Visual Representations of Temporal Relations between Events and Time Expressions in News Stories

Evelin Amorim[1,2], António Leal[4,3], Nana Yu[3], Purificação Silvano[3,1], Alípio Jorge[2,1]
[1]INESC TEC / Porto, Portugal
[2]University of Porto, FCUP / Porto, Portugal
[3]University of Porto, CLUP / Porto, Portugal
[4]University of Macau / Macau, China
evelin.f.amorim@inesctec.pt, antonioleal@um.edu.mo,
robertananayu@hotmail.com, msilvano@letras.up.pt, amjorge@fc.up.pt

Abstract

High-quality annotation is essential for the effective predictions of machine learning models. When annotations are dense, achieving accurate human labeling can be challenging since the most used annotation tools present an overloaded visualization of labels. Thus, we present Vitra (Visualizer of temporal relation annotations), a tool designed for viewing annotations made in corpora, specifically focusing on the temporal relations between events and temporal expressions. This tool aims to fill a gap in the available resources for this purpose. Our focus is on narrative text, which is a rich source for these types of elements. Vitra was developed to increase the human capacity for detecting annotation errors and uncover relations between narrative components or issues about the annotation scheme. To show how this can be done, we present an analysis of a subset of the Text2Story Lusa corpus, a dataset of Portuguese news stories. Such analysis focuses on the linguistic properties of the events and temporal expressions that occur in the annotated texts, in particular, of short news. We highlight that annotation is an iterative process that involves multiple rounds of revision, and our tool facilitates this process by helping users detect inconsistencies and improve the annotation scheme, thus offering added value to the community.

1 Introduction

Events and time expressions are essential elements in news stories. They both contribute to the narrative structure by linking actions, facts, and developments to specific points in time, helping readers grasp the sequence, causality, and context of events. Additionally, the relationship between events and time expressions plays a crucial role in establishing a timeline of occurrences.

Annotating events and time expressions and their relations is a well-known task in NLP (Doddington et al., 2004; Cassidy et al., 2014; Caselli and Vossen, 2017; Wang et al., 2022; Olsen et al., 2024). Therefore, several formats have been proposed to visually present this information to a lay target audience (Chen et al., 2023), like timelines (Nguyen et al., 2016; dos Santos Fernandes, 2023), infographics (Chen et al., 2019), and data comic (Zhao et al., 2021). However, these representations often lack the precision and focus needed for expert audiences, whose primary goal is to analyze data and inspect annotations in detail. Some visual representations, such as the Message Sequence Chart (MSC), have been used for this purpose (Hingmire et al., 2020; Amorim et al., 2021). However, these approaches do not specifically address the representation of events combined with time expressions in a visually intuitive manner, which would help in identifying annotation mistakes. This gap highlights the need for specialized visual tools that cater to the demands of expert users in tasks like annotation analysis and narrative inspection.

To address this gap, this work explores the following research questions:

RQ1 Which insights can be derived from visualizing the events and temporal expressions, and their temporal relationships?

RQ2 How effective is isolating time expressions and their related events, and arranging them on a timeline, for facilitating annotation inspection by expert audiences?

The first research question is whether the proposed visualization is suitable for representing temporal relations between events and time expressions in narratives. The usefulness of the visualization will be evaluated according to its capability to detect annotation errors and uncover relations between key narrative components or issues about the annotation scheme. Thus, we will characterize the Text2Story Lusa corpus, which is a narrative dataset manually annotated with Portuguese news

Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025), pages 250–263

stories (Silvano et al., 2023b; Nunes et al., 2024) using Vitra (Visualizer of temporal relation annotations), our proposed visualization. The corpus was annotated using the Brat annotation tool (Stenetorp et al., 2012), and then, for this investigation, we converted the Brat standoff file format to the JSON format, which is a more general format. To this conversion, we employ the text2story package (Amorim et al., 2024), which also offers conversions from other types of annotation files. The second research question concerns its usability for the annotator. We elaborated on a questionnaire with six claims that intend to evaluate the overall design of the Vitra tool. Linguist experts with annotation experience were called to answer the proposed questionnaire.

