Medical-FLAVORS-AECC: Spanish Oncological Metaphors Dataset

Lucía Pitarch[1], Jordi Bernad[1], Sergio-Luis Ojeda-Trueba[2]
Alec Sánchez-Montero[3], Maxim Ionov[1], Emma Anglés-Herrero[1]
Ángel Óscar Corona Beomont[4], Gemma Bel-Enguix[2]
Universidad de Zaragoza, Zaragoza (Spain)[1]
Universidad Nacional Autónoma de México, Ciudad de México (Mexico)[2]
Universidad Pompeu Fabra, Barcelona (Spain)[3]
Centro Universitario de la Defensa de Zaragoza, Zaragoza (Spain)[4]
{lpitarch,jbernad,mionov,eangles,acorona}@unizar.es
sergioluis0906@gmail.com, alecm.montero@gmail.com, gbele@iingen.unam.mx

Abstract

Metaphors play a central role in cancer narratives, helping patients and practitioners articulate complex experiences and technical concepts. While cancer metaphors in English have been extensively studied, Spanish remains underexplored in this regard, despite its global importance and rich cultural variation. This paper presents a new dataset of Spanish cancer metaphors designed to address these gaps. The resource comprises over 80K annotated words drawn from diverse forum posts, with detailed documentation of lexical units, contextual versus basic meanings, and inter-annotator agreements. To construct the dataset, we adapted the Metaphor Identification Procedure (MIP) for Spanish medical discourse, proposing methodological refinements to challenges such as defining lexical units or domain-specific Basic Meaning labels.

Keywords: Metaphor annotation, Spanish Language Resources, Oncological Metaphors, Cancer Narratives

1. Introduction

Metaphors are cognitive and linguistic devices that enable speakers to explain complex or abstract concepts through more concrete or familiar terms (Lakoff and Johnson, 1980). In highly technical and emotionally sensitive contexts, such as cancer communication, metaphors play a central role in shaping how patients, families, and professionals conceptualize the illness (Pinheiro et al., 2017; Landau et al., 2019; Gustafsson et al., 2019; Fadul et al., 2009). Building on the premise that metaphors are not inherently positive or negative, Semino and Demjen (2017) developed the Metaphor Menu to help patients and practitioners select more suitable language based on individual experiences and preferences, as understanding their use, potential impact, and the patient preferences is key to effective communication.

This paper presents a manually annotated subset from the AECC (Asociación Española Contra el Cancer) forums, a central place for communication among Spanish-speaking cancer patients and their family members. This dataset will be used to train and evaluate a model for the automatic extraction of metaphors from the rest of the corpus. The annotated subset comprises 72 posts (700 words each on average, including comments).

Following a Spanish adaptation of the MIPVU (Metaphor Identification Procedure Vrije Universiteit) (Steen et al., 2010) adjusted for the medical domain, the subset was annotated by five independent annotators. From over 83,000 words, approximately 2,035 phrases (ca. 2.45%) were labeled as metaphoric vehicles related to the cancer experience by at least one of the annotators. With an overall inter-annotator agreement (IAA) of 0.49 (F1 score), the resulting dataset is available upon request and can be used to fine-tune language models for metaphor identification (see Ge et al. (2023) for possible metaphor identification methods).

The main contributions of this paper are as follows:

- We present Medical-FLAVORS-AECC dataset, a large-scale Spanish-language resource on metaphors used by patients and their families to describe the cancer experience.
- We describe the annotation process and inter-annotator agreement, providing the first resource for the Spanish medical domain that explicitly documents the distinction between a lexical unit's basic and contextual meaning as established by the MIPVU guidelines.
- We provide a curated vocabulary of over 800 distinct metaphorical vehicles (grouped into 81 source and 46 target domains), useful for both qualitative linguistic analysis and NLP applications.

All annotation guidelines, metadata, and the curated vocabulary are released in a GitHub repos-

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 159–170

itory[1] to ensure reproducibility and support future corpus-based and computational studies on metaphor in the medical domain.

This paper begins by reviewing relevant precedents in Section 2. We then describe the data extraction from the AECC forums and present general statistics of the corpus in Section 3, followed by a detailed account of the annotation and curation process in Sections 4 and 5. Finally, in Section 6 we present an analysis of the resulting dataset, highlighting the insights it provides into the cancer experience as communicated by real people.

2. Related Work

Metaphors are not only pervasive in medical language (Casarett et al., 2010; Semino et al., 2017) but also highly relevant. In medical settings, metaphors translate specialized terminology into lay terms (Pinheiro et al., 2017), facilitate decision-making (Landau et al., 2019), support coping (Gustafsson et al., 2019), and address sensitive topics (Fadul et al., 2009).

Research on metaphors has largely focused on English, with limited attention to Spanish, despite its global reach. Authors such as Liu et al. (2024); Pitarch et al. (2024); Magaña (2020) highlight the need to adapt metaphor studies to different languages and varieties. While some resources already exist for Spanish, they are scarce and limited. For instance, the MetaNet project (Dodge et al., 2015) includes only a small and heterogeneous set of Spanish examples, while the first version of Medical-FLAVORS (Pitarch et al., 2024), a dataset of Spanish cancer metaphors, is biased toward Chilean Reddit data, with uneven cancer types and slang-heavy language. Both resources lack transparent annotation procedures, a crucial issue given the subjectivity of metaphor analysis (Sánchez-Montero et al., 2025a,b).

Our work addresses this gap by introducing Medical-FLAVORS-AECC, a new module for the Medical-FLAVORS dataset. For its annotation, we adapt the Metaphor Identification Procedure (MIP) (Pragglejaz Group, 2007) to Spanish medical discourse. This method, along with its extension MIPVU (Steen et al., 2010), is the most established framework for metaphor annotation and involves four steps: (1) segmenting the text into lexical units, (2) determining each unit's basic meaning (the more concrete, physical, or older sense), (3) identifying its contextual meaning, and (4) labeling the unit as metaphorical if a contrast exists between the two meanings.

