A Comparative Analysis of Adjectives in Image Descriptions in Brazilian Portuguese and English

Maucha Andrade Gamonal[1], Adriana Pagano[2], Felix Engler[2], André V. Lopes Coneglian[2], Ely E. S. Mattos[3]

[1]Universidade Estadual de Campinas, [2]Universidade Federal de Minas Gerais, [3]Universidade Federal de Juiz de Fora
mgamonal@unicamp.br, apagano@ufmg.br, felixenglera@gmail.com, coneglian@ufmg.br, ely.matos@ufjf.edu.br

Abstract. *This paper investigates how adjectives construe visual experience in Brazilian Portuguese and English descriptions from the Framed Multi30K dataset. Grounded in Systemic Functional Grammar and Frame Semantics, the study annotates adjectives as experiential Epithets, interpersonal Epithets, or Classifiers, and examines the frames they evoke. The results show that experiential Epithets predominate, suggesting that adjectives mainly construe visually salient properties. Frame annotation shows that* `Color` *is the most frequent frame. By identifying salient adjectival meanings in human-produced descriptions, the study provides evidence of which visual attributes tend to be prioritized by human observers.*

Resumo. *Este artigo investiga o modo como os adjetivos representam a experiência visual em descrições, em português brasileiro e em inglês, do conjunto de dados Framed Multi30K. Com base na Gramática Sistêmico-Funcional e na Semântica de Frames, o estudo classifica os adjetivos em Epítetos experienciais, Epítetos interpessoais e Classificadores, além de examinar os frames por eles evocados. Os resultados mostram que os Epítetos experienciais são predominantes, o que indica que os adjetivos são empregados principalmente para representar propriedades visualmente salientes. A anotação semântica também revela que* `Color` *é o frame mais frequente. Ao identificar os significados adjetivais mais salientes, o estudo contribui para compreender quais atributos visuais tendem a receber maior destaque por parte dos observadores.*

1. Introduction

Image descriptions play a central role in accessibility resources, such as audiodescription, enabling blind and visually impaired people to access visual content. However, they remain frequently absent in online environments: according to Movimento Web Para Todos (2024), only 42.83% of 26.3 million Brazilian websites include image descriptions. This gap has motivated the development of automatic description systems, which rely on computer vision and natural language generation to identify visual content and transform it into language. However, accessible description requires more than the recognition of entities in an image: it also involves decisions about which attributes and meanings are relevant enough to be verbalized. Improving such systems therefore depends on a better understanding of how human-produced descriptions are linguistically structured and

which kinds of meanings they tend to foreground. Adjectives are central to this issue because they help construe visually salient properties, including color, size, age, material, posture, physical appearance, and evaluation.

This study investigates the use of adjectives in image descriptions in Brazilian Portuguese and English, based on a subset of the Framed Multi30K dataset [Viridiano et al. 2024]. The analysis focuses on 300 sentences describing 30 images, a sample selected to enable detailed manual annotation and qualitative control. Adjectives are classified following Systemic Functional Grammar [Halliday and Matthiessen 2013] and analyzed in relation to the entities they modify. In addition, Frame Semantics [Fillmore 1982] is used to examine the semantic frames evoked by these adjectives, with support from the LOME FrameNet parser [Xia et al. 2021], whose outputs are critically assessed.

As an exploratory study, this paper does not aim at large-scale generalization, but seeks to identify tendencies in adjective use in human-produced image descriptions and to discuss how adjectival meanings contribute to the selection and construal of visually salient attributes in multimodal and accessibility-oriented contexts. As [Coneglian and Pagano 2025] have shown, the Framed Multi30K image description constitute a prime terrain for the investigation of how lexicon and grammar are mobilized to construe meaning in language.

2. Theoretical Background

2.1. Systemic-Functional Grammar

Systemic-Functional Grammar [Halliday and Matthiessen 2013] views language as organized around three metafunctions: ideational, interpersonal, and textual. The ideational metafunction is associated with the representation of experience through processes, participants, and circumstances; the interpersonal metafunction with interaction, stance, and evaluation; and the textual metafunction with the organization of meanings in discourse.

