

# A FrameNet Brasil Approach to Annotation of Pragmatic Frames Evoked by Turn Organization Gestures

## *Uma abordagem da FrameNet Brasil para anotação de frames pragmáticos evocados por gestos de organização de turno*

**Helen de Andrade Abreu**

Federal University of Juiz de Fora (UFJF)

Juiz de Fora | MG | BR

FAPEMIG

CNPq

[helen.abreu@estudante.ufjf.br](mailto:helen.abreu@estudante.ufjf.br)

<https://orcid.org/0000-0003-1228-7144>

**Ely Edison da Silva Matos**

Federal University of Juiz de Fora (UFJF)

Juiz de Fora | MG | BR

[ely.matos@ufjf.br](mailto:ely.matos@ufjf.br)

<https://orcid.org/0000-0002-9464-9315>

**Abstract:** This paper presents the manner in which FrameNet Brasil has developed annotation of pragmatic frames, with a focus on those evoked by interactive gestures used in turn organization during face-to-face conversation. As corpus, we chose the interviews shown in the episodes of the first season of the Brazilian television series *Pedro Pelo Mundo*. We used Webtool 4.0, the newest version of this tool, for annotation of a total of 47 interactive gestures. As a result, we were able to demonstrate that these interactive gestures evoke specific pragmatic frames – turn passing, turn confirmation and turn taking – and are tools which make conversation a multimodal means of communication. Our research shows that the annotation of pragmatic frames is not only possible, but also useful for pragmatic investigations and for expanding FrameNet multimodal analysis.

**Keywords:** pragmatic frames; conversational turn organization; multimodality.

**Resumo:** Este artigo apresenta como a FrameNet Brasil desenvolveu a anotação de frames pragmáticos, com foco nos que são evocados por gestos interativos usados na organização de turnos durante conversas face a face. Como *corpus*, escolhemos as entrevistas mostradas nos episódios da primeira temporada da série de televisão brasileira *Pedro Pelo Mundo*. Utilizamos a Webtool 4.0, a versão mais nova dessa ferramenta, para anotação de um total de 47 gestos interativos. Como resultado, fomos capazes de demonstrar que esses gestos interati-



vos evocam frames pragmáticos específicos – passagem de turno, confirmação de turno e tomada de turno – e são ferramentas que tornam a conversa um meio multimodal de comunicação. Nossa pesquisa mostra que a anotação de frames pragmáticos não é apenas possível, mas também útil para investigações pragmáticas e para expandir a análise multimodal da FrameNet.

**Palavras-chave:** frames pragmáticos; organização de turno conversacional; multimodalidade.

## 1 Introduction

Charles Fillmore first developed the concept of frames in a series of articles (Fillmore, 1975, 1982, 1985, 2008; Fillmore *et al.*, 2003a, 2003b, 2009) encompassing three decades. He defined a frame as a “system of concepts” (Fillmore, 1982, p. 111) which works in a way that, for a person to understand one of these concepts, it is necessary to understand the whole system or structure in which it is interconnected. In this way, when one of these concepts is evoked, all the other concepts in the system are automatically made available to our conceptualization. When a person mentions the fact that they have purchased something, for example, the whole structure (the frame) in which there is a buyer, goods, a seller, an amount of money used in the purchase, etc. (the frame elements) is made available to our conceptualization, allowing us to understand the topic of the conversation and, furthermore, to elaborate upon that.

Using his frame semantics theory as a basis, Fillmore (Fillmore *et al.*, 2003a, 2003b) decided to develop a lexicographic database which would permit annotation of frames that would in turn be used for both linguistic research and machine learning development. For this purpose, he created the FrameNet initiative in Berkeley, California, in 1997 (Fillmore *et al.*, 2003a, 2003b). Within the initiative, data were to be collected from corpora composed exclusively of real use of English. Since then, other FrameNets have been developed for studies in other languages (Subirats-Rüggeberg; Petruck, 2003; You; Liu, 2005; Boas; Ziem, 2018; Gruzitis *et al.*, 2018; Ohara *et al.*, 2018; Hahm *et al.*, 2020). To work with Brazilian Portuguese, FrameNet Brasil was developed at the Federal University of Juiz de Fora in 2009.

FrameNet Brasil developed its own annotation tool, Webtool (Torrent; Ellsworth, 2013; Torrent *et al.*, 2024), which has recently been updated to the version Webtool 4.0. Among the other FrameNets, FrameNet Brasil stands out, because it has developed into a multimodal database (Torrent; Ellsworth, 2013; Belcavello *et al.*, 2022; Viridiano *et al.*, 2022; Luz *et al.*, 2023). Webtool 4.0 can work with frames within texts as well as those evoked by visual images – both static and dynamic – thus making it possible for the database to work with textual frames and visual frames separately or together. This has greatly enriched the realm of research possibilities within FrameNet Brasil.

As part of FrameNet Brasil, ReINVenTA (Research and Innovation Network for Vision and Text Analysis) is a network comprised of research groups and laboratories from these universities: UFJF, UFMG, UFU and PUC-MG. This network was created for the development

of a computational model for annotation of multimodal objects. At present, ReINVenTA's dataset is composed of three corpora – Audition (Dornelas; Gamonal; Pagano, 2024); Framed Multi 30K (Viridiano *et al.*, 2022; Viridiano *et al.*, 2024); and Frame2 (Belcavello *et al.*, 2024), on which our project is based.

## 1.1 Frame<sup>2</sup>

Frame<sup>2</sup> is a multimodal dataset (Belcavello *et al.*, 2024). For the construction of this dataset, a Brazilian television series, *Pedro Pelo Mundo* (*Pedro Around the World*), was selected. FrameNet was granted written permission by the copyright owners of the series, so that the first season of the series could be used for research purposes.