While extensively used, the MIP poses significant challenges, particularly in defining text segmentation into relevant lexical units and selecting the most basic meaning of a word. These core decisions are often controversial (De Backer et al. (2023); Maudslay and Teufel (2022)), lack objective metrics, and are rarely documented in new corpora annotations. Furthermore, both MIP and MIPVU were initially developed for English, lacking fine-grained adaptations to Spanish. To our knowledge, the dataset presented in this paper is the first of its kind to address these issues for Spanish by providing explicit guidelines, annotations, and metrics in terms of lexical unit segmentation and meaning distinction.

3. Data Extraction

The data were extracted from the web forums of the Asociación Española Contra el Cáncer (AECC), founded in 1953 to provide assistance and support to cancer patients and their families. Although based in Spain, the forums are accessible worldwide, enabling Spanish-speaking users from diverse regions to share their experiences. It is organized into multiple topics, ranging from specific cancer types to prevention, volunteering, and events. While anyone may read the forums, registration is required to participate. The most active and general thread, "Testimonios de cáncer" (cancer stories), has received contributions continuously since 2010. Given the association's prominence and its broad coverage of patients, relatives, and experiences over time, the AECC forums were chosen as a representative and ethically appropriate source of cancer discourse in Peninsular Spanish.

We selected posts from three topics: "Testimonios" (general thread), "Cáncer de mama" (breast cancer), and "Cáncer de próstata" (prostate cancer). Their high incidence, together with their cultural and emotional salience (World Health Organization, 2025; Sociedad Española de Oncología Médica, 2025a,b) in patient discourse, make these two types particularly relevant for metaphor analysis in Spanish. The "Testimonios" thread, on the other hand, provides a more general overview of first-hand cancer experiences. Each post in the forums consists of a title, an initial message from the main author, and a thread of subsequent comments from other community members.

4. Annotation

The annotation process was carried out in two phases: a pilot testing phase for guideline refinement, followed by the main annotation phase. In the pilot testing, all annotators applied a preliminary draft of the guidelines to the first five texts. Questions and disagreements were discussed in a subsequent meeting and integrated into the revised guidelines, with special attention to the delimitation

[1] https://github.com/LuciaPitarch/MedicalFLAVORS-AECC/

of lexical units. To ensure consensus, a sixth text was co-annotated and reviewed by the entire team. For the main annotation, the remaining posts were distributed among annotators under the following conditions: (a) each post was annotated by two annotators, (b) the workload was balanced, and (c) all possible annotator pairs were assigned common posts. The annotation took place between May and October 2025, following the principles of the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). A summary of the annotators' profiles can be found in Table 3.

4.1. Annotation Platform and Guidelines

This section provides a summary of our annotation guidelines. The entire annotation was conducted on the INCEpTION platform, as illustrated in Figure 1. where we can observe how each annotated post follows a typical forum structure: an initial message from the main author followed by a series of comments from other community members. Figure 2 summarizes the annotation pipeline we applied.

4.1.1. Definition of Metaphor

Following the principles of MIP (Pragglejaz Group, 2007) and its extension MIPVU (Steen et al., 2010), we define a metaphor as a linguistic expression in which a more concrete, physical, or perceptually imaginable concept is used to refer to a more abstract domain of experience. For instance, describing CANCER AS A BATTLE constitutes a metaphor, as it draws on the tangible and culturally salient domain of warfare to frame the intangible and multifaceted experience of illness.

4.1.2. Lexical Unit Segmentation

Given the complexity of metaphorical expressions in our dataset, a flexible approach to lexical unit segmentation was required. Metaphorical meaning often arises not from isolated words, but from the interaction of multiple elements within a phrase or clause, or even from a text as a unit. It can sometimes only be comprehended by comparison with very similar clauses where different words interact.

For example, in constructions such as *voy paso a paso con el tratamiento* ("I go step by step with the treatment") or *lucho a diario contra el cáncer* ("I fight daily against cancer"), the metaphor is distributed across several words. Similarly, in expressions like *recibir tratamiento* ("to receive treatment") or pasar por tratamiento ("to go through treatment"), the metaphorical meaning, where treatment is conceptualized as an object or a milestone in a journey, emerges from the semantic interaction among the elements involved. In such cases, understanding how the person envisions the experience of being ill with cancer or how they conceptualize treatment only becomes possible by analyzing the full clause.

To account for this, we adopt Lynne Cameron's approach to metaphor annotation (Cameron, 2007), particularly the notion of *metaphorical vehicles*: clusters of words that collectively convey a metaphorical mapping. We annotate all lexical elements central to the metaphor, including verbs, nouns, adjectives, and multiword expressions.

Importantly, only metaphorical expressions that pertain to the experience of the illness (e.g., physical, emotional, or cognitive aspects of cancer) are annotated, which means our focus is on cancer metaphors. When the metaphor is confined to a single word, it is annotated in the `Metaphors` layer, as depicted in Figure 2.

4.1.3. Annotating Basic and Contextual Meaning

Although MIP and MIPVU highlight the centrality of the *basic meaning* (BM) and *contextual meaning* (CM) in metaphor identification, few concrete examples are provided in their documentation. Besides, BM and CM are not explicitly annotated in most publicly available datasets. To enhance transparency and reproducibility in our annotations, we chose to explicitly annotate the BM and CM for each metaphorical unit. To do so, we embedded the Spanish version of the Open Multilingual Wordnet (Vossen et al., 2016) into the INCEpTION platform, creating an integrated tool for annotation specifically adapted to our task.