By answering these research questions, the contributions of this work are the following:

1. Vitra, an open tool for visualization of events and time expressions to aid the iterative annotation process[1];

2. An analysis of how an effective visualization tool can aid in detecting annotation mistakes and improving multi-layer annotation schema quality;

3. A deep characterization of temporal relations between time expressions and events in a narrative dataset.

Our goal is to advance research on multi-layer dataset annotation by improving annotation quality through the integration of a visualization tool. Additionally, we aim to provide insights into the design and application of the proposed annotation scheme.

2 Related Work

The two fundamental concepts of our work are temporal relations between events and time expressions and their visualizations. Both have been extensively studied in recent years. Therefore, we divide this section into two parts to discuss research related to each topic. The first part discusses similar works about temporal relations. The second part is the visualization of temporal information.

[1]The code is available in https://github.com/evelinamorim/sentencevisual; a demo can be found in https://nabu.dcc.fc.up.pt/annotationinspector/

2.1 Temporal Relations

Several linguists, including Bell (1997) and Schokkenbroek (1999), argue that the narratives conveyed in news articles are inherently dependent on the temporal arrangement of events. The reconstruction of a narrative's timeline can be achieved through implicit temporal references, such as verb tense, or explicit temporal markers, namely time expressions (Filatova and Hovy, 2001). Time expressions are, in fact, essential for situating events within a temporal framework and determining the structural organization of a text. Moreover, they also play a crucial role in numerous downstream tasks in Natural Language Processing (NLP) and Information Retrieval (IR), such as timeline summarization, named entity recognition, temporal information retrieval, and question answering (Jatowt et al., 2022). Advancing these tasks, particularly the identification and extraction of time expressions (Lange et al., 2020; Sousa et al., 2023; Zhong and Cambria, 2023; Zhong et al., 2024), relies on the availability of annotated data and well-defined annotation schemes.

Various studies have proposed different annotation frameworks to represent not only the temporal information of events but also the characteristics of time expressions. One of the most significant contributions in this domain is the work of Pustejovsky et al. (2003), who introduced TimeML as an annotation specification designed to systematically encode time expressions, events, and their temporal relations in natural language texts (ISO-24617-1, 2012). In this framework, time expressions (labeled TIMEX3) are categorized into dates, times, durations, and sets, while the morphosyntactic and semantic properties of events (EVENT) are captured through attributes related to class, type, tense, part of speech, among others, and the temporal relations (TLINK) are represented by values like *before, after, during*, among others.

TimeML provides a robust methodology for encoding temporal information across various linguistic contexts, facilitating its application beyond English to languages such as Italian, Korean, Chinese, French, and Portuguese (Costa and Branco, 2012; Bittar, 2009; Silvano et al., 2024). Language-specific adaptations, including It-TimeML (Caselli et al., 2011) and KTimeML (Im et al., 2009), have been developed to address language-specific phenomena not adequately covered by ISO-TimeML. Based on these annotation frameworks, several an-

notated datasets have been created, encompassing a wide range of textual genres. Some examples include TimeBankPT—an adaptation of the English TimeBank — a Portuguese Annotated Dataset of news stories (Silvano et al., 2023b), and i2b2 (Sun et al., 2013), a dataset annotated with events and time expressions extracted from clinical narratives. Additionally, the NewsReader MEANTIME (Multilingual Event ANd TIME) corpus is a semantically annotated resource consisting of 480 news articles in English, Italian, Spanish, and Dutch (Minard et al., 2016). The dissemination of these annotated datasets has been facilitated through shared tasks, such as the TempEval series (Pustejovsky and Verhagen, 2009) and Clinical TempEval (Bethard et al., 2016), which target the extraction of three key tags: TIMEX3, EVENT, and TLINK.

Despite the availability of datasets containing annotated time expressions and their corresponding temporal relations with events, visualizing this information can often be challenging. In this regard, visualization tools play a critical role in facilitating linguistic analysis and validating annotation quality, thereby enhancing the interpretability and usability of annotated temporal data. Regarding the temporal analysis of news stories using visualization, Silvano et al. (2023a) and Silvano et al. (2024) analyzed temporal relations between events using a visualization called Bubble visualization. Our work aims to study the temporal relations of temporal expressions and their connected events, which means an analysis of a different annotation layer.