The metafunctional perspective operates across the rank scale, from the clause to the morpheme, including the intermediate ranks of group/phrase and word. At each rank, units are analyzed in terms of both class and function: class identifies the type of unit, while function specifies the role that the unit performs in the structure of a higher-ranking unit. At group level, nominal groups typically function within the clause as participants, but they also have their own internal organization. Within the nominal group, words and word groups perform different functions, such as identifying, classifying, describing, or quantifying the entity being referred to. Among the distinct words classes that make up a nominal group, adjectives may function as Epithets, contributing descriptive or evaluative meanings, or as Classifiers, helping to categorize the entity as a particular type [Halliday and Matthiessen 2013]. Epithets express qualities attributed to an entity and may be either experiential, describing objective properties, or interpersonal, expressing subjective evaluation. Classifiers, in contrast, define subclasses of entities and are not typically gradable. One of the sentences in our corpus illustrates these different functions:

(1) A skier is overlooking the beautiful white snow-covered landscape.

In this example, *beautiful* functions as an interpersonal Epithet, expressing subjective evaluation, while *white* and *snow-covered* function as experiential Epithets, des-

cribing objective properties of the entity.

2.2. Frame Semantics

Frame Semantics [Fillmore 1982] assumes that word meaning is understood in relation to structured background knowledge, or frames, which are evoked during language use. Each frame represents a conceptual scenario involving specific participants and relations, enabling a detailed account of meaning beyond general semantic roles. In this study, frame semantics is used to examine the semantic frames evoked by adjectives in image descriptions. Frame annotation is supported by the LOME FrameNet parser [Xia et al. 2021].

A single sentence may evoke multiple frames. For example:

(2) The old man bought a pie.

In this sentence, *bought* evokes a `Commerce` frame, *pie* evokes a `Food` frame, and the adjective *old* evokes an `Age` frame[1].

This illustrates how adjectives contribute to the activation of specific domains of meaning within a broader semantic structure.

Adjectives are therefore analyzed in terms of the frames they evoke—such as `Color`, `Age`, or `Movement`—highlighting how different domains of meaning are activated in each language.

By combining functional and frame-semantic perspectives, this study identifies cross-linguistic tendencies in adjective use and contributes to the linguistic description of Brazilian Portuguese in multimodal contexts, as well as to discussions on semantic annotation practices.

2.3. A comparative typology of adjective lexemes in Brazilian Portuguese and English

Adjective lexemes pose an interesting challenge for inter-language comparison. In comparing Brazilian Portuguese, there are three aspects of adjectives lexemes that must be considered: (i) the lexical nature of the class; (ii) the grammatical behavior of such lexemes in each language; (iii) other language-specific means through which 'adjectival functions' may be realized.

First, Brazilian Portuguese and English are languages that have an open class of adjectives, and these are languages that present a fairly robust morphological derivational paradigm for the coinage of adjective words from nouns and verbs. Secondly, adjective lexemes differ in their grammatical behavior in both languages: while Brazilian Portuguese adjectives are grammatically aligned with nouns, English adjectives present a behavior that differs from both nouns and verbs [Dixon 2010]. This difference between the two languages poses a significant challenge in comparing captions of image descriptions.

Thirdly, both languages present different constructions that express the adjectival function of modification of an entity. In both languages, participles may be lexicalized

[1]By convention, italics are used in this paper both for emphasis on lexical items discussed in the analysis and for English translations of sentences originally written in Brazilian Portuguese. Monospaced font is used to represent frame names exactly as they appear in the FrameNet database

into functioning as modifying adjectives. And, most significantly, both languages present a construction [Preposition + Noun] that functions as a modifier of a nouns, as in 'mesa **de madeira**', *table **of wood***, in Brazilian Portuguese. These cases will not be considered in our analysis, since our focus is on the lexical expression on modification.

3. Resources used

3.1. FrameNet Brasil

FrameNet Brasil [Torrent et al. 2022] is a lexicographic annotation project based on Berkeley FrameNet, organized around the description of meaning through frames and their associated elements. The resource includes lexical units, frames, frame elements, and relations between frames, as well as extensions to multimodal semantic annotation. In recent years, FrameNet Brasil has also contributed to the development of multimodal datasets and annotation methodologies integrating textual and visual information [Viridiano et al. 2024] [Belcavello et al. 2024] , expanding the application of Frame Semantics to image and audiovisual description contexts.

Lexical units correspond to pairings of a word with a specific meaning that evokes a frame. For example, the adjective *white* may evoke different frames depending on its meaning, such as `Color` or `Race descriptor`. Frames represent structured conceptual scenarios, including participants and relevant attributes. Each frame is associated with a set of frame elements, which encode semantic roles and are typically classified as core, peripheral, or extra-thematic [Ruppenhofer et al. 2016].