The annotators were instructed to use the BM annotation fields as follows:

- Enter the basic meaning in the `Basic-Meaning` field.
- When available, retrieve definitions from WordNet that correspond to the most concrete or prototypical sense of the lexical unit.
- If no appropriate WordNet entry exists, consult the *Diccionario de la lengua española* (RAE) (Real Academia Española, 2023) or other authoritative sources such as ASALE (Asociación de Academias de la Lengua Española, 2023).
- If neither resource provides a suitable definition, provide a manually formulated paraphrase based on your interpretation of the basic meaning.
- Use the `Contextual-Meaning` field to define the meaning of the lexical unit in the given utterance. If a figurative sense is listed in the

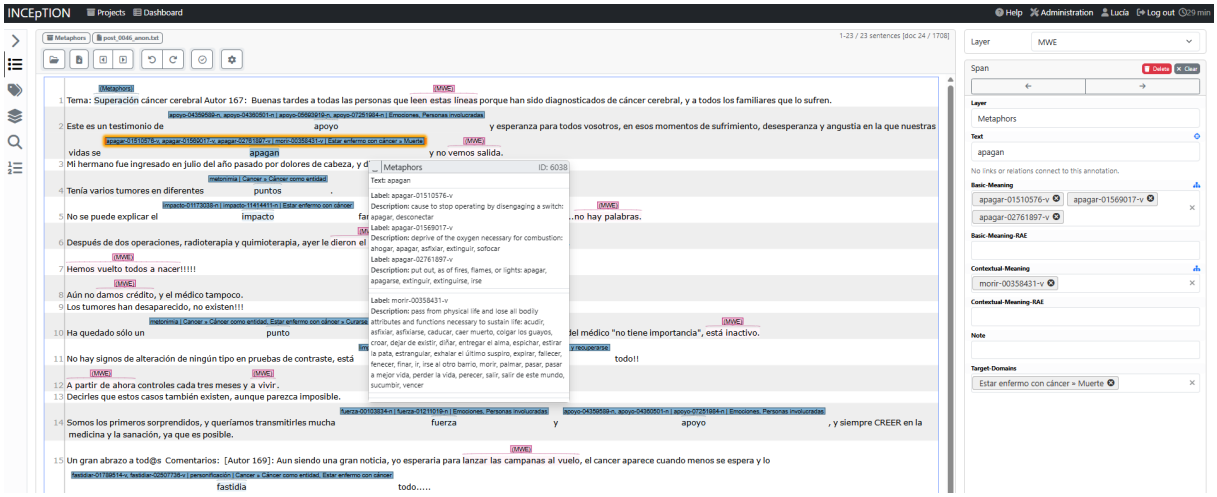

Figure 1: Annotation in INCEpTION, with the additional lexical resources at right

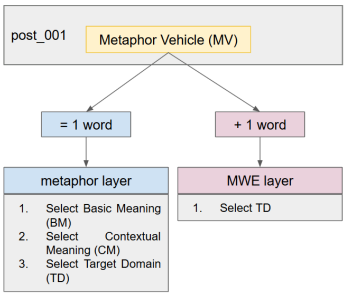

Figure 2: Schematic annotation pipeline

lexical resources, select it accordingly. When no direct match exists, but related entries reflect the intended sense, they may be adapted and cited.

These annotations can facilitate further semantic analysis by leveraging hypernymy/hyponymy relations and identifying recurring conceptual domains.

4.1.4. Identifying Cancer-Related Themes

We tagged each identified metaphor with its corresponding target domain(s) to classify the aspect of the cancer experience it addresses. Common domains include bodily experience, treatment, emotional impact, and social relationships.

Annotators selected the most appropriate theme(s) from a predefined list in the `Target-Domain` field. Multiple themes could be tagged if a metaphor spanned more than one dimension of experience. The topics were extracted from a relevant scoping review on cancer metaphors (Liu et al., 2024) and further adapted to gain a finer understanding of the experience through discussions with healthcare professionals and conversations with patients. The full list of topics is available on GitHub.

4.1.5. Notes and Issue Tracking

Annotators are encouraged to use the `Note` field (in both the `Metaphors` and `MWE` layers) to document uncertainties, interpretation challenges, or contextual ambiguities encountered during the annotation process. In the event of repeated content, textual errors, or systemic inconsistencies, annotators had to report them in the shared `Incidencias` log for follow-up and dataset revision.

4.2. Agreement Metrics

To report our inter-annotator agreement (IAA) we provide F1 score, following Boguslav and Cohen (2017), as it has been proven to be a more transparent and efficient metric for tasks that involve phrase extraction. Most widely used metrics to report IAA, such as Cohen's Kappa or Krippendorf's Alpha, will not work well on tasks where many spans are possible and the large number of possible negative spans can distort agreement scores.

Using F1 score implies that, for each post, one annotator's labels were treated as the gold standard and the other's as the prediction. An agreement, or a true positive, was defined by a partial overlap, meaning at least one word was shared between the spans marked by the two annotators in any layer (metaphor or MWE). For instance, if Annotator A marked "fight against cancer" and Annotator B marked only "fight", the overlap was considered as an instance of agreement. The overall F1 score was 0.49. A detailed breakdown of the results is presented in Tables 1 and 2.

Annotator	Mean F1
A1	0.60
A2	0.52
A3	0.48
A4	0.53
A5	0.49

Table 1: Mean IAA scores by annotator

L_1	L_2	F1	L_1	L_2	F1
A2	A1	0.64	A3	A1	0.62
A1	A4	0.60	A5	A1	0.55
A2	A3	0.54	A2	A4	0.47
A3	A5	0.47	A2	A5	0.46
A5	A4	0.46	A3	A4	0.45

Table 2: IAA scores by annotator pair (L_1,L_2)

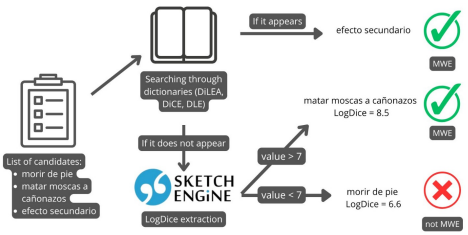

Figure 3: Pipeline MWE curation

5. Curation

The curation process of the annotated dataset focused on two main challenges: (1) clustering metaphorical vehicles (including whether to analyze them compositionally or holistically), and (2) refining the final annotations to determine which specific words would be labeled as metaphorical.

5.1. Curation of Metaphorical Vehicles

Metaphorical vehicles pose particular challenges for annotation due to their morphological variability, syntactic discontinuity, and position on a continuum from compositional to fixed expressions. In this work, we distinguish two main categories of multi-word metaphorical vehicles: (a) **Multi-Word Expressions (MWEs)**, understood as conventionalized lexical combinations with fixed or semi-fixed structures (Corpas Pastor and Alvar Ezquerra, 1996; Escartín et al., 2018), and (b) **complex grammatical constructions** that, although not MWEs in the strict sense, contribute semantically or pragmatically to metaphorical meaning.

