From Automatic Metaphor Processing in Spanish to a Multilingual Perspective: Annotation, Systems, and Evaluation

Elisa Sanchez-Bayona

HiTZ Center - Ixa, University of the Basque Country UPV/EHU, Donostia-San Sebastián, Spain

Abstract

There is very few work on Natural Language Processing (NLP) to automatically deal with the metaphorical use in Spanish and those developed so far rely on unsupervised approaches, which obtain a significantly lower performance than supervised systems developed for English. The main reason is the lack of a corpus with wide coverage and large size compared to English resources with metaphor annotations. This thesis project aims to collect and label the first corpus with sufficient magnitude to be able to develop NLP systems for the automatic processing of metaphor in Spanish. Thus, this thesis will analyze the linguistic metaphor present in everyday language from a corpus-based perspective. In order to achieve this purpose, a corpus of wide coverage of texts in Spanish will be gathered and annotated by means of appropriate guidelines. Idiosyncrasies of each language will be taken into account during the application of these guidelines, establishing new annotation criteria when necessary. The corpus will serve as a foundation for the development of metaphor processing systems in Spanish, both in detection and interpretation. Additionally, this thesis will also explore multilingual approaches for both tasks. Finally, an extrinsic evaluation will be considered to analyse the impact metaphor on final NLP tasks such as Machine Translation, Sentiment Analysis, Fact-checking and Hoax detection, as well as opinion mining and argumentative discourse generation.

Keywords

Metaphor, Computational Metaphor Processing, Spanish Metaphor, Metaphor Detection, Metaphor Interpretation, Natural Language Processing

1. Introduction

The use of metaphorical expressions is an ubiquitous phenomenon in our daily utterances. For this reason, its automatic processing through linguistic technology is essential for a large number of real NLP applications, such as Machine Translation (MT) [1], Sentiment Analysis (SA) [2, 3] or tasks that allow subsequent analysis of different types of discourse, such as fact-checking, or the detection of biased articles like hyper-partisan news [4], argumentative discourse [5] or hate speech [6, 7]. For instance, a metaphor not shared between two languages could be identified and translated by a MT system as a literal meaning expression. Likewise, a SA system can benefit from detecting and understanding those metaphorical expressions used to emphasize the valuation of a product. In addition, the presence of metaphorical expressions can help in the

Doctoral Symposium on Natural Language Processing from the PLN.net network 2022 (RED2018-102418-T), 21-23 September 2022, A Coruña, Spain.

elisa.sanchez@ehu.eus (E. Sanchez-Bayona)

analysis and characterization of texts from a specific domain, for example, political speeches, where a greater use of metaphors can imply an increase in the capacity of persuasion towards the public [8].

2. Related Work

Metaphorical use arises, in general terms, when a concept belonging to one domain is expressed in terms of a concept from a different domain. In other words, metaphors are frequently used in natural language to convey abstract concepts or ideas through specific experiences related to the real, physical world or to reinforce the ideas of a discourse, such as in these examples from CoMeta (https://ixa-ehu.github.io/cometa/, currently under development):

1. La *batalla* - ya lo sabemos - va a ser *dura*, pero con disciplina social, con resistencia, con unidad y con moral de victoria lo vamos a volver a lograr. ("The *battle* - we already know it - is going to be *hard*, but with social discipline, with resistance, with unity and with victory morale we are going to achieve it again.").
2. La ley es importante, pero hay otras *armas* aún más eficaces contra el virus. ("The law is important, but there are other even more effective *weapons* against the virus.").
3. Depende de nosotros y nosotras *levantar* un *muro* de unidad que *frene* al virus mientras disponemos de la vacuna que lo *destierre* para siempre. ("It is up to us and us to *build* a *wall* of unity that *stops* the virus while we have the vaccine that will *banish* it forever.").
4. *Grandes* intelectuales que saben muy bien interpretar las palabras. ("*Great* intellectuals who know very well how to interpret words.").
5. Realizan el catering con productos de *alta* calidad. ("They *cater* with high quality products.").