The first season of *Pedro Pelo Mundo* was aired in Brazil in 2016 on the cable television channel GNT. It was composed of 10 episodes with duration of 23 minutes each, which resulted in 230 minutes of video in total. Pedro Andrade, the presenter, visited a different country in each episode, focusing on locations that had recently undergone a social transformation that could be economic, cultural, governmental, and so forth. During each episode, Pedro Andrade introduced the viewer to beautiful locations, tried local food and drinks while describing his experience, interacted with local people, and interviewed people who had been living there – especially, although not exclusively, Brazilian people. For this reason, several of the interviews in the series were conducted in Portuguese. Other interviews, however, were conducted in English and a few in Spanish. The interviews occurred in different places, including restaurants and parks, and some of them occurred while Pedro Andrade and his interviewee were walking on the street, on their way to a specific location. This variety of settings proved to be an asset for our research.

This audiovisual resource was used to create a dataset composed of annotated data collected from all the verbal language in the series (using both the transcribed audio portion of the series and the subtitles present on screen when the interviews were not conducted in Portuguese), and also from the visual portion of the series. The transcribed audio portion of the ten episodes, as well as all the subtitles, considered together as text, have been lexically annotated for frames and their corresponding frame elements, generating 16,458 annotation sets. The visual portion of the series has been annotated for objects recognizable by computational vision and for the frames evoked by them, generating 7,520 bounding boxes.

As explained by Belcavello *et al.* (2020), images and text may work together in the production of meaning. Taking this into consideration, the video annotators for Frame<sup>2</sup> dataset used a multimodal approach in completing their task (Belcavello *et al.*, 2024): the annotation of images started only after the text annotation was completed; when the visual annotators were working on creating the image dataset, they were instructed to first watch the video (with the audio on); after that, they were required to read each annotated sentence and then watch the video sequence related to that sentence, looking either for the objects described in it or for objects that were somehow related to these sentences. For example, Belcavello *et al.* (2024) describe how, during the annotation process, a graffiti on a wall was considered the visual manifestation of Behavior, a frame element in the frame Mental\_property, evoked by the lexical unit *criativo* ('creative') in the sentence *o povo voltou a ser criativo* ('the people became

creative once more'). In this way, it was demonstrated that this multimodal approach adds layers of meaning to FrameNet capabilities.

For its multimodal capabilities, Frame<sup>2</sup> dataset proved to be invaluable for our project.

## 1. 2 Pragmatic Frames

In the article "Beyond Lexical Semantics", Czulo, Ziem and Torrent (2020) discuss the usefulness of creating a way for pragmatic frames to be annotated and become part of the FrameNet dataset. As the authors point out, Berkeley FrameNet, founded by Fillmore, was created to deal with lexicographical problems (Fillmore; Baker, 2009, for example), and as other FrameNets for other languages were created, their aim remained the same. For this reason, they do not have the necessary apparatus for annotation of the pragmatic aspects of terms and expressions. However, Czulo, Ziem and Torrent (2020) demonstrate the necessity of recognizing and working with the pragmatic nature of certain frames.

Czulo, Ziem and Torrent (2020) make it clear that their proposal is not to modify any existing theory. Both Frame Semantics (Fillmore, 1982) and Frame Analysis (Goffman, 1974) allow for the possibility of the existence of pragmatic frames. Moreover, the proposal is not to cover the entire field of pragmatics, but only the study of linguistic expressions that have a conventionalized pragmatic use.

The authors begin their discussion with an analysis of tag questions, demonstrating that pragmatic meaning does not lie in each separate lexeme that is part of such expressions. Instead, it is the construction itself that performs the different pragmatic functions of tag questions, depending on the context in which they are used and fulfilling an interactive role between interlocutors (Czulo; Ziem; Torrent, 2020, p. 2). "Tag questions are multi-word expressions that evoke frames in a holistic rather than a compositional fashion, in that the building blocks of the expressions cannot be considered units carrying frame-semantic information on their own" (Czulo; Ziem; Torrent, 2020, p. 2).

The authors use, as a basis for their arguments, observations of the analysis of the results obtained by the Global FrameNet Shared Annotation Task. Annotation of the text of the TED talk "Do Schools Kill Creativity?", presented by Sir Ken Robinson (2006), was made using the translations (provided with the video) in Brazilian Portuguese, German and English, using Berkeley FrameNet 1.7. While undertaking this task, the annotators were allowed to include new lexical units to the dataset; however, they were not allowed to include new frames or to modify the existing ones. The first thirty sentences annotated in this way were, then, analyzed by an evaluation system (Czulo; Ziem; Torrent, 2020).

Comparing the annotations made for these three languages, Czulo, Ziem and Torrent (2020) demonstrate that some differences in the three annotation sets were caused by lack of an appropriate form to annotate pragmatic aspects of terms and expressions such as question tags and greetings. 'Good morning', for example, was not annotated in either Portuguese or English, as annotators recognized the pragmatic nature of the expression and did not have the means to indicate it. German annotators, however, chose to annotate the German term *Guten Morgen* ('Good morning'), as two separate lexical units – *Guten*, evoking the Desirability frame, and *Morgen*, evoking the Calendric\_unit frame – which caused the expression to lose the pragmatic meaning of a greeting. In fact, 'good morning' (and its equivalent expressions

in other languages) is a multi-word unit (like tag questions) that, to work as a greeting, has to meet certain conditions: the fact that the speaker is involved in an interaction with their interlocutor, the fact that it is morning, and the fact that it is the first time that the interlocutors meet that morning (Czulo; Ziem; Torrent, 2020, p. 3).

