

Annotation Guidelines: Linguistic metaphors in a corpus of popular science in Mexican Spanish

Development and revision: Alec Misael Sánchez Montero, Gemma

Bel-Enguix, Sergio Luis Ojeda Trueba

English translation: Diana Luz Hernández Solís

General overview

In this guide, some guidelines for carrying out the annotation of metaphorical expressions in the corpus of tweets provided will be presented. The objective is to identify the linguistic units that are used as metaphors and annotate them within each tweet, using the available tags.

Corpus

The corpus is made up of 3734 tweets from authors dedicated to popular science in Mexico. These texts were published between January 2020 and May 2023. The topics of the tweets cover areas such as astronomy, biology, physics, computer science, health sciences, technology, neurosciences, history of science, and mathematics, among others.

Definition of metaphor

In the context of this annotation, an expression is considered metaphorical if it establishes a conceptual relationship between a source domain and a target domain, usually to describe or explain abstract or difficult-to-understand concepts in a way that is more accessible or familiar. The source domain usually belongs to a more concrete and familiar concept, which is taken from our everyday experience and used as a starting point to describe or understand the target domain. Meanwhile, the target domain refers to the image or concept —often more abstract— that we try to explain or understand.

Types of metaphors

Three types of metaphor will be considered:

1. **Standard metaphor:** These metaphors establish a direct comparison between the **source** domain and the **source** domain. The relationship between both domains are presented explicitly using **signs**, such as words or phrases indicating a comparison. It consists of **three parts**: **Source unit**, **target unit**, and **sign**.

Do you usually check your phone before going to sleep?

BZZZZZZZT!

This is why **this device** **is** our **worst ally** when trying to fall asleep.

#RediscoverScience on #WorldSleepDay! <https://t.co/wDvzhO7HVE>

2. Do dogs see only in black and white?

During the day, their vision is somewhat blurry, but at night it's better than ours! They also have a **layer** of light-reflecting, **mirror-like tissue** behind the retina, called tapetum lucidum, which improves their night vision considerably.

Figure 1 Example of tweets containing standard metaphors.

2. **Implied metaphor:** In this type of metaphor, there is an **implicit comparison** between the source domain and the target domain, so there are no words or signs indicating the presence of metaphor, unlike a direct metaphor. Instead, only the **source unit** is presented. That is, the one that belongs to the domain we use to make the comparison with the target concept. In implied metaphors, the target domain is not explicitly mentioned in words, which requires the reader to make an inference to grasp the implicit relationship between the two domains. In these cases, the meaning of the source unit in the text should be contrasted with its more conventional meaning; this means that the meaning of the unit may be lexicalized, but its use in a different or unexpected sphere is what gives it its status as a metaphor, as can be seen in the following examples.

New numerical simulations of the distribution of matter in the **cosmic web** and the background radiation of the early universe shed light on the period of inflation of the cosmos, when it was less than a second old. (Universe Today)
<https://t.co/avKiT7yVOp>

Today is one of our favorite days:

International Day of Women and Girls in Science

To celebrate, learn all about our favorite female scientists in our **thread**.

Retweet so it information reaches more young science lovers!

Figure 2 Example of tweets containing implied metaphors.

3. **Anthropomorphism:** Anthropomorphism means comparing or describing something non-human in terms of human characteristics in order to create a more concrete image. This type of metaphor holds a special relation between the **anthropomorphized object** (usually a noun, a proper noun, or a noun phrase) and **a word or phrase that anthropomorphizes** (usually a conjugated verb or a verb in an impersonal form, although other types of words whose conceptual meaning implies human or animate features may also be included).

Some celestial bodies produce X-rays, high-energy radiation that does not pass through the atmosphere. The new **IXPE space telescope** **will analyze** this radiation generated by stellar explosions, matter jets generated by disks and jets around black holes, and much more. (NASA) <https://t.co/84LjMjoXjU>

<p>In 1986, a disease that attacked the nervous system of cows emerged in the United Kingdom. In 1996, the world was shocked by the “mad cow disease”.</p> <p>25 years ago, it appeared in the Revista Ciencias: https://t.co/2jk2EP2Ma2 https://t.co/TF2MlAtUmW</p>

Figure 3 Example of tweets containing anthropomorphism.