Metaphors are put into words in a wide variety of forms and are classified from different points of view [9], although the most common distinction is that between **conventional** metaphors, already lexicalized, and **novel** metaphors. Both types are used in everyday language. Thus, *hard* in 1 is a conventional metaphor, since hardness, understood as physical resistance, is equated with the difficulty of the circumstances. On the contrary, the metaphorical expressions of 1, 2 and 3, *battle, weapons* are an example of novel metaphors that recount the pandemic perceived as a war, where the virus is the enemy and society the army that must fight and defeat it, through tools such as a *wall of unity*.

The publication of Lakoff and Johnson [10] has been perhaps the most influential theoretical work in establishing that metaphor is not only a rhetorical mechanism of language, but a cognitive-linguistic phenomenon highly common in everyday speech. In this approach (and in its many extensions), metaphor is a conceptual mapping that recasts an entire domain of experience (source) in terms of a different domain (target).

A **conceptual metaphor** can be materialized in natural language through multiple **linguistic metaphors**. The most common types are **lexical metaphors** (examples from 1 to 5), **multiword metaphors**, and **extended metaphors**, which cover larger fragments of speech. In turn, metaphors can be classified according to the grammatical category to which they belong, the most common are **verbal metaphors** (*levantar* ("to build"), *frene* ("to stop") and *destierre* ("to

banish") in 3), **adjectival** (*dura* ("hard") in 1, *grandes* ("large") in 4 and *alta* ("high") in 5) and the **nominal metaphors** (*batalla* ("battle") in 1, *armas* ("weapons") in 2 and *muro* ("wall") in 3).

Automatic processing of metaphor can be divided into three different tasks: (i) **detection** of metaphorical expressions in everyday text, (ii) their **interpretation**, that is, the identification of the literal meaning expressed by the linguistic metaphor, and (iii) the **generation** of new metaphorical expressions. This thesis will focus on the first two points and will be framed within an empirical approach based on real data. In other words, the characterization of the linguistic metaphor in Spanish will be carried out by the compilation of existing texts from various sources and domains. The annotation will follow the MIP methodology [11], later extended to MIPVU [12].

The vast majority of work on metaphor processing has focused on English texts, due to the greater availability of manually annotated data. The most widely used corpus for the characterization of linguistic metaphor is the VU Amsterdam Metaphor Corpus (VUAMC) [12], a labeled dataset in English with several typologies of metaphor based on the VU Metaphor Identification Procedure (MIPVU, for its acronym in English), subsequently adapted to other languages [13]. Regarding Spanish metaphor processing, due to the lack of manually labeled corpora to develop supervised systems, previous work consisted mainly in unsupervised approaches, which present lower performance than the supervised systems developed for English [14, 15]. The few works carried out focused either on a very specific type of conceptual metaphor [16], or on the annotation of datasets bounded to a specific domain and too reduced in size to train systems in detection and interpretation of metaphors [17]. The recent development of CoMeta (https://ixa-ehu.github.io/cometa/) aims to alleviate this shortage, providing the largest dataset of general domain texts with metaphorical annotations in Spanish that, despite not reaching the size of the VUAMC, can serve as a base and be expanded with the collection of more texts.

Regarding detection, there are corpus-based works [18, 19, 20]. The most recent approaches for English address the task as sequential labeling usually based on deep learning, neural networks and word embeddings [21, 22], with a variety of syntactic-semantic features (WordNet, FrameNet , VerbNet, dependency analysis, morphology, etc.). Most notorious improvements derive from the celebration of several evaluation tasks around the detection of metaphors using the VUAMC dataset [23, 24], although top results were achieved at classifying most conventional metaphors [5, 25].