We consider it interesting to notice that the equivalent expression in Brazilian Portuguese, *bom dia*, means literally ‘good day’, as in (1):

(1) Brazilian Portuguese

Bom	dia.
Good	day
‘Good morning.’	

*Dia* (‘day’), in Portuguese, either refers to the whole period of twenty-four hours, or to the part of this period during which there is sunlight, similarly to the word ‘day’ in English. Nevertheless, in Portuguese, the use of the greeting *bom dia* is restricted to the morning period, exactly the way ‘good morning’ is used in English and *Guten Morgen* is used in German. This fact reinforces the pragmatic value of the expression *bom dia*.

Even though German annotators made the choice of annotating *Guten Morgen* separately during the Global FrameNet Shared Annotation Task, German FrameNet and Constructicon was the first FrameNet to start working on pragmatic frame definition (Ziem; Willich; Triesch, 2023). Taking the analysis and observations of Czulo, Ziem and Torrent (2020), as well as the work of German FrameNet and Constructicon as departing points, FrameNet Brasil has started the development of creation and annotation of pragmatic frames, which are now part of our database.

### 1.3 Interactive Gestures

As part of our work within ReINVenTA, it seemed logical to develop pragmatic frame annotation as a multimodal task. The first step, then, was to choose a specific topic of investigation. The topic chosen was the strategies used by communicators, during face-to-face conversation, to manage conversational turns. More specifically, we chose the gestures used by these communicators to indicate different choices such as keeping the turn, passing it, asking for it, and so forth. As Sacks, Schegloff and Jefferson (1974) and Bavelas (2022) observed, face-to-face conversation is the most basic form of communication. Nevertheless, research in turn management is far from abundant.

Janet Bavelas (Bavelas *et al.*, 1995; Bavelas, 2021, 2022) is probably the most prolific author in the study of gestures used by communicators for face-to-face conversation. Her research involves different aspects of this form of communication, and one of these aspects is the types of gestures used by communicators during conversation. She separates these gestures, which she calls “co-speech gestures” (Bavelas, 2022, p. 13), into two different categories: topic gestures and interactive gestures. Topic gestures are those used to illustrate the topic of conversation – for example, when a person indicates with a hand a height level while saying something like, “the bookshelf was this high”. Interactive gestures, however, refer to a person’s understander. These gestures are used to refer, for example, to something one’s communica-

tor mentioned previously. Bavelas *et al.* (1992, p. 469) give as an example a situation in which the utterer says, “and Customs is DEFINITELY is career oriented”. When the utterer mentioned “Customs”, they made a brief movement towards their understander, which the authors, through observation and analysis of the video, considered to mean something like “which YOU suggested” (Bavelas *et al.*, 1992, p. 471).

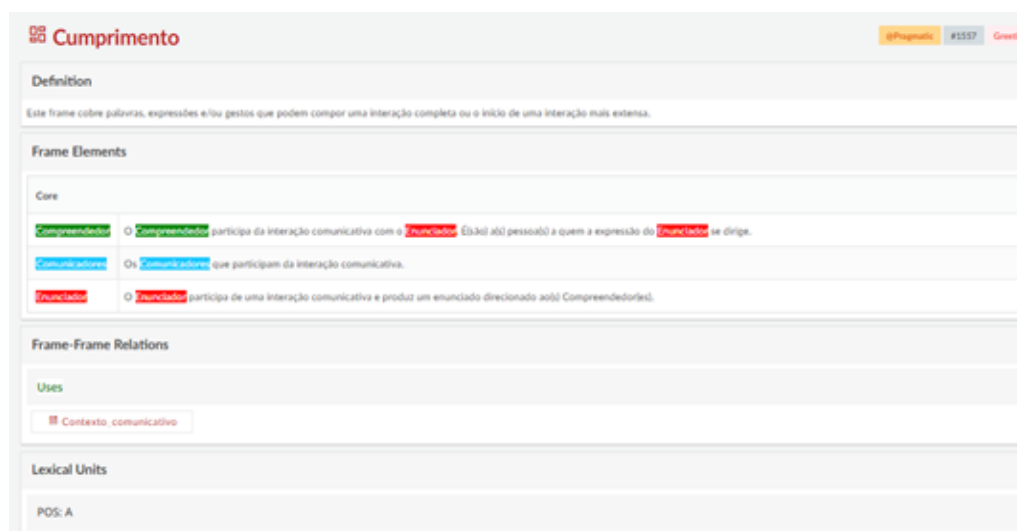
Interactive gestures are also used for the management of conversational turns (Bavelas *et al.*, 1995; Bavelas, 2021, 2022), proving to be an important tool that helps communicators indicate their intentions of keeping the conversational turn, passing it to a chosen person, asking permission to take it and so forth, while the topic of conversation continues to flow. When an utterer “passes” the turn to another person, for instance, a typical turn-passing gesture is one that is similar to the gesture of passing a real object to someone. This gesture may be used when only two people are participating in conversation, but it can be especially useful when three or more people are involved. It is important to note that the message conveyed through these gestures is not mentioned in the conversation. The gestures work as a second layer of communication, which makes face-to-face conversations multimodal.

The following sections will present how our dataset was developed, a discussion of our observations and our conclusions.

## 2 Dataset Development

The first step in our work was to develop the annotation of pragmatic frames within Webtool. The first two pragmatic frames to be inserted into Webtool were *Contexto comunicativo* (‘Communicative\_context’) and *Cumprimento* (‘Greeting’), which is demonstrated by Figure 1.

Figure 1- Greeting



Source: FrameNet Brasil Lab (c2014-2025d).