2.2 Visualizations of Temporal Information

Arranging temporal information in a visual timeline is a natural form of organizing events, time expression, and participants. For example, Gonçalves et al. (2023) presents a platform that provides a user's query search for related news stories in a database. The information is presented in a timeline of news stories, in temporal groups, among other representations for non-temporal information. Ye et al. (2024) uses the GPT model to annotate text, and then the main events are presented in a timeline. The authors tested the proposed approach using two use cases, one with a fictional book and another with a movie script. Most users who experimented with it found the tool easy to use and helpful in understanding the narratives.

Tang et al. (2018) proposed iStoryline, a tool that was built to generate hand-drawn narrative storylines. The input is a structured file with the entities and their relations in a time order. Then, a timeline of the story is built in a hand-drawn style. The authors also based the tool on extensive research of the relevant visual elements that design experts commonly employ when creating timelines of stories. Tang et al. (2020) also proposed a timeline tool, PlotThread, which generates timelines of stories and enhances them through reinforcement learning. In this platform, the user defines a storyline, and then an AI agent proposes alternatives to the user's storyline. Consequently, the user can improve the visualization. In the proposed visualization, the timelines of the participants can be inspected along with some remarkable events in which they participated. Wang et al. (2024) proposed another timeline visualization called E^2Storyline that presents entities and their relations using a novel matrix color system designed to convey relationships between entities in narratives. The authors tested the visualization with human users who reported easily identifying information from stories and understanding the relations between entities. None of these tools, however, focuses on the analysis of annotation and linguistic patterns. Usually, their goal is to improve the experience of narrative understanding for a lay user or, at most, provide a high-level analysis of narrative patterns for an expert.

Lai (2023), differently, focused on a deep analysis of annotations. The author proposed an R package to process annotated data from Rezonator, an annotation tool for discourse and grammar, and conversation analysis, among others. The package builds cliques of causal structures, Gants charts, co-reference chains, and many more visual devices to allow comparisons between participants in a dialog. Our visualization, nonetheless, is designed to portray the relations of temporal information and their connected events. This type of annotation can occur in different domains of texts, and as far as we know, this type of tool has not yet been proposed. Thus, our tool intends to fill this gap.

3 Methodology

Our methodology comprises two main steps that we detail below: data analysis of a subset from a Portuguese news stories dataset and the visual tool.

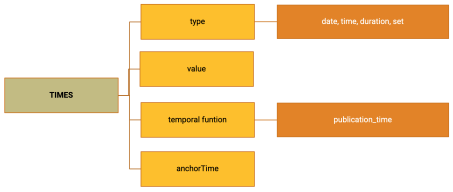


Figure 1: Attributes of the tag Time

Figure 2: Attributes of the tag Events

3.1 Dataset and annotation

The corpus analyzed in this study has 67 news articles in European Portuguese, predominantly published between October and December 2020, sourced from a Portuguese news stories dataset (Silvano et al., 2023b). The articles were selected based on their narrative nature and a word count ranging from 100 to 200 words. The dataset covers diverse topics, including accidents, homicides, and robberies. Annotation was performed using the Brat Rapid Annotation Tool (Brat) (Stenetorp et al., 2012), adhering to the annotation scheme developed by Silvano et al. (2021) and Leal et al. (2022).

The annotation scheme used in the employed dataset integrates four levels of the ISO-24617 standard: the temporal level (ISO-24617-1, 2012), the referential level (ISO-24617-9, 2019), the spatial level (ISO-24617-7, 2020), and the semantic roles level (ISO-24617-4, 2014). The scheme is structured into two primary components: (1) the entity structure, encompassing labels for events, temporal expressions, participants, and spatial elements, and (2) the link structure, representing relationships such as temporal, objectal, spatial, and semantic role links. The annotation scheme has demonstrated coherence and interoperability through testing on the same dataset, yielding favorable results (Silvano et al., 2023a, 2024).

This study specifically focuses on temporal annotation, emphasizing the labels and attributes associated with the entity structures *Time* (Figure 1) and *Event* (Figure 2). These labels are utilized to identify and characterize temporal expressions and events. Additionally, the analysis incorporates the *Temporal Link* structure, which captures relationships among events, between events and temporal expressions, and among temporal expressions. Temporal links include attributes such as *Before*, *After*, *Includes*, *Is_included*, *During*, *Simultaneous*, *Identity*, *Begins*, *Ends*, *Begun_by*, and *Ended_by*.