Regarding metaphor interpretation, most successful approaches tackle the task as a paraphrase of the metaphorical expression into its literal counterpart [26, 27, 28], exploiting the existing syntactic-semantic relationships between source and target domains of a metaphorical expression [29]. However, these approaches do not account for the features of the target domain present in the metaphorical expressions. For this reason, metaphor interpretation should take into account the complex role linguistic metaphors play with respect to the communicative intent, for instance, in the scenario of political argumentation [8] or opinion mining, in which metaphor expressions illustrate ideas more clearly and emphasize the message to be conveyed.

In each and every one of the advances in metaphor processing mentioned so far [5], the presence of the Spanish language is remarkably scarce. This thesis project would therefore constitute a first and novel contribution to place metaphor processing in Spanish, one of the most widely spoken languages in the world, at a similar level of development in terms of linguistic resources publicly available. Moreover, it will empirically explore the influence of

metaphorical language applied to tasks with different communicative intents.

3. Research Proposal

The subject of research is novel and ambitious at the same time for several reasons: (i) to the best of our knowledge, currently there are not datasets of wide coverage annotated with linguistic metaphors for Spanish such as the one we propose in this thesis; (ii) the development of the corpus involves characterizing the metaphorical language in everyday Spanish texts. The process requires an adaptation of the MIPVU method [12] originally designed for English, which is a considerable scientific challenge [13]; (iii) there are currently no NLP systems available for automatic detection and interpretation of metaphorical texts in Spanish that achieve similar performance to those trained for English.

The achievement of the main objective of the thesis will have two main benefits. First, the results of the thesis will allow us to better understand the various linguistic mechanisms underlying metaphorical expressions in Spanish. Second, the potential impact of the development of NLP systems for multilingual metaphor detection and interpretation (Spanish and English) on other NLP tasks. Something that has not been thoroughly analysed so far. To achieve these objectives, a series of intermediate tasks and experiments are proposed that will contribute to the general vision of the thesis:

- Apply the MIPVU method pointing out those aspects that could differ from English, based on the work done for other languages [13].
- Annotate a corpus of wide coverage in Spanish, extending the initial work of CoMeta. The labeling process will be carried out among several annotators by means of a crowdsourcing tool. Afterwards, the consistency of the dataset will be evaluated through inter-annotator agreement metrics, which will also give an account of the inherent subjectivity of the task.
- Examine context-based multilingual vector representations (context-based word embeddings) for metaphorical expressions in neural language models such as mBERT, XLM-RoBERTa, mDEBERTA [30], and those developed specifically for metaphor processing [25], such as MelBERT [31] and MIss RoBERTa WiLDe [32]. The availability of CoMeta together with the English dataset (VUAMC) will allow us to study multilingual approaches to metaphor detection and interpretation, as well as to assess how certain types of metaphors are shared among languages.
- Explore zero- and few-shot approaches combined with cross-lingual word embeddings for metaphorical knowledge transfer. This analysis would leverage the development of automatic metaphor processing systems for a language with scarce resources using the existing corpora for other languages, such as English. The previously mentioned multilingual models learn in one or more languages and make predictions in the target one.
- In connection with these goals, the organization of a shared evaluation task will be proposed, based on CoMeta and VUAMC. Following the lead of the celebrated tasks for metaphor detection [23, 24], it will consitute the first evaluation task for multilingual

metaphor processing. This methodology will promote the assessment and discussion of techniques and ideas in the field of NLP for the detection (first phase) and interpretation (second phase) of metaphor.
- Study the impact of metaphor detection and interpretation on final tasks previously mentioned. This would allow us to examine how the usage of metaphorical expressions alters the communicative intent compared to similar literal utterances and how that is reflected quantitatively in the performance of NLP systems trained for other tasks.

4. Methodology & Experiments

From the point of view of the data, the development of the project requires the compilation and annotation of a corpus in Spanish with annotations of metaphorical expressions present in real texts from diverse domains. We take advantage of the most thorough guideline published for metaphor annotation, MIPVU [12], that resulted in the VUAM Corpus.