The decision to start with these frames was based on the conclusions reached by Czulo, Ziem and Torrent (2020) about expressions such as ‘good morning’ having pragmatic mea-

ning, as presented in section 1, and the fact that human conceptualization also makes use of the communicative context frame together with the greeting frame. The relation between these two frames is indicated in the “Frame-to-Frame Relations” area, which can be observed in Figure 1 above. As indicated, the frame *Cumprimento* (“Greeting”) uses the frame *Contexto comunicativo* (“Communicative\_context”).

The frame elements (FEs from now on) in the frame Greeting are the people involved in the communicative scene: *enunciador* (‘utterer’, the person who talks or gestures), *compreendedor* (‘comprehender’, the person to whom the communicative act is directed) and *comunicadores* (‘communicators’, encompassing utterer and comprehender together, as a unit). On the upper right-hand corner of the screen, the orange label indicates the type of frame it is – in this case, a pragmatic one.

As we had in mind the type of pragmatic frames we would be looking for in the videos, based on the paper by Czulo, Ziem and Torrent (2020) and on the research on interactive gestures presented in section 1, the next frames to be inserted into Webtool were: *Organization\_of\_conversation* (of which the next frames would be sub-frames); *Turn\_passing* (in which the utterer indicates the passage of conversational turn to the comprehender); *Turn\_taking* (in which a communicator takes the communicative turn); *Turn\_confirmation* (in which the comprehender indicates the utterer should keep the turn); and *Turn\_keeping* (in which the utterer indicates the wish to continue communicating). Figure 2 shows the frame *Organização\_da\_conversa* (‘Organization\_of\_conversation’). In its Frame-Frame Relations panel we can observe its subframes and the fact that it uses the frame *Contexto comunicativo* (‘Communicative\_context’). After these frames were ready for annotation, we proceeded to the next phase of our work.

Figure 2- *Organization\_of\_conversation*



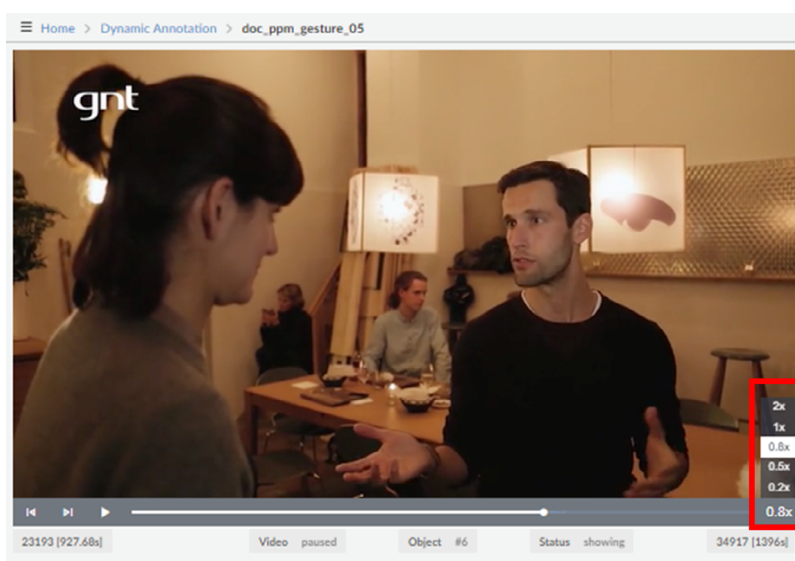
Source: FrameNet Brasil Lab (c2014-2025e).

Our corpus, *Pedro Pelo Mundo - Gestures*, was developed as a subdataset within Frame<sup>2</sup>. This subdataset is composed of the same episodes of the first season of the series *Pedro Pelo Mundo* that had already been fully annotated for lexical frames and the visual

objects related to them. Pedro Pelo Mundo - Gestures, however, was created within Webtool 4.0 separately, so that we could annotate each gesture that evokes a pragmatic frame.

The next step was to look for the interactive gestures that evoked each of the pragmatic frames previously mentioned, as they were used by the communicators during all the interviews. Webtool works in such a way that in order to look for the gestures we have the option of watching the video at normal speed, double the speed, or we can watch it in slower motion (0.8, 0.5 or 0.2 times in relation to the normal speed). The speed can be altered in the right-hand corner of the bar underneath the video panel on the screen, as shown by Figure 3, marked by a red box.