5.1.1. Multi-Word Expressions (MWEs)

MWEs were identified through a manual, iterative clustering process based on structural patterns described by Corpas Pastor and Alvar Ezquerra (1996), Escartín et al. (2018), and the *Nueva Gramática de la Lengua Española* (NGLE, 2009). The following types were considered relevant for metaphor annotation:

Light Verb Constructions (LVCs). LVCs[2] were considered for metaphor analysis when the choice of light verb introduced a paradigmatic contrast that shaped the metaphorical framing of the event. Common light verbs include *dar* (to give), *tener* (to have), *tomar* (to take), *recibir* (to receive), and *pasar por* (to go through, to undergo). For instance, in expressions such as *dar miedo* (to scare, lit. to give fear) or *recibir noticias* (to receive news), abstract entities like emotions and information are conceptualized as manipulable objects that can be given or received. Similarly, in *tocar*, *recibir*, or *pasar por tratamiento* (to have, to receive, or to go through treatment), the noun *tratamiento* ('treatment') is metaphorically construed as a tangible entity, an impinging external force, or a milestone within a journey. In these cases, LVCs were annotated as part of the metaphorical vehicle because the figurative meaning resides in the interaction between the light verb and its abstract complement, rather than in either element alone.

Fixed and Formulaic Expressions. Highly frequent and context-bound collocations such as greetings, discourse routines, or conventional phrases (e.g., *¿cómo te va?* 'how is it going?' [journey] vs. *¿qué tal estás?* 'how are you?' [neutral]) were annotated when they contributed metaphorical shading to discourse (Dickins, 2020).

Idiomatic Expressions and Collocations. Idioms (e.g., *comerse el coco* 'to overthink', 'to eat one's head', from lit. 'to eat the coconut') were treated as single analytical units due to their semantic opacity. Potential collocations were evaluated for **institutionalization** and **frequency**, and analyzed as single lexical units when they met these criteria. Institutionalization was verified using **DiLEA** (Penadés Martínez, 2020), **DiCE** (Vincze et al., 2011), and **RAE** resources. Frequency was assessed using the **logDice** metric (Rychlý, 2008), robust to corpus size (Gablasova et al., 2017). Following Pinto et al. (2021), co-occurrences with logDice ≥ 7 in the *esTenTen23* corpus (Sketch Engine) (Sketch Engine, 2025) were considered valid phraseological units. See Figure 3 for more details.

Grammatical Constructions Included in the Vehicle. Certain grammatical combinations, while not metaphorical in themselves, modulate the interpretation of metaphors through modality, temporality, or aspectuality. These were included within the metaphorical vehicle but not further analyzed for metaphorical content:

- **Modal verb constructions:** Expressions

[2]We take the definition and LVC selection from Alonso Ramos (2004)

with *haber que/de*, *tener que*, *poder*, *deber*, or *querer* (e.g., *debo tomar precauciones* 'I must take precautions') adjust the degree of the speaker's modality, obligation or possibility (Magaña, 2020).

- **Temporal verb constructions:** Forms such as *ir a*, *ir + gerundio*, or *haber + participio* frame metaphorical processes temporally (e.g., *va a pasar por algo parecido* 'he's going to go through something similar').
- **Copulative verbs:** *Ser*, *estar*, and *parecer* frame metaphors by marking states or attributes (*soy una persona fuerte* 'I am a strong person', *estoy débil* 'I am weak'). For instance, *la alimentación y el deporte son aliados* ('nutrition and exercise are allies') encodes a direct metaphorical comparison.

5.1.2. Non-MWE Multiword Metaphorical Vehicles

In addition to MWEs, certain multiword combinations were annotated as single metaphorical vehicles when they contributed meaningfully to metaphor construction. Four main cases were identified:

Frame-interaction metaphors. Following Stickles et al. (2016), we included cases where metaphoricity arises from **mismatches between semantic roles** or **cross-frame alignments** rather than individual lexical items, e.g., *lo único que acapara en mi cabeza es la muerte* ('the only thing that takes over my head is death') or *remamos en el mismo barco* ('we row in the same boat').

Explicit metaphors. Following the MIPVU framework and extensions to Spanish proposed by Sánchez-Montero et al. (2025a), explicit metaphors and similes containing overt markers or signals (*como si*, *como que*, 'as if', 'as though') were annotated as multiword vehicles (e.g., *vivo como si mañana no fuera a estar* 'I live as if tomorrow I would not be here').

Extended frame metaphors. In some cases, metaphorical content extended coherently across an entire clause or frame, encompassing several lexical items that together instantiate the metaphorical domain (e.g., *es una carrera de obstáculos y hay que ir pasándolos* 'it is an obstacle race and one must keep overcoming them').

Hyperbolic expressions. Recurrent hyperboles (*no hay nada que puedas hacer* 'there's nothing you can do', *no pasa nada* 'it's fine', *has hecho todo lo que ha estado en tu mano* 'you've done everything within your reach') were annotated as multi-word vehicles when functioning as intensified metaphors rather than literal exaggerations.

5.2. Vocabulary Curation

Given the restricted domain of oncology, many metaphorical expressions were highly recurrent, often used with consistent basic and contextual meanings across posts. To analyze these patterns systematically, we compiled a unified vocabulary. The process began by merging all annotations, which contained an initial list of 2035 candidate items. Subsequent clustering and curation refined this list to a final vocabulary of approximately 800 distinct metaphorical vehicles.

Leveraging the annotated synsets, word senses were clustered according to hypernymy relations to identify shared semantic domains. Once these domains were established, lexical items were reviewed and discussed within their respective clusters. This process facilitated the evaluation of similar semantic behaviors and ensured coherent annotation decisions.

One illustrative case involved weak verbs within the semantic field of movement or transition (e.g., *seguir* ('to follow/continue'), *comenzar* ('to begin'), *terminar* ('to end'), *pasar* ('to go through'), *pasar por* ('to undergo')). These verbs are highly grammaticalized, resulting in considerable polysemy and contextual variability, which made it difficult to establish a stable basic sense, leading to the following decisions:

- *Seguir* ('to follow') was marked as metaphorical when referring to pursuing or following someone, but not when it meant 'to continue'.
- *Comenzar* ('to begin') and *terminar* ('to end') were marked as non-metaphorical.
- *Pasar* and *pasar por* ('to go through', 'to undergo') were generally marked as metaphorical due to their abstract uses to describe experiences or phases.