Available labels

No.	Label	Represents	Description
[1]	md_fuente	Standard metaphor: Source Unit	Use this label to identify and label the source unit in a standard metaphor. The source unit is the word or group of words that belong to the source domain and are used to do the comparison.
[2]	md_meta	Standard metaphor: Target	Use this label to identify and label the target unit in a standard metaphor. The target unit is the word or group of words that belong to the source domain and represent the concept with which the source domain is compared.
[3]	md_señal	Standard metaphor: Sign	Use this label to identify and label the sign or indicator in a standard metaphor. The sign is the word or phrase that signals explicitly the comparison between both domains. It can be a word or phrase including terms like: “like [a/an]”, “as if”, “the same way as”, “[something being] reminiscent of [something else]”, “in the manner of”, “metaphorically”, “figuratively speaking”, “just like”, “as [...] as”, etcetera, or punctuation marks with a comparative function (such as a comma, quotation marks or a colon).
[4]	m_indirecta	Implied metaphor:	Use this label to identify and label the target unit in an implied metaphor. In this case, there will not be an explicit sign indicating the metaphor, so you will have to identify and label the source unit only, as it was previously described.


[5]	pers_obj	Anthropomorphism: Anthropomorphized object	Use this label to identify and label the anthropomorphized object in case of anthropomorphism. The anthropomorphized object is the inanimate entity to which we attribute human or animate characteristics.
[6]	personificador	Anthropomorphism: Anthropomorphizing element	Use this label to identify the word or phrase that acts as the anthropomorphizing element. It could be a conjugated verb, an impersonal form of the verb or some other type of word that attributes human or animate characteristics to the anthropomorphized object.

Table 1: Available labels for the task

Annotating process

For the annotating process, you will need to identify every unit (word or group of words) involved in the detected metaphor within a tweet and label it according to the provided classification and the corresponding label, as shown in the table in the previous section. It is crucial to consider the context to decide whether an expression is metaphorical or not, which is why you should read the whole tweet to comprehend the meaning of the message and the particular contextual meaning of the words.

Keep in mind that it is possible to find more than one type of metaphor within a single tweet. Therefore, the system allows you to use the six types of label as many times as needed in a single text entry. More than one label can be applied to the same section of words, as long as the length of the second label is greater than the first. In some cases, certain words will be involved in more than one metaphor, which is why this function is useful to indicate their different metaphorical meanings in the text. You must be weary, though, of the order in which you place all the labels.

When you are done annotating a tweet, you will need to validate the labels you have applied to each text. For doing so, you must click the “ **Validate**” button corresponding the registration you annotated. It is important to save the changes this way, otherwise, your modifications will be lost when changing pages. After validating the annotation, it is still possible to delete or modify the labels, so you do not have to worry if you make a mistake.

IMPORTANT NOTE: Due to previous problems with the platform, it is recommended to wait a few seconds for the changes to be registered in case it does not allow you to validate the registrations. If the platform continues to be unresponsive, it is best to continue annotating other records and return to the problematic text later.

The “Validate” button should not be pressed more than once, since it could delete the changes in the registration.

Regarding the other buttons:

- Do not press the “Discard” button.
- Only press the "Clear" button if you plan to remove all tags from a tweet, in case it is necessary to correct a very large error in the annotation.
- Do not press the “Reset” button.

What should I do with the tweets that don't contain metaphors?

For tweets that do not contain metaphors, there are no additional labels. In other words, if you come across a metaphor-free tweet, validating the register is enough. No further action is needed.

What are the limits for my annotations?