With regard to previously developed systems and algorithms, deep learning neural systems have the leading role in the state of the art of NLP [33, 34]. Nevertheless, the vast majority of previous publications are restrained to metaphor detection in English. Our methodological proposal will consist of the following novelties: (i) the development of multilingual and cross-lingual approaches for metaphor detection; (ii) the interpretation of the metaphor framed within the Natural Language Inference (NLI) or Textual Entailment (TE) task, unlike previous proposals based on generating paraphrases [27, 35]; (iii) an evaluation framework will be proposed to assess the impact of metaphorical language on other NLP applications; (iv) synergies between other research groups will be encouraged through the organization of the first shared evaluation task for multilingual metaphor processing in international forums such as SemEval of FigLang.

All compiled data and developed software will be publicly distributed through free licenses to facilitate the reproduction of results and the advancement of scientific knowledge.

4.1. Datasets Development

The first version of CoMeta is the largest dataset with metaphorical annotations at token level in Spanish texts labeled by means of MIPVU methodology. Currently, we are working on its augmentation with more texts from various domains and sources, as well as with its annotation through crowdsourcing tools. The definite version will contain samples of the daily use of metaphor. In addition, issues that arise in the labeling process due to the adaptation of MIPVU to Spanish will also be reported.

For the compilation of this dataset, texts from multiple domains will be taken into account, such as reviews, transcripts of dialogues, news, political discourse, minutes of regional and national parliaments, blogs, wiki, etc. Most of these resources will be extracted from existing datasets developed for other specific tasks. Following in the model of the publication Metaphor Identification in Multiple Languages: MIPVU Around the world [13], it will be established which aspects of MIPVU, originally developed for English, are (i) valid for Spanish, (ii) which must be adapted and, (iii) which ones will be proposed specifically for Spanish. In this first phase, annotators will carry out the labeling task in order to calculate the inter-annotator agreement.

The first augmented and reviewed version of CoMeta will be used as a test bed to generate a first approximation to the automatic detection of the metaphor in experiments detailed in 4.2.

To explore metaphor interpretation within the evaluation frame of NLI, the data from CoMeta will be exploited along with NLI datasets. The annotation process will follow the same procedure through crowd-sourcing tools.

4.2. Metaphor Detection

First experiments for the task of metaphor detection will involve developing a baseline system using the test bench described in the previous section. To do so, we will take advantage of state-of-the-art neural systems based on contextual word embeddings, including pre-trained language models (mBERT, XLM-RoBERTa, mDEBERTA [30, 34] and models pre-trained specifically in the task of metaphor detection, such as MelBERT [31] or MIss RoBERTa WiLDe [32]).

A second iteration of experimentation with deep learning systems will be carried out, as well as in the annotation process, but in the reverse order of the work carried out in the first year. This will allow, based on an error analysis of the results of the first batch, to improve the metaphor detection system and, as a result, refine the corpus characterization of metaphorical language in Spanish.

Some issues involved in the error analysis process imply the need to investigate the correctness of the lexical units, how the metaphor annotation is related to the morphosyntactic information and to the grammatical structure of the text. Additionally, morphosyntactic structures through which metaphors are manifested will be studied in more detail, e.g. subject-verb-object in copulative sentences, or adjective-noun phrases. This characterization will lead to a classification of metaphors based both on the observation of these structures, theoretical metaphor approaches and the semantic features of the terms involved in the metaphorical expression. According to this error analysis, the annotations of the developed datasets will be reviewed.

At this stage, we will perform first multilingual experiments for metaphor detection in Spanish and English. This system will have as its starting point the representations of cross-lingual words learned in neural systems based on multilingual language models.

4.3. Metaphor Interpretation

In this thesis we approach metaphor interpretation from an inferential point of view, since systems trained for this task seem to encounter difficulties when dealing with metaphor [26, 36, 37] . In this case, we will make use of the resulting dataset mentioned in 4.1.

As a novelty, the task will be modeled to learn to infer whether a metaphorical expression can be inferred from a literal expression. Given two fragments, the task consists in deciding whether a hypothesis is an entailment of a premise, which could contain a metaphorical expression. For this, the same multilingual language models that have been applied in detection can be used and fine-tuned.

