-Contradiction dataset to measure the accuracy of the approach in terms of distinguishing between compatible or contradiction information, assuming that the information is related. The results are shown in Table 4. The approach obtains similar results in both predicted classes. This is due to the quality of the training examples and the balanced number of examples from each class in this dataset.

Table 4
Results obtained from Experiment 2: Detecting between compatible and contradiction information when the texts are related.

	F_1 **Score (%)**		F_1m **(%)**	**Acc (%)**
System	Compatible	Contradiction		
BETO-Contra_Comp	88.63	88.75	88.69	88.69

---

[5]https://pytorch.org/, accessed on 15 June 2021

4.3. Detecting Specific Types of Contradictions

This experiment aims to analyze the detection capability of the approach by contradiction types. Table 5 shows the results obtained exclusively for the detection of contradiction types.

Table 5
Results obtained from Experiment 3: Detecting each specific type of contradiction treated.

	F_1 **Score (%)**				$F_1 m$ **(%)**	**Acc**
System	Con-Neg	Con-Ant	Con-Num	Con-Str		
BETO-Type of contradictions	97.90	93.20	92.39	68.75	88.06	93.78

The structural Contradiction class (Con_Str) is the one that obtains the lowest accuracy results and $F_1 m$. This contradiction type is considered one of the most complicated contradictions to detect compared with the other contradictions [11], which is in line with our results. In addition, the Con_Str class, due to the scarcity of training examples, contains the lowest number of examples in this dataset, so the model can learn more about other more representative classes.

5. Discussion

The experiments carried out have validated the effectiveness of a contradiction detection system using the ES-Contradiction dataset. However, the experiments evidenced one of the deficiencies in the ES-Contradiction dataset namely, the lack of examples of the Con_Str class. It would also be interesting to include other types of contradictions that are not being taken into account in this dataset.

For the purpose of improving the results of the Contradiction class, we can test by including resources that detect antonyms and synonyms in line with [23]. Furthermore, including syntactic and semantic information could improve the detection of other more complex contradictions, such as structural ones, without the need for such large datasets.

It is highly likely that contradictions such as the structure contradiction need external semantic knowledge to improve detection results, similar to the introduction of SRL [10] and the use of Wordnet relations [9], both of which improve the results of deep learning models.

6. Conclusions

This work has built the ES-Contradiction dataset, a new Spanish language dataset that contains **Contradiction**, **Compatible**, and **Unrelated** information. Unlike other datasets, in the ES-Contradiction dataset, contradictions are annotated with a fine-grained annotation. We used the BETO model to create an automated contradiction detection system.

The results obtained by our system show that the created Spanish contradictions dataset is a good option for generating a language model that is able to detect contradictions in the Spanish language. In order to create a powerful contradiction detection system in Spanish, it is necessary to extend our dataset with other types of contradictions and add specific features.

Finally, the creation of this dataset will make it possible to validate the effectiveness of a contradiction detection architecture in the Spanish language that will be created in future works.

Acknowledgments

References

[1] G. Tsipursky, F. Votta, K. M. Roose, Fighting Fake News and Post-Truth Politics with Behavioral Science: The Pro-Truth Pledge, Behavior and Social Issues 27 (2018) 47–70. URL: https://doi.org/10.5210/bsi.v27i0.9127. doi:10.5210/bsi.v27i0.9127.

[2] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, Journal of economic perspectives 31 (2017) 211–36.

[3] S. Tavernisen, As fake news spreads lies, more readers shrug at the truth, New York Times (2019). URL: http://nyti.ms/2lw56HN.

[4] A. Alonso-Reina, R. Sepúlveda-Torres, E. Saquete, M. Palomar, Team GPLSI. approach for automated fact checking, in: Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 110–114. URL: https://www.aclweb.org/anthology/D19-6617. doi:10.18653/v1/D19-6617.