Figure 3: Video speed



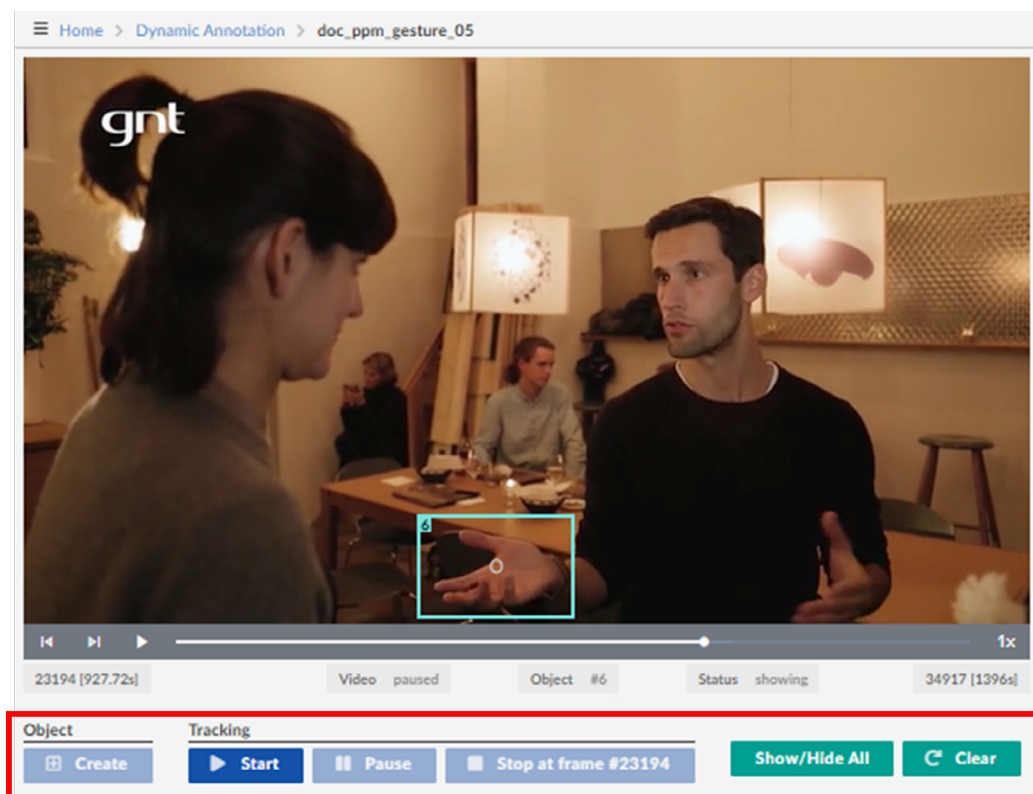
Source: FrameNet Brasil Lab (c2014-2025g).

We observed that the best speed for identifying the gestures is 0.8, which is almost normal speed, and yet slow enough for the gestures to be more easily recognized. Once the gesture is recognized, slower speeds help locate the precise moment of the beginning and the end of the movement. After pinpointing the moment the gesture begins, we start the process of creating a bounding box by pressing the button “create” on the left-hand side of the screen, under the video panel. After pressing this button, we can move the cursor and choose a place in the paused video image, click there and drag the cursor to create the bounding box. The box can then be moved, and its sides can be adjusted as necessary. Then we can press the button “start” and the bounding box will automatically follow the object as it moves within the video, until we press the button “pause” or “stop at frame #XXXX”.<sup>1</sup> We also have the option of using the button “move forward”, to the left of the “play” button, to move the video slowly, frame by frame, and make any adjustments to the bounding box that may be necessary. The buttons are shown in Figure 4, within a red box.

<sup>1</sup> It is important to notice that this button indicates the movie frame, which is a term used in the film industry and is unrelated to frames as in the frame theory. “Frame” as used in the film industry refers to each “picture” that, together with all the others, and in motion, compose a video.



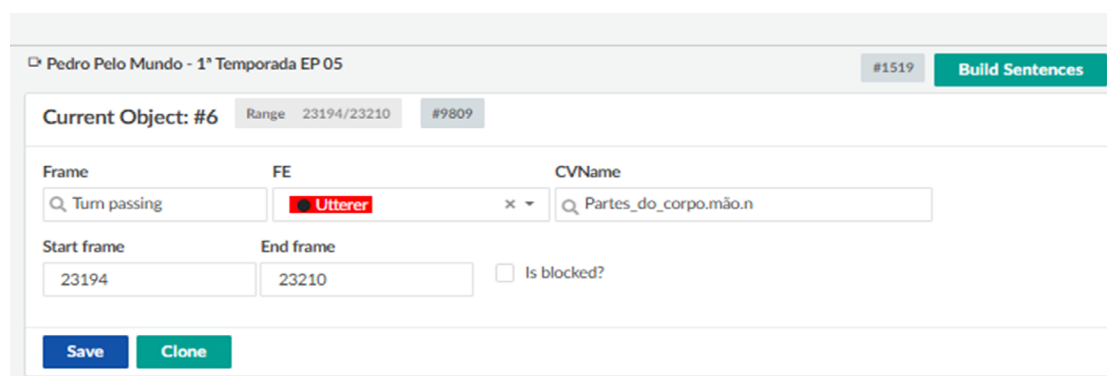
Figure 4: Bounding box creation



Source: FrameNet Brasil Lab (c2014-2025b).

Once the bounding box is created, we proceed to the top right-hand side of the screen, as depicted in Figure 5, to manually insert the chosen frame, the FE and the CVName (Computer-vision Name). In the case of gestures, the CVNames are the lexical identification of the visual object marked within the bounding box. In our example in Figure 4, it is the utterer's hand, marked by the bounding box number 6.

Figure 5: Completing information



Source: FrameNet Brasil Lab (c2014-2025c).

Webtool 4.0 can be seen in Brazilian Portuguese or in English. In the example shown by Figure 5, in English, the chosen frame is Turn\_passing; the FE is 'utterer', as it is the utterer's hand that is marked by the bounding box; and the LU is *mão* ('hand'), within the frame *Partes\_do\_corpo* ('Body\_parts'). The start frame and the end frame numbers are provided automatically; however, they can be manually adjusted if necessary. Once these options are chosen, we press the button "save" and the object appears within a panel in the bottom right quadrant of the screen. We can see the individual object more clearly in Figure 6:

Figure 6: Annotated object information

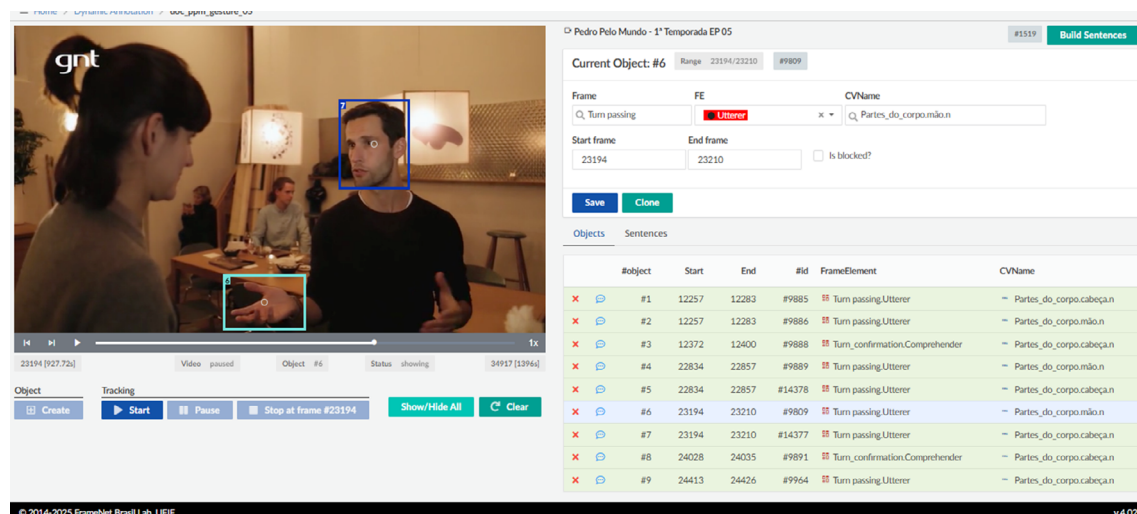
✕	🗨	#5	22834	22857	#14378	Turn passing.Utterer	Partes_do_corpo.cabeça.n
✕	🗨	#6	23194	23210	#9809	Turn passing.Utterer	Partes_do_corpo.mão.n
✕	🗨	#7	23194	23210	#14377	Turn passing.Utterer	Partes_do_corpo.cabeça.n

Source: FrameNet Brasil Lab (c2014-2025a).

This panel contains such information as the number of the object within the episode (in this case, number 6), the start and end movie frames of the bounding box (23194 and 23210, respectively), the object number in relation to the whole dataset (object number 9809), the frame (Turn\_passing), the FE (Utterer) and the CVName (*mão*, 'hand', which is part of the frame Body\_parts).

It is important to mention that turn passing gestures may involve not only the utterer's hand or hands, but also the person's head. Sometimes it is only with the head that the utterer indicates passage of turn. In the example depicted in Figure 4, Pedro, the utterer, used his right hand and his head to make the gesture. However, we can only annotate each body part separately. In this case, his hand is object 6, and his head is object 7. By clicking on the button "show/hide all", we can make all bounding boxes visible at the same time. Figure 7 depicts the whole screen, with the video paused at a moment in which we can see the passing of turn, the two parts of the utterer's body involved in the gesture (head and hand), as well as all the other objects (parts of the body involved in interactive gestures) annotated in this whole episode. It also shows how the whole screen is visualized on Webtool 4.0.

Figure 7: Whole screen



Source: FrameNet Brasil Lab (c2014-2025h).

### 3 Discussion

This corpus was selected for its advantages, as presented in section 1; however, it has some limitations that should be considered. First of all, the series is not composed only of interviews. There are long sequences in which Pedro Andrade is showing the viewers different places, trying different types of food, and so forth, as mentioned before. The interviews are only one of the multiple parts of each episode. As a result, of the 23 minutes that compose one episode, there may be less than 10 minutes of actual interview time. Secondly, because it is a television program, the camera angles chosen are not always the best for simultaneously observing both communicators. Sometimes only one of them is visible on the screen, and it is not possible for us to observe all the expected gestures or the responses to them which would be necessary to confirm these gestures as interactive ones. At other times, the camera is focused on the communicator's face and the gestures they might be making are hidden from the viewer. Thirdly, during the interviews, the video sequence at times shows not the communicators themselves but objects, people, places or scenes that are related to the subject being discussed.

Although this presentation format is a great resource for multimodal annotation of frames in general (Belcavello *et al.*, 2024), it was impossible for us to observe what gestures the communicators were using at those times. For all these reasons, in each episode there are only a few minutes of fully observable interactions between communicators. Nevertheless, the corpus is significant enough in a way that we were able not only to observe and confirm the existence of gestures that had been previously recorded by other

researchers, but also to make our own observations which were added to the body of knowledge about interactive gestures.

One of these observations was the fact that the gestures are not always directed to the utterer's comprehender. It is our understanding that the context in which interactions were recorded in previous work (interviews and conversations filmed in laboratories) did not allow for the observation of what communicators do in other settings. For example, some of the interviews conducted by Pedro Andrade happened while he and his communicator were walking on the street. In these situations, while both communicators were walking side by side, seemingly paying partial attention to where they were walking, some of the turn passing gestures were not pointed at the utterer's communicator. Instead, these gestures were made with the utterer's hand extended to his front and still they could be easily recognizable. Figure 8 shows one of these interactions.

Figure 8 : Turn passing gesture



Source: FrameNet Brasil Lab (c2014-2025f).

## 4 Conclusions

This paper aimed to present pragmatic frames as they are now being developed and annotated by FrameNet Brasil. Czulo, Ziem and Torrent (2020) had already demonstrated the opportunities for annotating pragmatic meaning and the advantages that it would entail. This paper has shown that FrameNet Brasil, through its Webtool, has been able to produce

such annotations and confirmed the views of Czulo, Ziem and Torrent (2020). It has also shown that pragmatic frames may be evoked lexically or visually.

Moreover, our research has demonstrated that, through annotation of pragmatic frames, we were able to confirm that the management of conversational turns is multimodal. This fact proves that the annotation of these frames is a useful tool for research in linguistics as well as other areas involved with language, such as, but not limited to, conversation analysis and Natural Language Processing.