Another discussion centered on adjectives describing physical characteristics (e.g., *débil* ('weak'), *fuerte* ('strong'), *duro* ('hard'), *grave* ('serious'), *agudo* ('acute')). The team questioned whether the physical-quality sense (e.g., *fuerte* as the ability to exert physical resistance) was more basic than its use to describe a person's body (e.g., physical strength, muscular build). Since neither could be clearly established as more fundamental without resorting to etymological analysis, which was beyond the scope of this study, these adjectives were only annotated as metaphorical when used to describe mental or emotional processes

(e.g., *una persona fuerte* 'a strong person' in the sense of emotional resilience).

Similarly, *controlar* ('to control') was marked as metaphorical in expressions such as *controlar el dolor*, *el miedo*, or *la enfermedad* ('to control pain', 'fear', or 'the illness'). Although *controlar* has a broad meaning, examining its noun form *control* ('control') revealed a more basic sense related to exerting physical power, which supported its metaphorical interpretation in these contexts.

A further case involved *fe* ('faith'). A distinction was made between *fe* as religious belief (not annotated as metaphorical) and *fe* as confidence or trust in something (annotated as metaphorical), as in *tener fe en el tratamiento* ('to have faith in the treatment').

Curation decisions were finalized through a consensus-based review of concordance lines in group sessions, which was essential for resolving borderline cases. To provide transparency, the complete vocabulary file of over 800 metaphorical vehicles, including basic and contextual meanings, is openly available. To the best of our knowledge, this constitutes the largest vocabulary of annotated basic and contextual meanings to date, and the only domain-specific one. Beyond transparency, the explicit annotation of basic meanings also enhances interoperability: since senses are aligned with WordNet (Miller, 1995), the resource can be linked to other semantic frameworks such as Framester (Gangemi, Aldo and Presutti, Valentina and Navigli, Roberto and others, 2017).

Moreover, basic meaning (BM) annotation is a crucial component in several automated metaphor identification methods. However, the lack of manually annotated resources has often led such systems to rely on shortcuts, such as selecting the first dictionary sense, an approach that does not always correspond to the most appropriate basic meaning in language use. By providing a large, systematically annotated BM inventory, this resource can help improve the accuracy and conceptual grounding of future automated metaphor identification efforts. For a general-domain counterpart, see Maudslay and Teufel (2022).

6. Data Analysis

In terms of the annotation itself, we explored two main aspects: (1) the distribution of IAA across different domains to assess which concepts were more subjective or prone to disagreement, and (2) the relationship between metaphorical content and disagreement, based on the hypothesis that words which can be used either metaphorically or literally in the cancer context may lead to greater ambiguity for both manual and automated annotation.

Regarding agreement, as shown in Figure 4, the most frequent case was that of disagreement. This outcome is not unexpected given the complexity of metaphor annotation and is consistent with previous reports (Sánchez-Montero et al., 2025a). Interestingly, the second most frequent case was that of unanimous agreement, particularly in domains related to emotions. We hypothesize that agreement tends to increase for highly conventionalized metaphors, which are more easily recognized by annotators.

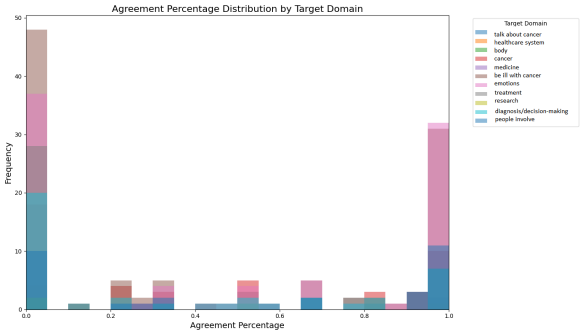

Figure 4: Distribution of agreement across semantic domains.

Secondly, we conducted a thematic analysis to address questions such as: *What are the most metaphorically discussed themes? Are war and journey the only ways of conceptualizing cancer or do other source domains prevail for particular topics?*

To understand how **target domains** (cancer-related topics such as treatment, diagnosis, or emotions) are expressed through **source domains** (the metaphorical images used to represent them, such as war or journey), we leveraged the BM annotations linked to WordNet synsets. These were then queried through the Open Multilingual WordNet (Vossen et al., 2016) to automatically group them into semantic domains.

Figure 5 displays the relationships between cancer-related topics (green nodes) and their metaphorical domains (blue nodes) as a network graph. The network's central node is the general topic of *being ill with cancer*, which connects to experiences like pain, death, healing, and relapse. This is followed by *emotions*, *people involved* (including emotional support, a key dimension in the forums as users sought shared experiences and advice), *cancer* itself, *diagnosis*, and *decision making*. Less frequently discussed topics, at least metaphorically, included *treatment*, *body*, *research*, and the *healthcare system*. This does not necessarily imply that these topics are less discussed overall, but rather that they appeared less often in this annotated sample of AECC posts.

A closer analysis of the largest nodes, *being ill*

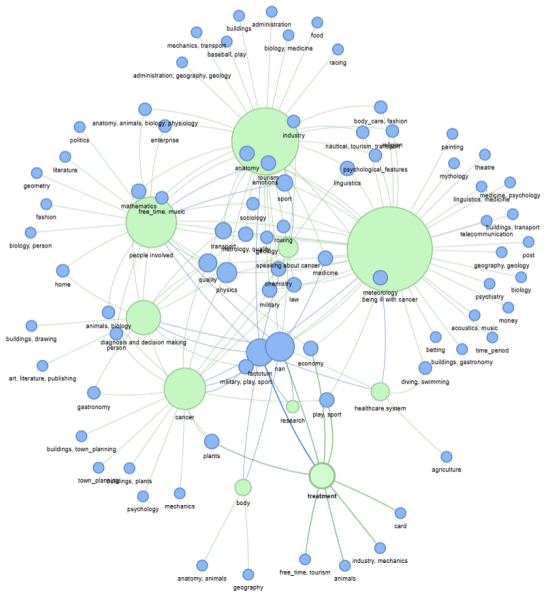

Figure 5: Relations between target (green) and source (blue) metaphorical domains.

with cancer and *emotions*, reveals that symptoms and side effects were among the most frequently discussed subtopics, followed by life changes, healing, and death. Regarding emotions, beyond general distress, positive emotions such as *optimism* and *hope* were notably prominent, reflecting the supportive tone of the forums, where users actively encourage others seeking solidarity and understanding.