In this study, FrameNet Brasil provides the semantic framework used to analyze meanings evoked by adjectives in image descriptions, supporting the investigation of cross-linguistic semantic variation in multimodal contexts.

3.2. LOME FrameNet Parser

LOME FrameNet parser [Xia et al. 2021] is a multilingual information extraction system designed to identify semantic frames, frame elements, and their corresponding trigger spans directly from raw text. Unlike parsers conditioned on predefined targets, LOME performs full-structure prediction, extracting all frames and roles present in a sentence. The model was trained on annotated data in multiple languages, including Portuguese and English.

In this study, LOME was used to re-analyze the selected image descriptions in both languages, with the specific goal of identifying frames evoked by adjectives. Its outputs were subsequently reviewed to ensure analytical reliability, particularly in cases of misidentification or omission. This procedure enabled both the examination of frame distribution across languages and the assessment of the parser's performance in a multimodal annotation context.

3.3. Framed Multi30k dataset

Framed Multi30K [Viridiano et al. 2024] is a multimodal dataset that extends Flickr30K by incorporating multilingual descriptions and frame-semantic annotations. The original Flickr30K dataset contains 31,783 images of everyday scenes, each paired with five English descriptions. Framed Multi30K expands this resource by including both translated

and original descriptions in Brazilian Portuguese, as well as frame and frame element annotations aligned with visual regions.

An example from the dataset is presented in Figure 1 with two English and two Brazilian Portuguese descriptions of the same image:

Figura 1. Example of image descriptions (Flickr30K 1003420127)

(3) A group of adults, inside a home, sitting on chairs arranged in a circle, playing musical instruments.

(4) Five musicians, a man and four women, practicing sheet music (using flutes) in a living room.

(5) Cinco pessoas em uma sala tocam flauta sentadas em cadeiras.
(Five people in a room play the flute sitting on chairs.)

(6) Um grupo de músicos reunidos em círculo estão sentados em cadeiras em uma sala clara e residencial segurando instrumentos e próximos a estantes de partitura.
(A group of musicians gathered in a circle are sitting on chairs in a light, residential room holding instruments and near sheet music stands.)

4. Methodology

A sample of 30 images was randomly selected from the Framed Multi30K dataset [Viridiano et al. 2024]. The dataset includes five original descriptions in English and five in Brazilian Portuguese per image, resulting in a total of 300 sentences. The selection of this subset allows for detailed manual annotation and controlled comparison across languages, prioritizing analytical depth over large-scale generalization.

All descriptions were organized on a spreadsheet. Adjectives were manually identified and classified according to Systemic Functional Grammar [Halliday and Matthiessen 2013] into experiential Epithets, interpersonal Epithets, and Classifiers. In addition, adjectives were annotated according to the type of entity they modify (person, clothing, object, location, animal, or other).

Following this step, all sentences were processed using the LOME FrameNet parser [Xia et al. 2021]. Parser outputs were extracted and manually reviewed to identify errors, including missing annotations and incorrect frame assignments. Only lexical unit labels were retained for analysis, as they directly indicate the frames evoked by adjectives, as indicated in §2.3.

5. Results and Discussion

5.1. Adjective Functions

A total of 182 adjectives were identified in the Brazilian Portuguese descriptions and 165 in the English ones. In both languages, experiential Epithets predominate (85.71% in Portuguese; 81.21% in English), followed by Classifiers, while interpersonal Epithets are rare and occur only in English (1.21%).

This distribution reflects the descriptive nature of image captions, which prioritize observable and visually salient properties. It may also be related to the controlled conditions under which the descriptions were produced, as annotators were asked to describe the images without contextual information (see [Viridiano et al. 2024]. (7) and (8) are descriptions of the same image and illustrate experiential epithets in the corpus:

(7) Um cachorro *preto* e *branco* salta para pegar um brinquedo no gramado de um quintal.
(A black and white dog jumps to get a toy on the lawn of a yard.)

(8) A *black* and *white* dog jumps to get the Frisbee.

The adjectives *preto*, *branco*, *black*, and *white* describe observable physical properties that help distinguish the central entity in the image. Such adjectives are particularly frequent in image descriptions because they provide visually accessible information that is not necessarily specified by the noun itself.