Further developments of our work will contribute to the expansion of FrameNet multimodal analysis to include turn organization pragmatic frames, which is fundamental for computational representation of different multimodal text genres, such as interviews and debates. Also, further developments of our work will contribute to pragmatic investigations, through the introduction of a computational implemented tool for analysis of pragmatic phenomena in a multimodal context.

Such contributions will increase both the analytical and theoretical impact of the multimodal turn in FrameNet development. From the analytical side, including the annotation of pragmatic frames evoked by gestures opens up the possibility to include in the resulting datasets a myriad of meaning making interrelations between communicative modes yet to be fully explored by the FrameNet Brasil model.

From the theoretical side, they bring an additional opportunity to extend the application of the theory of Frame Semantics to a domain that, despite being included in Fillmore's original proposals for a Semantics of Understanding (Fillmore, 1985), is yet to be fully explored by practitioners in the field.

## Acknowledgments

This paper is a project within ReINVenTA – Research and Innovation Network for Vision and Text Analysis of Multimodal Objects. ReINVenTA is funded by FAPEMIG grant RED 00106/21, and CNPq grants 408269/2021-9 and 420945/2022-9. Helen de Andrade Abreu's research is funded by FAPEMIG grant RED-00106-21.

## References

- BAVELAS, J. B. *Face-to-Face Dialogue: Theory, Research, and Applications*. Oxford: Oxford University Press, 2022.
- BAVELAS, J. B.; CHOVIL, N.; COATES, L.; ROE, L. Gestures Specialized for Dialogue. *Personality and Social Psychology Bulletin*, v. 21, n. 4, p. 394-405, 1995.
- BAVELAS, J. B.; CHOVIL, N.; LAWRIE, D. A.; WADE, A. Interactive Gestures. *Discourse Processes*, v. 15, n. 4, p. 469-489, 1992.
- BAVELAS, J. Pragmatics of Human Communication 50 Years Later. *Journal of Systemic Therapies*, v. 40, n. 2, p. 3-25, 2021. DOI: 10.1521/jsyt.2021.40.2.3.

- BELCAVELLO, F. *et al.* Frame<sup>2</sup>: A FrameNet-Based Multimodal Dataset for Tackling Text-image Interactions in Video. *In: JOINT INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, LANGUAGE RESOURCES AND EVALUATION (LREC-COLING 2024)*, 2024, Torino. *Proceedings [...]*. Torino: European Language Resources Association (ELRA)/ ICCL, 2024. p. 7429-7437.
- BELCAVELLO, F.; VIRIDIANO, M.; COSTA, A. D. da; MATOS, E. E.; TORRENT, T. T. Frame-Based Annotation of Multimodal Corpora: Tracking (A)Synchronies in Meaning Construction. *In: LREC INTERNATIONAL FRAMENET WORKSHOP*, 2020, Marseille. *Proceedings [...]*. Marseille: ELRA, 2020. p. 23-30.
- BELCAVELLO, F.; VIRIDIANO, M.; MATOS, E.; TORRENT, T. T. Charon: A FrameNet Annotation Tool for Multimodal Corpora. *In: LINGUISTIC ANNOTATION WORKSHOP (LAW)*, 16., 2022, Marseille. *Proceedings [...]*. Workshop presented at LREC 2022. Marseille: ELRA, 2022. p. 91-96.
- BOAS, H. C.; ZIEM, A. Constructing a Constructicon for German. *In: LYNGBELT, B.; BORIN, L.; OHARA, K.; TORRENT, T. (ed.). Constructicography: Constructicon Development Across Languages*. Amsterdam: John Benjamins Publishing Company, 2018. p. 183-228.
- CZULO, O.; TORRENT, T. T.; MATOS, E.; COSTA, A. D. da; KAR, D. Designing a Frame-Semantic Machine Translation Evaluation Metric. *In: WORKSHOP ON HUMAN-INFORMED TRANSLATION AND INTERPRETING TECHNOLOGY (HiT-IT)*, 2., 2019, Varna. *Proceedings [...]*. Varna: Incoma, 2019. p. 28-35.
- CZULO, O.; ZIEM, A.; TORRENT, T. T. Beyond Lexical Semantics: Notes on Pragmatic Frames. *In: LREC INTERNATIONAL FRAMENET WORKSHOP*, 2020, Marseille. *Proceedings [...]*. Marseille: ELRA, 2020. p. 1-7.
- DORNELAS, L.; GAMONAL, M. A.; PAGANO, A. S. Análise semântica de audiodescrição em curta-metragem: uma abordagem multimodal a partir da Semântica de Frames. *Domínios de Linguagem*, Uberlândia, v. 18, e1866, 2024.
- FILLMORE, C. J. An Alternative to Checklist Theories of Meaning. *In: COGEN, C. THOMPSON; H. THURNGOOD, G.; WHISTLER, K. (ed.). Proceedings of the Berkeley Linguistic Society*. Berkeley: Berkeley Linguistics Society, 1975. p. 123-131.
- FILLMORE, C. J. Frame Semantics. *In: LINGUISTIC SOCIETY OF KOREA (ed.). Linguistics in the Morning Calm*. Seoul: Hanshin Publishing, 1982. p. 111-137.
- FILLMORE, C. J. Frames and the Semantics of Understanding. *Quaderni di Semantica*, v. 6, p. 222-254, 1985.
- FILLMORE, C. J. The Merging of Frames. *In: FAVRETTI, R. R. (ed.). Frames, Corpora, and Knowledge Representation*. Bologna: Bononia University Press, 2008. p. 1-12.
- FILLMORE, C. J.; BAKER, C. A Frames Approach to Semantic Analysis. *In: HEINE, B. et al. (ed.). The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2009. p. 791-816.
- FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. Background to FrameNet. *International Journal of Lexicography*, v. 16, n. 3, p. 235-250, 2003a.
- FILLMORE, C. J.; LEE-GOLDMAN, R.; RHOMIEUX, R. A. FrameNet in Action: The Case of Attaching. *International Journal of Lexicography*, v. 16, n. 3, p. 297-332, 2003b.
- FRAMENET BRASIL LAB. *Annotated Object Information*. [S. l.], c2014-2025a. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1519>. Accessed on: 10 Jan. 2025.