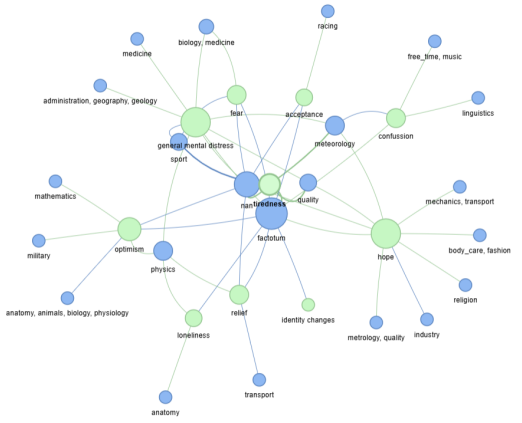

Figure 6: Source domains related to emotions.

In terms of source domains, the **military domain** was indeed central and frequently used across several topics (including emotions, illness, and cancer itself). However, other metaphorical frames were equally or more salient in specific contexts.

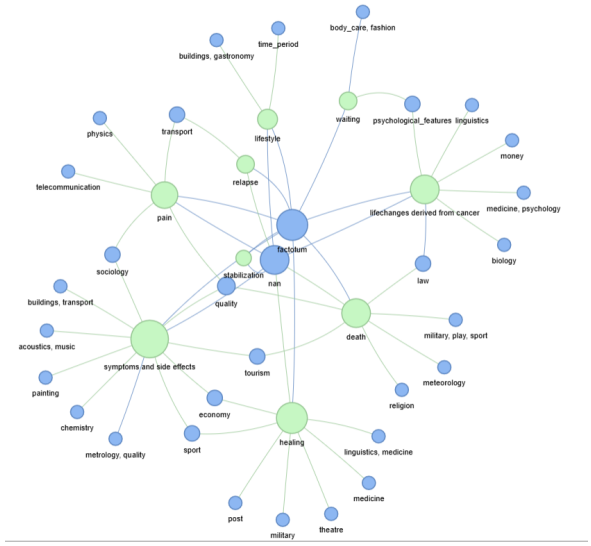

Figure 7: Source domains related to *being ill with cancer*.

For instance, metaphors related to **buildings, construction, and home** appeared recurrently when discussing emotions, emotional support, and even diagnosis. Likewise, **transport and mechanics** metaphors were common in discussions about emotions, cancer, and treatment. These findings demonstrate that Spanish cancer discourse, specifically narratives of cancer experience, draws from a far more diverse metaphorical repertoire than is typically acknowledged in the literature, which opens new avenues for the study of metaphors in Spanish biomedical communication.

7. Conclusions

The paper presents Medical-Flavors-AECC, a dataset of cancer narratives in Spanish extracted from digital forums. We propose a manual annotation framework based on distinguishing basic and contextual meaning, following MIP (Pragglejaz Group, 2007) and MIPVU (Steen et al., 2010), through the use of various lexical resources, such as WordNet, in different annotation layers.

This resource is intended for the automatic detection of metaphors, especially in the medical field, as well as for the linguistic study of cancer discourse. To this end, we offer an analysis of the metaphorical contents in Spanish cancer narratives, as well as the emotions involved in the production of these metaphors and the degree of lexicalization across different conceptual domains.

This research has significant social and humanistic impact. The study's results can help physicians and healthcare personnel improve their commu-

nication strategies when discussing the disease. At the same time, our results provide psychologists with an empirical basis for investigating how metaphors shape the perception of illness.

As a future work, we will seek to test alternative forms of annotation frameworks that allow for more extensive research into the subjective scope of metaphor and lexicalization processes in cancer narratives. Another area of current development is the annotation of a larger dataset using LLMs, including hybrid labeling approaches that combine human-AI collaboration. Future work will involve extending the annotation scheme with multiclass labels to distinguish between different metaphor types. This will enable a more fine-grained analysis of the specific metaphors characteristic of each conceptual domain.

8. Data Statement

We follow the guidelines specified by (Bender and Friedman, 2018) for creating a Data Statement to mitigate bias and ensure transparency in collection and annotation of the data.

A. Curation Rationale Posts were collected from the public forums of the Spanish Association Against Cancer (AECC). Given the association's prominence and its broad coverage of patients, relatives, and experiences across time, the AECC forums offered an ethically appropriate and representative source of cancer discourse in Spanish.

B. Language Variety The publications are written in Peninsular (European) Spanish. Orthographic and lexical variation was preserved during the annotation process to maintain linguistic authenticity, as the orthography employed in this type of public forum often deviates from normative conventions. Non-textual content such as images or links was removed, while the original punctuation, capitalization, and emoticons were retained.

C. Forums Authors Demographic The authors of the posts are patients, relatives, and caregivers participating in AECC online forums between 2010 and 2024. Since all forums contributions were publicly posted and pseudonymized, no personally identifiable or demographic information (e.g., age, gender, location) was collected. We expect diversity in socioeconomic background, education level, and regional origin among contributors since the AECC's broad reach in Spain.

D. Annotator Demographic The annotation team comprised five independent annotators, all native Spanish speakers previously trained in the Metaphor Identification Procedure (MIP) and with experience in metaphor annotation. Four were trained linguists, and one had a background in Natural Language Processing (NLP), Table 3 summarizes their profiles. All annotators provided informed consent, and the annotation process adhered to the principles of the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).

Categories	Data
Sex	2 female 3 male
Native Language	Spanish
Nationality	4 Spanish 1 Mexican
Residence	Spain and Mexico
Education level	University

Table 3: Annotator demographic

E. Speech Situation The forums posts are about personal experiences with cancer, including diagnosis, treatment, recovery, and emotional support. Topics vary from medical advice and coping strategies to interpersonal communication and testimonies. Posts were written asynchronously in a semi-formal or non-formal tone, frequently used in online patient support environments.