We will also explore zero-shot approaches combined with cross-lingual embeddings for metaphorical knowledge transfer between languages. This will allow us to empirically demonstrate the results obtained from the theoretical comparison between languages carried out in previous steps.

4.4. Extrinsic Evaluation

We will evaluate the impact of automatic metaphor processing on other NLP tasks (extrinsic evaluation) such as Machine Translation, Sentiment Analysis or the detection of biased articles and/or hoaxes, among others previously mentioned. For instance, in the context of MT, a metaphor not shared by another language could be identified and translated as a literal sentence. On the other hand, for the analysis of reviews it can be useful to recognize the metaphorical expressions used when evaluating a product. Likewise, texts belonging to certain domains, such as politics, are characterized by a higher presence of metaphorical expressions that can help in the identification of this type of discourse. Like the proposed framework for metaphor interpretation, these tasks can be formulated from an inferential point of view, which would make it easier to assess the impact of metaphor on these applications.

5. Conclusions

The main target of this thesis is to contribute to the development of technological tools to process metaphor in Spanish, one of the most widely spoken languages in the world, yet with few available resources and applications for public use. For this reason, this research will revolve around metaphor, both in Spanish and from a multilingual approach that will enable the exploitation of systems developed to process metaphor in other languages; as well as the application of state-of-the-art deep learning techniques in the field of Natural Language Processing. Thus, two areas of knowledge are combined to broaden accessible tools for the automatic treatment of Spanish language.

References

[1] R. Mao, C. Lin, F. Guerin, Word embedding and WordNet based metaphor identification and interpretation, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 1222–1231.

[2] L. Zhang, Metaphor interpretation and context-based affect detection, in: COLING, 2010.

[3] V. Rentoumi, G. Giannakopoulos, V. Karkaletsis, G. A. Vouros, Sentiment analysis of figurative language using a word sense disambiguation approach, in: Proceedings of the International Conference RANLP-2009, Association for Computational Linguistics, Borovets, Bulgaria, 2009, pp. 370–375.

[4] F. Liu, Lexical metaphor as affiliative bond in newspaper editorials: a systemic functional linguistics perspective, Functional Linguistics 5 (2018).

[5] X. Tong, E. Shutova, M. Lewis, Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2021, pp. 4673–4686.

[6] B. van Aken, J. Risch, R. Krestel, A. Löser, Challenges for toxic comment classification: An in-depth error analysis, 2018.