- FRAMENET BRASIL LAB. *Bounding Box Creation*. [S. l.], c2014-2025b. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1519>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Completing Information*. [S. l.], c2014-2025c. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1519>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Greeting*. [S. l.], c2014-2025d. Available at: <https://webtool.frame.net.br/report/frame/1557>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Organization\_of\_conversation*. [S. l.], c2014-2025e. Available at: <https://webtool.frame.net.br/report/frame/1561>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Turn Passing Gesture*. [S. l.], c2014-2025f. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1516>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Video Speed*. [S. l.], c2014-2025g. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1515>. Accessed on: 10 Jan. 2025.
- FRAMENET BRASIL LAB. *Whole Screen*. [S. l.], c2014-2025h. Available at: <https://webtool.frame.net.br/annotation/dynamicMode/1519>. Accessed on: 10 Jan. 2025.
- GRUZITIS, N.; NESPORE-BERZKALNE, G.; SAULITE, B. Creation of Latvian FrameNet Based on Universal Dependencies. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 11., 2018, Miyazaki. *Proceedings* [...]. Edited by T. T. Torrent, L. Borin and C. F. Baker. Miyazaki: ELRA, 2018. p. 23-27.
- HAHM, Y.; NOH, Y.; HAN, J. Y.; OH, T. H.; CHOE, H.; KIM, H.; CHOI, K. S. Crowdsourcing in the Development of a Multilingual FrameNet: A Case Study of Korean FrameNet. In: LANGUAGE RESOURCES AND EVALUATION CONFERENCE (LREC), 12., 2020, Marseille. *Proceedings* [...]. Marseille: ELRA, 2020. p. 236-244.
- LUZ, A. C. L.; BRAZ, G.; RUIZ, L. P.; PINTO, M. C.; BELCAVELLO, F.; SIGILIANO, N. S.; TORRENT, T. T. Anotação do Dataset Multimodal da ReINVenTA. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (STIL), 14., 2023, Porto Alegre. *Anais* [...]. Porto Alegre: SBC, 2023. p. 360-364.
- OHARA, K.; KAWAHARA, D.; SEKINE, S.; INUI, K. Linking Japanese FrameNet with Kyoto University Case Frames Using Crowdsourcing. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 11., 2018, Miyazaki. *Proceedings* [...]. Edited by T. T. Torrent, L. Borin and C. F. Baker. Miyazaki: ELRA, 2018. p. 56-61.
- SACKS, H.; SCHEGLOFF, E.; JEFFERSON, G. A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language*, v. 50, n. 4, p. 696-735, 1974.
- SUBIRATS-RÜGGERBERG, C.; PETRUCK, M. R. Surprise: Spanish FrameNet! In: WORKSHOP ON FRAME SEMANTICS, 2003, Prague. Workshop presented at the XVII International Congress of Linguists (CIL). *Proceedings* [...]. Edited by E. Hajivcová, A. Kotévsocová, and J. Mirovsky. Prague: Matfyzpress, 2003.
- TORRENT, T. T. *et al.* A Flexible Tool for a Qualia-Enriched FrameNet: The FrameNet Brasil WebTool. *Lang Resources & Evaluation*, 2024. DOI: <https://doi.org/10.1007/s10579-023-09714-8>.
- TORRENT, T. T.; ELLSWORTH, M. Behind the Labels: Criteria for Defining Analytical Categories in FrameNet Brasil. *Veredas*, v. 17, n. 1, p. 44-65, 2013.

VIRIDIANO, M. *et al.* Framed Mult30K: A Frame-Based Multimodal-Multilingual Dataset. *In: JOINT INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, LANGUAGE RESOURCES AND EVALUATION (LREC-COLING)*, 2024, Torino. *Proceedings* [...]. Torino: ELRA/ICCL, 2024. p. 7438-7449.

VIRIDIANO, M.; TORRENT, T. T.; CZULO, O.; LORENZI, A.; MATOS, E.; BELCAVELLO, F. The Case for Perspective in Multimodal Datasets. *In: WORKSHOP ON PERSPECTIVIST APPROACHES TO NLP*, 1., 2022, Marseille. *Proceedings* [...]. Workshop presented at LREC 2022. Marseille: ELRA, 2022. p. 108-116.

YOU, L.; LIU, K. Building Chinese FrameNet Database. *In: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING AND KNOWLEDGE ENGINEERING*, 2005, Wuhan. *Proceedings* [...]. Wuhan: IEEE, 2005. p. 301-306.

ZIEM, A.; WILlich, A.; TRIESCH, S: Expanding the German FrameNet: Pragmatic Frames Across Lexicon and Grammar. *In: INTERNATIONAL COGNITIVE LINGUISTICS CONFERENCE*, 16., 2023, Düsseldorf. *Proceedings* [...]. Düsseldorf: ICLC, 2023. Available at: [https://iclc16.github.io/abstracts/ICLC16\\_BoA.pdf](https://iclc16.github.io/abstracts/ICLC16_BoA.pdf). Accessed on: 13 Feb. 2025.