[5] R. Sepúlveda-Torres, M. Vicente, E. Saquete, E. Lloret, M. Palomar, Exploring summarization to enhance headline stance detection, in: E. Métais, F. Meziane, H. Horacek, E. Kapetanios (Eds.), Natural Language Processing and Information Systems, Springer International Publishing, Cham, 2021, pp. 243–254.

[6] R. Sepúlveda-Torres, Identification of fake news by contradiction detection in texts, CEUR Workshop Proceedings 2802 (2020) 38–43.

[7] S. R. Bowman, G. Angeli, C. Potts, C. D. Manning, A large annotated corpus for learning natural language inference, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, 2015, pp. 632–642. URL: https://www.aclweb.org/anthology/D15-1075.

[8] A. Williams, N. Nangia, S. Bowman, A broad-coverage challenge corpus for sentence understanding through inference, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 1112–1122. URL: https://www.aclweb.org/anthology/N18-1101.

[9] Q. Chen, X. Zhu, Z.-H. Ling, D. Inkpen, S. Wei, Natural language inference with external knowledge, arXiv preprint arXiv:1711.04289 42 (2017) 43.

[10] Z. Zhang, Y. Wu, H. Zhao, Z. Li, S. Zhang, X. Zhou, X. Zhou, Semantics-aware bert for language understanding, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 9628–9635.

[11] M.-C. de Marneffe, A. N. Rafferty, C. D. Manning, Finding contradictions in text, in: Proceedings of ACL-08: HLT, Association for Computational Linguistics, Columbus, Ohio, 2008, pp. 1039–1047. URL: https://www.aclweb.org/anthology/P08-1118.

[12] J. Canete, G. Chaperon, R. Fuentes, J. Pérez, Spanish pre-trained bert model and evaluation data, PML4DC at ICLR 2020 (2020).

[13] S. Harabagiu, A. Hickl, F. Lacatusu, Negation, contrast and contradiction in text processing, in: Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1, AAAI'06, AAAI Press, 2006, p. 755–762.

[14] P. Lendvai, U. D. Reichel, Contradiction Detection for Rumorous Claims, ArXiv (2016). URL: http://arxiv.org/abs/1611.02588.

[15] R. Sifa, M. Pielka, R. Ramamurthy, A. Ladi, L. Hillebrand, C. Bauckhage, Towards contradiction detection in german: a translation-driven approach, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 2497–2505.

[16] A. Conneau, G. Lample, R. Rinott, A. Williams, S. R. Bowman, H. Schwenk, V. Stoyanov, XNLI: evaluating cross-lingual sentence representations, CoRR abs/1809.05053 (2018). URL: http://arxiv.org/abs/1809.05053.

[17] Y. Takabatake, H. Morita, D. Kawahara, S. Kurohashi, R. Higashinaka, Y. Matsuo, Classification and acquisition of contradictory event pairs using crowdsourcing, in: Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 99–107. URL: https://www.aclweb.org/anthology/W15-0813.

[18] Z. Rahimi, M. ShamsFard, Contradiction Detection in Persian Text (2021). URL: https://arxiv.org/abs/2107.01987v1. arXiv:2107.01987.

[19] R. Sepúlveda-Torres, A. Bonet-Jover, E. Saquete, "here are the rules: Ignore all rules": Automatic contradiction detection in spanish, Applied Sciences 11 (2021) 3060. URL: https://doi.org/10.3390%2Fapp11073060. doi:10.3390/app11073060.

[20] J. Kuiken, A. Schuth, M. Spitters, M. Marx, Effective headlines of newspaper articles in a digital environment, Digital Journalism 5 (2017) 1300–1314.

[21] J. Cohen, A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement 20 (1960) 37.

[22] Q. Zhang, S. Liang, A. Lipani, Z. Ren, E. Yilmaz, From stances' imbalance to their hierarchical representation and detection, in: The World Wide Web Conference, ACM, 2019, pp. 2323–2332.

[23] X. Kang, B. Li, H. Yao, Q. Liang, S. Li, J. Gong, X. Li, Incorporating synonym for lexical sememe prediction: An attention-based model, Applied Sciences 10 (2020) 5996.