Framed Multi30K

Authors

Marcelo Viridiano[1], Arthur Lorenzi Almeida[1] Tiago Timponi Torrent[1 2], Ely Edison da Silva Matos[1], Adriana Silvina Pagano[2 3], Natália Sathler Sigiliano[1], Maucha Gamonal[1 3], Helen de Andrade Abreu[1], Lívia Vicente Duarte[1 4], Mairon Samagaio[1], Mariane Carvalho[1], Franciany Campos[1], Gabrielly Azalim[1], Bruna Mazzei[1], Mateus Oliveira[1], Ana Carolina Luz[1], Livia Padua Ruiz[1], Júlia Bellei[1], Amanda Pestana[1], Josiane Costa[1], Iasmin Rabelo[3], Anna Beatriz Silva[3], Raquel Roza[3], Mariana Mota[3], Igor Oliveira[3], and Márcio Freitas[3].

Affiliations

1. FrameNet Brasil, Federal University of Juiz de Fora
2. Brazilian National Council for Scientific and Technological Development – CNPq
3. LETra, Federal University of Minas Gerais
4. Masters in Language Technology, Gothenburg University

A Frame-Based Multimodal-Multilingual Dataset

Built upon benchmark image-caption datasets

FLICKR30K EXTENSIONS

Multi30K
Elliott et al., 2016

+ 31K German translations;
+ 158K original German descriptions (5 per image).

Flickr30K Entities
Plummer et al., 2015

+ 244K coreference chains;
+ 276k manually annotated bounding boxes.

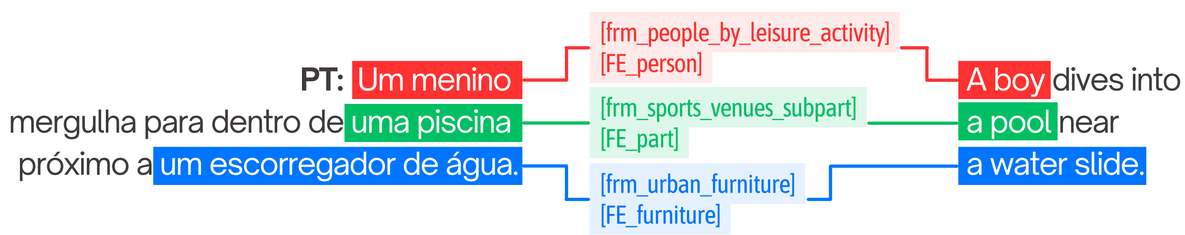
EN: A boy dives into a pool near a water slide.
DE: Ein Junge taucht in der Nähe einer Wasserrutsche in ein Schwimmbecken.

OUR NEW DATASET

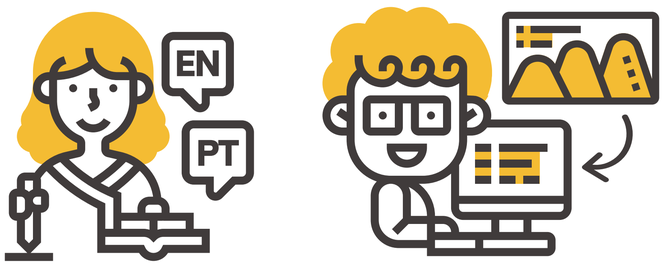
Framed Multi30K
Viridiano et al., 2024

+ 31K Brazilian Portuguese translations;
+ 158K Brazilian Portuguese descriptions (5 per image);
+ 4 Million frames and Frame Element labels to the English descriptions, and Brazilian Portuguese descriptions and translations.

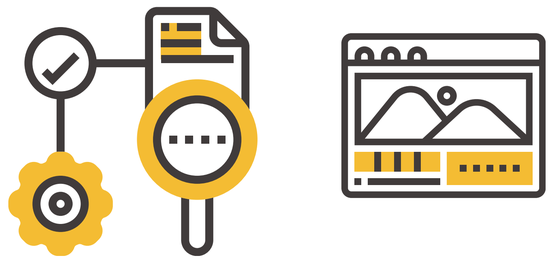
+ 169K manually annotated frames and Frame Elements correlations to the existing phrase-to-region correlations of Flickr30K Entities.

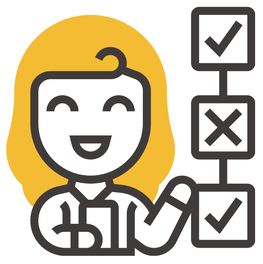


Created and validated by humans



- 15-hour workshop on the annotation tool and methodology;
- Hands-on practice sessions with test subsets of the original corpus;
- Weekly alignment meetings with other annotators and the authors.





31K Portuguese translations + 158K originaldescriptions
created by Brazilian undergraduate and graduate students with advanced proficiency in English.

Frames and Frame Element information
added via LOME (Xia et al., 2021) to the English descriptions and Portuguese originals, and manually to the Portuguese translations.

Quality of annotations
evaluated both manually, by periodically checking subsets, and with automated methods.

Developed for

1. Expanding FrameNet beyond text
Integrating image annotation within FrameNet marks a departure from a three-decade tradition of solely text-based annotation.