[7] J. Lemmens, I. Markov, W. Daelemans, Improving hate speech type and target detection with hateful metaphor features, in: Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, Association for Computational Linguistics, 2021, pp. 7–16.
[8] V. Prabhakaran, M. Rei, E. Shutova, How metaphors impact political discourse: A large-scale topic-agnostic study using neural metaphor detection (2021).
[9] S. Rai, S. Chakraverty, A Survey on Computational Metaphor Processing, ACM Comput. Surv. 53 (2020).
[10] G. Lakoff, M. Johnson, Metaphors We Live By, 1980.
[11] G. Pragglejaz, MIP: A method for identifying metaphorically used words in discourse, Metaphor and symbol 22 (2007) 1–39.
[12] G. Steen, A. Dorst, J. Herrmann, A. Kaal, T. Krennmayr, T. Pasma, A method for linguistic metaphor identification. From MIP to MIPVU., 2010.
[13] S. Nacey, A. Dorst, T. Krennmayr, W. Reijnierse, Metaphor Identification in Multiple Languages: MIPVU around the world, Converging Evidence in Language and Communication Research, John Benjamins Publishing Company, 2019.
[14] Y. Tsvetkov, L. Boytsov, A. Gershman, E. Nyberg, C. Dyer, Metaphor Detection with Cross-Lingual Model Transfer, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 248–258.
[15] E. Shutova, L. Sun, E. D. Gutiérrez, P. Lichtenstein, S. Narayanan, Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning, Computational Linguistics 43 (2017) 71–123.
[16] J. T. Williams Camus, et al., Get the metaphor right! Cancer treatment metaphors in the English and Spanish press (2016).
[17] F. Martínez Santiago, M. A. García Cumbreras, M. C. Díaz Galiano, A. Montejo Ráez, Etiquetado de metáforas lingüísticas en un conjunto de documentos en español, 2014-09.
[18] J. Charteris-Black, Corpus Approaches to Critical Metaphor Analysis, Springer, 2004.
[19] H. Skorczynska, A. Deignan, Readership and Purpose in the Choice of Economics Metaphors, Metaphor and Symbol 21 (2006) 87–104.
[20] E. Semino, Corpus linguistics and metaphor, The Cambridge Handbook of Cognitive Linguistics (2017) 463–476.
[21] C. Wu, F. Wu, Y. Chen, S. Wu, Z. Yuan, Y. Huang, Neural Metaphor Detecting with CNN-LSTM Model, in: Proceedings of the Workshop on Figurative Language Processing, New Orleans, LA, 2018.
[22] Y. Bizzoni, M. Ghanimifard, Bigrams and BiLSTMs Two neural networks for sequential metaphor detection, in: Proceedings of the Workshop on Figurative Language Processing, 2018, pp. 91–101.
[23] C. W. B. Leong, B. Beigman Klebanov, E. Shutova, A Report on the 2018 VUA Metaphor Detection Shared Task, in: Proceedings of the Workshop on Figurative Language Processing, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 56–66.
[24] C. W. B. Leong, B. Beigman Klebanov, C. Hamill, E. Stemle, R. Ubale, X. Chen, A Report on the 2020 VUA and TOEFL Metaphor Detection Shared Task, in: Proceedings of the Second Workshop on Figurative Language Processing, Association for Computational Linguistics, Online, 2020.

[25] A. Neidlein, P. Wiesenbach, K. Markert, An analysis of language models for metaphor recognition, in: COLING, 2020, pp. 3722–3736.
[26] R. Agerri, Metaphor in Textual Entailment, in: COLING, 2008, pp. 3–6.
[27] E. Shutova, S. Teufel, A. Korhonen, Statistical Metaphor Processing, Computational Linguistics 39 (2013) 301–353.
[28] D. Bollegala, E. Shutova, Metaphor Interpretation Using Paraphrases Extracted from the Web , PloS one 8 (2013) e74304.
[29] P. Xiao, K. Alnajjar, M. Granroth-Wilding, K. Agres, H. Toivonen, et al., Meta4meaning: Automatic metaphor interpretation using corpus-derived word associations, in: Proceedings of the Seventh International Conference on Computational Creativity, Sony CSL Paris, 2016.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, 2017.
[31] M. Choi, S. Lee, E. Choi, H. Park, J. Lee, D. Lee, J. Lee, MelBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories, 2021.
[32] M. Babieno, M. Takeshita, D. Radisavljevic, R. Rzepka, K. Araki, MIss RoBERTa WiLDe: Metaphor Identification Using Masked Language Model with Wiktionary Lexical Definitions, Applied Sciences 12 (2022).
[33] A. Akbik, D. Blythe, R. Vollgraf, Contextual string embeddings for sequence labeling, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 1638–1649.
[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 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), Association for Computational Linguistics, Minneapolis, Minnesota, 2019.
[35] S. Mohammad, E. Shutova, P. Turney, Metaphor as a medium for emotion: An empirical study, in: Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics, Association for Computational Linguistics, Berlin, Germany, 2016, pp. 23–33.
[36] M. Mohler, M. Tomlinson, D. Bracewell, Applying Textual Entailment to the Interpretation of Metaphor , in: 2013 IEEE Seventh International Conference on Semantic Computing, IEEE, 2013, pp. 118–125.
[37] T. Chakrabarty, D. Ghosh, A. Poliak, S. Muresan, Figurative language in recognizing textual entailment, arXiv, 2021, pp. 3354–3361.