Most of the annotations will correspond to nouns or noun phrases, so it is necessary to indicate the limits of the annotation in order to apply the adequate label. For proper names, all the words that make up the name and refer to specific entities should be marked, whether they are in Spanish or in another language, for example "James Webb Space Telescope" instead of "telescope". In the case of syntagms, it is considered relevant to annotate in the same label, in addition to the noun in the head, **the semantic information directly related to it**. In syntactic terms, almost all indirect modifications —that is, those introduced by a preposition, relative pronoun or certain conjunctions— are discarded, while direct modifications, usually of adjectival or even nominal category, must be included in the annotation. NOTE: Only in exceptional cases may indirect modifications introduced by prepositions such as "of" or "to" be considered, although these cases should be reported with their corresponding arguments.

As a general rule, determiners that may accompany the noun or modification —articles, numerals, possessives, demonstratives— should be left out of the annotation. It is only necessary to annotate determiners if they belong to a standard metaphor sign, for example, in expressions similar to "like a/an", just as detailed in a previous section.

Likewise, in the case of verbs, the set of words that make up the *head of the verbal phrase* should be recorded under the same label, whether they are simple verbs (e.g. *he ate, we listened*), compound verbs (e.g. *he had eaten, he was listening, he is talking*), verbal periphrases (e.g. *we need to know, we have to talk*) or phrasal verbs (e.g. *they found out, he fell in love, it blew up*). VERY IMPORTANT NOTE: In the case of verbs involved in anthropomorphism, close attention must be paid to semantics to corroborate that it is indeed a verb requiring human or animate features that attributes such features to an inanimate or abstract object. When in doubt, it is recommended to consult a dictionary. SECOND VERY IMPORTANT NOTE: In particular, for the verb "to be" in standard metaphors, one has to read in detail to make sure that it is not being used to attribute or to define. Its function has to be strictly comparative in order to score it as a sign.

Regarding punctuation, as a general rule, only those cases are considered relevant where a punctuation mark plays the role of a signal in the direct metaphor, as already mentioned.

One strategy that can be followed to delimit the annotations is to follow a substitution criterion. If a given metaphorical unit can be replaced by another without altering the meaning of the text, it can be considered to contain the semantic information essential for the annotation.

Should I include emojis in the annotations?

No.

Clarification about terminology

In the context of this annotation, we consider that a scientific term or specialized terminological unit may be metaphorical in nature, even if it is lexicalized or used with a very particular meaning in colloquial or non-specialized contexts. For example, a planet's "rings", family "trees", DNA "chains" or neural "networks" would be considered metaphors, even though they belong to scientific terminology. In other words, a word or phrase that appears to have a literal, specialized or fixed meaning in the general context of the tweet may be considered metaphorical if it establishes a comparison between conceptual domains.

Comprehensiveness and consistency

All the metaphors identified in the tweet should be marked, even if they are repeating metaphors or if a metaphor is repeated in the same tweet. Likewise, all types of metaphor identified in the tweets should be annotated, even when a word has several functions according to the labels described above. Always keep in mind that for each standard metaphor there should be three annotations, two for each anthropomorphism, and only one for each implied metaphor.

Dictionary use

When in doubt about the meaning of a lexical unit, you can consult a dictionary to clarify its meaning. This step is not always needed.

Problem cases and guideline review

If you come across a tweet where there is a conflict over the status of the metaphor or the relationship between the label and the word is judged to be complicated, you will be required to keep the ID of the tweet and report it. In the event of detecting special cases that might be of interest for this task, which are not specified by this annotation guide, they should be reported at the time, together with the respective examples, in order to refine the annotation guidelines.

One last comment

Do not discard your intuition as a native Spanish speaker. Sometimes, it is enough to think about the contrast of conceptual domains to verify our annotations.

Table of contents

General overview.....	1
Corpus	1
Definition of metaphor.....	1
Types of metaphors	1
Available labels	3
Annotating process	4
What should I do with the tweets that don't contain metaphors?	5
What are the limits for my annotations?	5
Should I include emojis in the annotations?	6
Clarification about terminology	6
Comprehensiveness and consistency	6
Problem cases and guideline review.....	6
One last comment	6