The dataset was annotated by a PhD student in linguistics who was trained in the Brat annotation tool and the annotation scheme guidelines under the supervision of a senior linguist researchers. The annotation process followed a structured sequence of steps: (1) Temporal expressions, along with their corresponding attributes and values, were annotated across all news items; (2) Events associated with these temporal expressions were identified and annotated, including their attributes and values; (3) Temporal relationships between each event and its corresponding temporal expression were established, with directionality specified from the event to the temporal expression; (4) Temporal relationships between all temporal expressions were annotated, with directionality defined from the last temporal expression in the linear discourse order to the preceding one. The PhD student and the senior Linguistics researcher conducted multiple consensus meetings following the training phase to ensure the reliability of the annotations. These meetings aimed to ensure that the annotation complied with the manual as the student progressed through the news items. In cases where there were doubts, solutions were found that were based on linguistic theory. After this annotation phase, a second senior Linguistics researcher knowledgeable about the annotation procedures checked and validated the results.

3.2 Visualization

The visualization methodology was designed with two main objectives: (1) ensuring that narrative components — events, temporal expressions, and their relations — are easily identifiable by experts, and (2) structuring the information to facilitate the recognition of annotation mistakes and specific patterns.

To develop the Vitra tool, the team collaborated with linguists to understand the requirements for temporal structure annotation. Initially, we adopted a design similar to Brat (Stenetorp et al., 2012),

where labeled elements appear within the raw text, highlighted by a color-coding system with relational links. However, this approach did not meet the linguists' needs, as it replicated the existing Brat interface, which does not isolate the relevant information and does not offer additional benefits. Consequently, we explored an alternative approach: isolating key information (time expressions, events, and relations) from the raw text using manual annotation. This separation improved visualization, aiding pattern identification. Given the central role of time expressions in this research, presenting them along a timeline was a natural choice. Events associated with time expressions were positioned to the left of the timeline, as they typically involve one or two instances at most.

The Figure 3 shows the final format after two more rounds of refinement with three linguist experts. Vitra was developed using programming languages such as Python, D3.js (Javascript), and the markup language HTML. The instructions are on the left side of the browser since it is the usual place for menus or referential information on a website. The sentences are separated in white blocks, thus, it is possible to highlight the current sentence under analysis by a human annotator. This functionality is activated after the human inspector clicks on the time expression he/she wants to analyze, and the corresponding sentence is highlighted. Different types of events and time expressions are assigned different borders and colors, as the instruction panel explains.

4 Results and Discussion

Our results are divided into data characterization regarding temporal information and the assessment of the visual tool proposed in this work.

4.1 Data characterization

The corpus contains an average of 175.97 tokens and 5.35 sentences per news article (cf. Table 1[2]).

	Tokens	Sentence
Avg.	175.97 ± 37.82	5.35 ± 1.29
Max.	239	9
Min.	82	3
Total	11,966	364

Table 1: Tokens and Sentences per News story

[2]We use the model pt_core_news_lg from the spacy library to tokenize the texts.

Regarding temporal expressions, the analysis of the attribute *Type* reveals that the most frequent temporal expressions correspond to *Date* (226), *Time* (43), and *Duration* (16), as shown in Table 2. The predominance of temporal expressions such as *Date* and *Time* is closely linked to the nature of the text analyzed. This type of text generally revolves around answering the central questions: 'Who?', 'What?', 'Where?', and, most importantly for our analysis, 'When?'. These results align with expectations given the analyzed text, which consists of brief news reports covering one or a few related events. Consequently, the temporal information is relatively straightforward, as temporal expressions typically indicate the relevant time interval related to the described situations. This is primarily achieved using *Date* expressions, which specify the day of the events, while *Time* expressions are used to a lesser extent to denote parts of the day. Example 4.1 illustrates this type of occurrence.