Classifiers occur less frequently because they typically designate stable and well-defined classes of entities. In Flickr30K, however, many of the entities are already basic, visually identifiable participants, such as men, women, children, and animals. As a result, they are more often modified by adjectives expressing qualities, appearance, size, color, age, or evaluation than by classifiers assigning them to specialized taxonomic classes. Examples from our corpus are:

(9) Uma equipe *médica* realiza uma operação.
(A medical team performing an operation.)

(10) *Surgical* team performing surgery.

In (9) and (10), *médica* and *surgical* define an institutional role associated with the noun.

Interpersonal Epithets appear only in English descriptions.

(11) Ballet dancers in a studio practice jumping with *wonderful* form.

In sentence (11), the adjective *wonderful* expresses a subjective evaluation rather than a directly observable property. Unlike experiential Epithets, which tend to describe perceptible attributes, interpersonal Epithets foreground the annotator's stance toward the image.

However, even experiential Epithets involve interpretative variation. Although they are treated as descriptors of observable properties, their use still depends on perceptual categorization. The same visual property may be categorized differently by different annotators, especially in borderline cases such as color distinctions. This indicates that even descriptions that appear to be objective involve choices about how visual information is selected and verbalized.

5.2. Adjectives Across the Two Languages

There were 47 instances of one-to-one adjective correspondences across descriptions of the same image in the two languages. Most correspondences involved experiential Epithets, especially color adjectives, suggesting that visually salient properties tend to be consistently selected by annotators in both languages.

Some examples include:

(12) Criança de vestido *rosa* entrando em uma pequena casa de madeira. *(Child in a pink dress going into a small wooden house.)*

(13) A little girl in a *pink* dress going into a wooden cabin.

The predominance of color adjectives among adjective correspondences suggests that color constitutes one of the most consistently verbalized types of visual information across languages. This tendency reinforces the relevance of visually prominent attributes in image descriptions.

However, equivalence at the level of meaning does not necessarily entail equivalence in grammatical realization. In several cases, meanings realized by adjectives in one language were encoded through other grammatical resources in the other language. From an SFL perspective, this shows that similar experiential meanings may be construed through different configurations of the nominal group or redistributed across other elements of the clause.

For example, in (12) the Portuguese prepositional phrase *de madeira* realizes the same experiential meaning as the English adjective *wooden* in (13). Similar shifts in lexicogrammatical realization occur in descriptions of posture, bodily position, and material composition. These meanings may be construed through different resources across the two languages, including adjectives within the nominal group, verbal processes in the clause, or prepositional phrases functioning as Qualifiers. The following image, Fig. 2, and its descriptions illustrate this tendency:

Figura 2. Image 1000919630 from the Flickr30K dataset)

(14) Um rapaz *sentado* em uma cadeira brinca com um leão *de pelúcia*.
(A guy sitting on a chair plays with a stuffed lion.)

(15) Um jovem *sentado* em um banco e vestindo bermuda, tênis e camiseta segura um leão de pelúcia e mexe na boca do brinquedo.
(A youth sitting on a stool and wearing shorts, sneakers and a T-shirt holds a stuffed lion and touches the toy's mouth.)

(16) A man sits in a chair while holding a large *stuffed* animal of a lion.

(17) A man is sitting on a chair holding a large *stuffed* lion.

These examples show a difference in grammatical realization. The meaning of material or composition is expressed in Portuguese through the prepositional phrase *de pelúcia*, while in English it is realized adjectivally through *stuffed*.

These patterns indicate that cross-linguistic differences in the distribution of adjectives are not only a matter of lexical preference. They also reflect broader differences in how comparable experiential meanings are mapped onto lexicogrammatical resources in each language.

5.3. LOME FrameNet Parser Performance and Limitations

Following the manual annotation of adjective functions and described entities, all sentences were processed using the LOME FrameNet parser [Xia et al. 2021] in order to identify semantic frames evoked by adjectives. The parser generated 445 labels for the 347 adjectives identified in the dataset, including both lexical unit and frame element annotations.

Most labels were correctly assigned, although three main error types were identified during manual revision: missing annotations, incorrect frame assignments, and incorrect frame element assignments. In total, 102 parser errors were identified, including 71 unlabeled adjectives, 26 incorrect frame assignments, and 5 incorrect frame element assignments. Although only 6.97% of the generated labels were incorrect, 20.46% of the adjectives in the dataset were left unlabeled.

One example of incorrect frame assignment occurs in the following sentence:

(18) A bunch of people in racing gear and helmets racing bicycles around a corner on *wet* pavement.