F. Text Characteristics The dataset includes 7,964 forum posts (5,641,005 words[3]), of which 74 posts (88,196 words) were manually annotated for metaphor use. Posts cover three main topics: breast cancer, prostate cancer, and personal testimonies. Texts were pseudonymized before annotation, with user mentions, personal names, and identifiable references removed or replaced following GDPR-compliant pseudonymization protocols (Parlamento Europeo y Consejo de la Unión Europea, 2016; Búrdalo, 2025; López-García et al., 2023). Additionally, each of the 74 posts includes comments that respond directly to the main topic introduced in the main post.

G. Recording Quality The text was processed into UTF-8 plain text format, ensuring consistency in encoding and tokenization. We scraped the posts from the AECC online forums. Descriptive statistics are presented in Table 4.

H. Ethical Considerations The data collection and annotation procedures were reviewed and approved by three independent entities: the AECC,

[3]Text splitting into words was done automatically by INCEpTION software.

Metric	Total	Annotated
Number of words	5,641,005	83,316
Number of posts	7,964	72
Number of users	2,976	442
Avg. words per post	708.30	
Avg. interactions (main author)	2.58	
Avg. interactions (comments)	12.20	
Avg. words per topic	Breast cancer = 504.5 Prostate cancer = 784.7 Testimonies = 749.8	
Covered years	2010-2024	

Table 4: Descriptive statistics of Medical-FLAVORS AECC full corpus and the presented annotated dataset.

the University of Zaragoza, and the EU project funding the research. All posts used were publicly available and pseudonymized to protect participants' identities. Annotators were fully briefed about the study's purpose and informed that the content could be emotionally sensitive. They were free to withdraw from participation at any point. Finally, this procedure followed the General Data Protection Regulation of the European Union (GDPR; Regulation (EU) 2016/679) (Parlamento Europeo y Consejo de la Unión Europea, 2016), established guidelines for anonymizing Spanish medical texts (Búrdalo, 2025), and existing anonymization tools for Spanish (López-García et al., 2023).

9. Limitations

We acknowledge some limitations in our study. First, although the dataset comprises over 83,000 manually annotated words, inter-annotator agreement on the curation could not be quantified, and the annotations result from consensus between two annotators and are then curated by a single curator.

Second, the calculation of metaphorical density relied in part on automatic lemmatization using SpaCy. During manual inspection, several lemmatization errors were identified, particularly involving irregular verb forms and clitic pronouns, which may have affected the precision of metaphor frequency estimates. Consequently, the metaphoricity percentages reported should be interpreted as approximate rather than exact.

Third, not all lexical items in the vocabulary could be assigned a WordNet synset. In some cases, no synset adequately captured the sense of a word as used in the dataset. These items were therefore excluded from the semantic domain analysis. Moreover, while WordNet domains are suitable for nominal concepts, many verbal synsets are assigned to the generic domain *factotum*, limiting their discriminative value. Future work should explore frame-based mappings (e.g., FrameNet) or manual domain annotation to improve coverage and granularity for verbs and other non-nominal items.

Finally, subsequent iterations of this resource will benefit from integrating domain-adapted lemmatization models and incorporating peer-reviewed quality control to further enhance reliability and reproducibility.

10. Acknowledgements

This paper has been supported by PAPIIT project IG-400325, by the I+D+i projects PID2024-159530OB-I00 (funded by MCIN/AEI/10.13039/501100011033), the EU research and innovation program HORIZON Europe in the "4D PICTURE" project under grant agreement 101057332 and UZ2024-lyA-02 (funded by Univ. Zaragoza), and by DGA Government predoctoral fellowship.

11. Bibliographical References

Margarita Alonso Ramos. 2004. *Las construcciones con verbo de apoyo*. Visor Libros.

Asociación de Academias de la Lengua Española. 2023. Diccionarios de la lengua española. Accessed: 2025-10-19.

Emily M Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Mayla Boguslav and Kevin Bretonnel Cohen. 2017. Inter-annotator agreement and the upper limit on machine performance: evidence from biomedical natural language processing. In *MEDINFO 2017: Precision Healthcare through Informatics*, pages 298–302. IOS Press.

Luis Búrdalo. 2025. D4.2 final harpocrates demonstrators - appendix i. Consultado: 2025-10-19.

Lynne Cameron. 2007. The affective discourse dynamics of metaphor clustering. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, (53):041–062.

David Casarett, Amy Pickard, Jessica M. Fishman, Stewart C. Alexander, Robert M. Arnold, Kathryn I. Pollak, and James A. Tulsky. 2010. Can metaphors and analogies improve communication with seriously ill patients? *Journal of Palliative Medicine*, 13(3):255–260. PMID: 19922170.

Gloria Corpas Pastor and Manuel Alvar Ezquerra. 1996. *Manual de fraseología española*. Gredos.

Laurence De Backer, Renata Enghels, and Patrick Goethals. 2023. Metaphor analysis meets lexical strings: finetuning the metaphor identification procedure for quantitative semantic analyses. *Frontiers in Psychology*, 14:1214699.

James Dickins. 2020. An ontology for collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs. *Linguistica Online*, 23:29–72.

Ellen Dodge, Jisup Hong, and Elise Stickles. 2015. MetaNet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49, Denver, Colorado. Association for Computational Linguistics.

Carla Parra Escartín, Almudena Nevado Llopis, and Eoghan Sánchez Martínez. 2018. Spanish multiword expressions: Looking for a taxonomy. *Multiword expressions*, page 271.

Nada Fadul, Ahmed F. Elsayem, J. Lynn Palmer, Egidio G Del Fabbro, Kay Swint, Zhijun Li, Valerie A Poulter, and Eduardo Bruera. 2009. Supportive versus palliative care: What's in a name? *Cancer*, 115.

Dana Gablasova, Vaclav Brezina, and Tony McEnery. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1):155–179.