2. Adding Fine-Grained Semantics and Perspective
FrameNet labels should augment sensibly the granularity and informativeness of both Multi30K and Flickr30K Entities datasets.

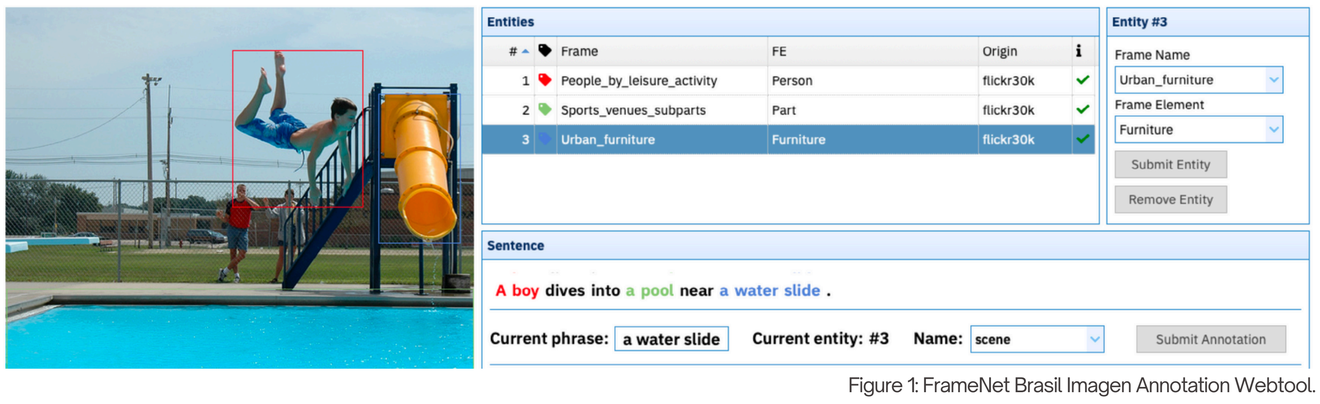
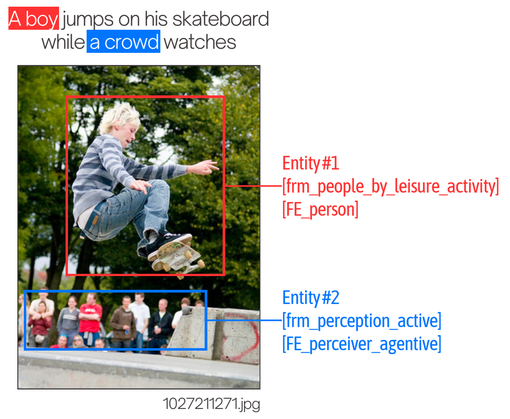
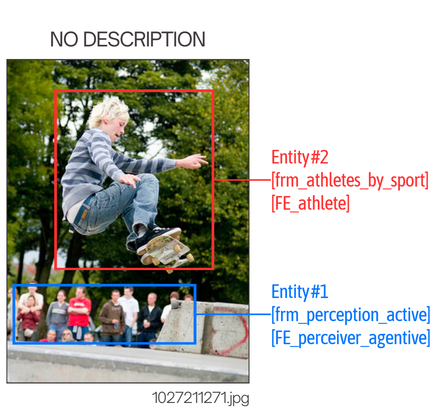


Figure 1: FrameNet Brasil Imagen Annotation Webtool.





3. Increasing the representation of Portuguese in NLP
Including Brazilian Portuguese into Multi30K adds one of the top ten most spoken languages in the world to the Multi30K dataset family.

Next Steps

1. **Annotations for Event Frames:** These would further link entities within the same image as participants of the same scene and improve the semantic descriptions of the situation being presented.

2. **English-Portuguese description alignment:** By aligning the noun phrases from the English original descriptions with their respective Brazilian Portuguese translations, we expect to be able to use the manually annotated bounding boxes to also assign frames and frame elements to image-text correlations in our translated descriptions.

3. **Frame-Based Multimodal MT Model:** Upon completing the annotation for Event Frames and the alignment tasks between English and Portuguese, we plan to train MT models using various combinations of this dataset, followed by assessment of the performance of each model.

Acknowledgements
Authors acknowledge the support of the Grad- uate Program in Linguistics at the Federal Uni- versity of Juiz de Fora. Research presented in this paper is developed by ReINVenTA - Re- search and Innovation Network for Vision and Text Analysis of Multimodal Objects. ReIN- VenTA is funded by FAPEMIG grant RED 00106/21, and CNPq grants 408269/2021-9 and 420945/2022-9. Viridiano's research was funded by CAPES PROBRAL PhD exchange grant 88887.628830/2021-00 and CAPES PROEX PhD Grant 88887.816219/2023-00. Lorenzi's research was funded by CAPES PROBRAL PhD ex- change grant 88887.628831/2021-00 and CAPES PROEX PhD Grant 88887.816228/2023-00. Tor- rent is an awardee of the CNPq Research Pro- ductivity Grant number 315749/2021-0. Pagano is an awardee of the CNPq Research Produc- tivity Grant number 313103/2021-6. Gamonal's research was funded by CAPES/PRINT grant 88887.936139/2024-00.

Visit **viridiano.com** to learn more and **download FM30K**