In this case, the adjective *wet* was incorrectly associated with the frame `Emptying` instead of `Being_wet`.

The parser also failed to identify some adjectives entirely. For example, in the sentence:

(19) Two *young* guys with *shaggy* hair look at their hands while hanging out in the yard.

While *young* was correctly associated with the frame `Age`, the adjective *shaggy* received no annotation.

Despite these limitations, the parser proved useful for large-scale semantic annotation, particularly when combined with manual curation. The results also highlight challenges involved in automatic frame-semantic annotation in multimodal contexts, especially in cases where semantic interpretation depends on visual or contextual information not directly available in the textual input.

5.4. Frames evoked by Adjectives

After removing incorrect annotations, only lexical unit labels were considered in the analysis of frames evoked by adjectives, since they directly indicate the semantic frames associated with each lexical unit. A total of 41 distinct frames were identified across the dataset.

The most frequent frame in both languages was `Color` (49.50% in English; 39.54% in Portuguese), reflecting the prominence of color-related information in image descriptions. Other recurrent frames include `Age` and `Size` in English, and `Body movement` and `Change Posture` in Brazilian Portuguese.

The predominance of the `Color` frame is consistent with the high frequency of color adjectives identified throughout the dataset, indicating that color is one of the most consistently verbalized types of visual information across languages.

While the two languages showed relatively similar distributions of adjective functions, frame analysis revealed more substantial semantic differences. Functional categories such as experiential Epithets capture broad grammatical behavior, whereas frame-semantic analysis highlights finer distinctions in the types of meanings associated with adjectives across languages. Portuguese descriptions showed greater frame diversity overall, with 32 distinct frames, compared to 16 in English. English descriptions, in contrast, concentrated a larger proportion of adjectives within a smaller set of semantic domains, particularly `Color` and `Age`. More substantial differences emerge in descriptions involving bodily position and posture. In Brazilian Portuguese, these meanings are frequently encoded through adjectives:

> (20) Uma pessoa, de bermuda *azul*, *deitado* no chão ao lado de carros.
> *(A person, in blue shorts, lying on the ground next to cars.)*
>
> In English, equivalent semantic content is often realized through verbal constructions:
>
> (21) A man in *blue* shorts is *lying* in the street.

In sentence (20), the adjective *deitado* evokes the `Change Posture` frame, whereas in (21) the same semantic information is encoded through the verb *lying*. As a result, certain frames become associated with adjectives in Portuguese, but not in English, despite being conceptually present in both descriptions. These patterns suggest that cross-linguistic differences in frame distribution may be related to semantic, typological, and syntactic differences in how visual information is linguistically encoded.

6. Conclusions

This study investigated the use of adjectives in image descriptions in Brazilian Portuguese and English by combining Systemic Functional Grammar and Frame Semantics in a multimodal analytical framework. Rather than focusing exclusively on grammatical categories, the study examined how adjectival meanings contribute to the semantic representation of visual information across languages.

The results showed that experiential Epithets predominate in both datasets, reflecting the descriptive and perceptually oriented nature of image captions. At the semantic level, the analysis revealed the prominence of visually salient domains such as `Color`,

while also identifying relevant cross-linguistic differences in the distribution of semantic frames.

The findings also demonstrate that semantically equivalent information is not necessarily realized through equivalent grammatical structures across languages. Information expressed adjectivally in one language may be encoded through verbal or prepositional constructions in the other, indicating that cross-linguistic variation in image descriptions involves not only lexical choice, but also broader semantic and typological preferences in how visual experience is linguistically structured.

From a methodological perspective, the study suggests that the combination of functional and frame-semantic analyses provides complementary insights into multimodal language description. While Systemic Functional Grammar captures broad grammatical tendencies, Frame Semantics enables a more fine-grained analysis of the conceptual domains activated by adjectives and their relation to visual interpretation. This distinction becomes particularly relevant in multimodal contexts, where different linguistic strategies may encode similar perceptual information.

The results also reinforce the relevance of frame-based semantic annotation for multimodal resources and accessible image description. Understanding how languages differ in the semantic organization of visual information may contribute to the development of more linguistically sensitive multimodal annotation practices and automatic caption generation systems.

Although exploratory and based on a limited dataset, this study contributes to discussions on cross-linguistic multimodal description, semantic annotation, and the linguistic representation of visual experience. Future research may expand the corpus, include additional image genres, and further investigate how semantic frames interact with translation, accessibility, and multimodal language technologies.