Gangemi, Aldo and Presutti, Valentina and Navigli, Roberto and others. 2017. Framester: A wide-coverage linguistically and cognitively grounded knowledge graph. Accessed: 2025-10-20.

Mengshi Ge, Rui Mao, and Erik Cambria. 2023. A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artif. Intell. Rev.*, 56(S2):1829–1895.

Anna W. Gustafsson, Charlotte Hommerberg, and Anna Sandgren. 2019. Coping by metaphors: the versatile function of metaphors in blogs about living with advanced cancer. *Medical Humanities*, 46:267 – 277.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. The University of Chicago Press.

Mark Jordan Landau, Linda D. Cameron, Jamie Arndt, W. Hamilton, Trevor James Swanson, and Michael N. Bultmann. 2019. Beneath the surface: Abstract construal mindset increases receptivity to metaphors in health communications. *Social cognition*, 37 3:314–340.

Yufeng Liu, Elena Semino, Judith Rietjens, and Sheila Payne. 2024. Cancer experience in metaphors: patients, carers, professionals, students – a scoping review. *BMJ Supportive & Palliative Care*, 14(e3):e2366–e2376.

Guillermo López-García, Francisco J. Moreno-Barea, Héctor Mesa, José M. Jerez, Nuria Ribelles, Emilio Alba, and Francisco J. Veredas. 2023. Named entity recognition for de-identifying real-world health records in spanish. In *Proceedings of the 23rd International Conference on Computational Science (ICCS 2023)*, volume 14073 of *Lecture Notes in Computer Science*, pages 172–183. Springer. Consultado: 2025-10-19.

Dalia Magaña. 2020. Local voices on health care communication issues and insights on latino cultural constructs. *Hispanic Journal of Behavioral Sciences*, 42(3):300–323.

Rowan Hall Maudslay and Simone Teufel. 2022. Metaphorical polysemy detection: Conventional metaphor meets word sense disambiguation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 65–77.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The belmont report: Ethical principles and guidelines for the protection of human subjects of research. Technical report, U.S. Department of Health, Education, and Welfare, Washington, D.C. Publication No. OS 78-0012.

Parlamento Europeo y Consejo de la Unión Europea. 2016. Reglamento (ue) 2016/679 del parlamento europeo y del consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Consultado: 2025-10-19.

Inmaculada Penadés Martínez. 2020. Diccionario de locuciones idiomáticas del español actual. Accessed: 2025-10-19.

Ana P. M. Pinheiro, Rachel H. Pocock, Margie D. Dixon, Walid L Shaib, Suresh S. Ramalingam, and Rebecca D. Pentz. 2017. Using metaphors to explain molecular testing to cancer patients. *The oncologist*, 22 4:445–449.

Paula Tavares Pinto, Geraint Paul Rees, and Ana Frankenberg-Garcia. 2021. Identifying collocation issues in english l2 research article writing. In Maggie Charles and Ana Frankenberg-Garcia,

editors, *Corpora in ESP/EAP Writing Instruction: Preparation, Exploitation, Analysis*, chapter 7, page 24. Routledge.

Lucia Pitarch, Emma Angles-Herrero, Yufeng Liu, Daisy Monika Lal, Jorge Gracia, Paul Rayson, and Judith Rietjens. 2024. Medical-FLAVORS: A figurative language and vocabulary open repository for Spanish in the medical domain. In *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*, pages 109–114, Torino, Italia. ELRA and ICCL.

Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

Real Academia Española. 2023. Diccionario de la lengua española, 23.ª edición. Accessed: 2025-10-19.

Real Academia Española and Asociación de Academias de la Lengua Española. 2009. *Nueva Gramática de la Lengua Española*. Espasa Libros / RAE-ASALE, Madrid, Spain. Accessed: 2025-10-19.

Pavel Rychlý. 2008. A lexicographer-friendly association score. In *RASLAN*, pages 6–9.

Alec Sánchez-Montero, Gemma Bel-Enguix, Sergio-Luis Ojeda-Trueba, and Gerardo Sierra. 2025a. Disagreement in metaphor annotation of Mexican Spanish science tweets. In *Proceedings of Context and Meaning: Navigating Disagreements in NLP Annotation*, pages 155–164, Abu Dhabi, UAE. International Committee on Computational Linguistics.

Alec Sánchez-Montero, Gemma Bel-Enguix, Sergio-Luis Ojeda-Trueba, and Gerardo Sierra. 2025b. Prompting metaphoricity: Soft labeling with large language models in popular communication of science tweets in Spanish. In *Proceedings of the 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II)*, pages 45–56, Vienna, Austria. Association for Computational Linguistics.

Elena Semino and Zsofia Demjen. 2017. The cancer card: Metaphor, intimacy, and humor in online interactions about the experience of cancer. *Metaphor: Embodied cognition and discourse*, page 181.

Elena Semino, Zsófia Demjén, Andrew Hardie, Sheila Payne, and Paul Rayson. 2017. *Metaphor, cancer and the end of life: A corpus-based study*. Routledge.

Sketch Engine. 2025. Estenten corpus. Accessed: 2025-10-20.

Sociedad Española de Oncología Médica. 2025a. Cáncer de mama. Accessed: 2025-10-19.

Sociedad Española de Oncología Médica. 2025b. Los avances en cáncer de próstata. Accessed: 2025-10-19.

Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*, volume 14 of *Converging Evidence in Language and Communication Research*. John Benjamins Publishing Company.

Elise Stickles, Oana David, and Eve Sweetser. 2016. Grammatical constructions, frame structure, and metonymy: Their contributions to metaphor computation. In *Proceedings of the 11th meeting of the High Desert Linguistics Society*, pages 317–345. High Desert Linguistics Society Albuquerque, NM.

Orsolya Vincze, Elena Mosqueira, and Manuel Alonso Ramos. 2011. An online collocation dictionary of spanish. In *Proceedings of the 5th International Conference on Meaning-Text Theory*, pages 275–286, Barcelona, Spain.

Piek Vossen, Francis Bond, and John McCrae. 2016. Toward a truly multilingual GlobalWordnet grid. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 424–431, Bucharest, Romania. Global Wordnet Association.

World Health Organization. 2025. Breast cancer. Accessed: 2025-10-19.