7. Acknowledgments

The authors thank two anonymous reviewers whose comments helped us improve this article. The usual disclaimer applies. Maucha A. Gamonal was funded by CAPES (88887.015648/2024-00) and is currently supported by H.IAAC 01245.003479/2024-10. Adriana Pagano has grants from the National Council for Scientific and Technological Development (CNPq 404722/2024-5; 313103/2021-6) and Minas Gerais State Agency for Research and Development (FAPEMIG). André V. Lopes Coneglian's research is funded by CNPq (153972/2025-4).

8. Limitations

The study is exploratory and has some limitations. The analysis is based on a small sample of 30 images from the Framed Multi30K dataset, resulting in 300 sentences, and the findings should therefore be interpreted as tendencies rather than as statistically generalizable patterns. In addition, the sample does not cover the full range of possible image types, descriptive styles, or annotator variation. The frame-semantic annotation was supported by the LOME parser and manually revised, so the results depend both on the parser's coverage and on interpretive decisions made during revision. Finally, the focus on adjectives necessarily excludes other lexicogrammatical resources through which

similar visual meanings may be realized, such as verbal and prepositional constructions. Future work should expand the sample and examine how visual attributes are distributed across different grammatical resources.

9. Ethics statement

This study does not involve collecting data from human participants or sensitive information of any kind. The Framed Multi30K dataset is openly accessible. All tools used for corpus compilation and annotation, code generation, and data analysis are freely available for research use.

10. Generative AI usage

No scientific content, analyses, or interpretations were produced by AI.

11. References

Referências

Belcavello, F., Torrent, T. T., Matos, E. E., Pagano, A. S., Gamonal, M., Sigiliano, N., Dutra, L. V., de Andrade Abreu, H., Samagaio, M., Carvalho, M., Campos, F., Azalim, G., Mazzei, B., de Oliveira, M. F., Loçasso Luz, A. C., Pádua Ruiz, L., Bellei, J., Pestana, A., Costa, J., Rabelo, I., Silva, A. B., Roza, R., Souza, M., and Oliveira, I. (2024). Frame2: A FrameNet-based multimodal dataset for tackling text-image interactions in video. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 7429–7437, Torino, Italia. ELRA and ICCL.

Coneglian, A. V. L. and Pagano, A. (2025). A gramática em datasets multimodais: um estudo de caso de legendas de imagens. *Caligrama: Revista de Estudos Românicos*, 30(1):24–51.

Dixon, R. M. (2010). *Basic linguistic theory volume 2: Grammatical topics*, volume 2. OUP Oxford.

Fillmore, C. J. (1982). Frame semantics. In Linguistics Society of Korea, editor, *Linguistics in the Morning Calm*, pages 111–138. Hanshin Publishing Co., Seoul, South Korea.

Halliday, M. A. K. and Matthiessen, C. M. I. M. (2013). *Halliday's Introduction to Functional Grammar*. Routledge.

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., Baker, C. F., and Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, CA.

Torrent, T. T., Matos, E. E. d. S., Belcavello, F., Viridiano, M., Gamonal, M. A., Costa, A. D. d., and Marim, M. C. (2022). Representing context in framenet: A multidimensional, multimodal approach. *Frontiers in Psychology*, 13.

Viridiano, M., Lorenzi, A., Torrent, T. T., Matos, E. E., Pagano, A. S., Sathler Sigiliano, N., Gamonal, M., de Andrade Abreu, H., Vicente Dutra, L., Samagaio, M.,

Carvalho, M., Campos, F., Azalim, G., Mazzei, B., Fonseca de Oliveira, M., Luz, A. C., Padua Ruiz, L., Bellei, J., Pestana, A., Costa, J., Rabelo, I., Silva, A. B., Roza, R., Souza Mota, M., Oliveira, I., and Pelegrino de Freitas, M. H. (2024). Framed Multi30K: A frame-based multimodal-multilingual dataset. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 7438–7449, Torino, Italia. ELRA and ICCL.

Xia, P., Qin, G., Vashishtha, S., Chen, Y., Chen, T., May, C., Harman, C., Rawlins, K., White, A. S., and Van Durme, B. (2021). LOME: Large ontology multilingual extraction. In Gkatzia, D. and Seddah, D., editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 149–159, Online. Association for Computational Linguistics.