Example 4.1 *Um homem [. . .] morreu* ***hoje*** *na sequência do despiste do ciclomotor que conduzia [. . .] Os bombeiros receberam às* ***19:08*** *o alerta para o acidente (Lusa 40)*

A man died today after the moped he was driving skidded off the road. Firefighters received the alert for the accident at 7:08 pm.

The first temporal expression, categorized as *Date*, locates the event of "morrer"(to die) within a specific time interval that corresponds to a calendar day. The second temporal expression ("19:08"), classified as *Time*, provides additional information about the timing of the "receber" (to receive) event. This event is located within a narrower time interval, which is a subset of the timeframe indicated by the initial *Date* expression. These two temporal expressions are linked by a TLink described as *isIncluded*.

Example 4.2 illustrates the cases of *Duration*, which occur less frequently. The reason for the low occurrence of this type is that these expressions do not denote chronologically identifiable time intervals; that is, they do not answer the question 'When?'. As a result, they are not essential for understanding the primary information in this type of news.

Example 4.2 *Ali Bongo Ondimba esteve* ***vários meses*** *em convalescença (Lusa 346)*

Ali Bongo Ondimba spent several months convalescing

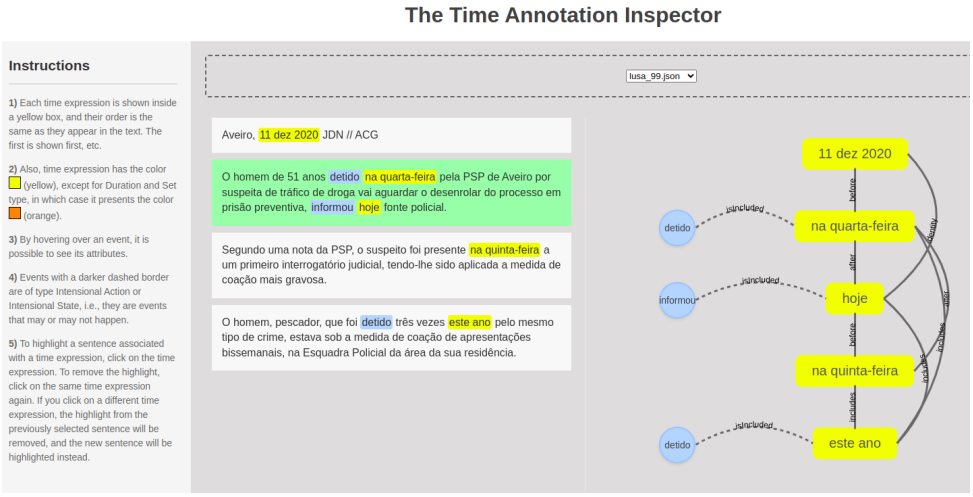


Figure 3: The Time Inspector Visualization

Concerning the temporal relations between events and temporal expressions, Figure 4a illustrates that, in most cases, the time interval denoted by the temporal expressions typically includes the time interval in which the situations are located (75%). The second most common relation observed is one where these two time intervals are simultaneous (13,2%). These findings support the earlier conclusion that, in these news stories, temporal expressions are usually dates (or expressions that function as dates) that locate the narrated situations within well-defined time intervals. Example 4.1 demonstrates a situation where the *isIncluded* relation is established between the expressions "morrer" (died) and "hoje" (today), while the *Simultaneous* relation occurs between "receber"(receive) and "19:08".

Other temporal relations between events and temporal expressions occur infrequently. Some expressions contribute to the temporal location of situations through the definition of the initial or final boundary of the relevant time interval (links *begunBy* (5.9%) and *endedBy* (1.5%)). Additionally, temporal expressions that have an aspectual role in measuring situations are also rare (only 4.4% of *During*) (cf. Figure 4a).

The analysis of temporal relations between temporal expressions reveals a significant variation in the results, as shown in Figure 4b. The most frequent relation between them is when the second temporal expression in the linear order of the discourse denotes a time interval that temporally precedes the time interval denoted by the first expression in the linear order of the discourse, accounting for 31.5%. This is followed by cases where both expressions refer to the same time interval (Identity), which makes up 28.1% of the results. In fourth place, we find the posteriority relation, where the temporal order matches the sequence of the expressions in the discourse, representing 13.4% of cases. Lastly, there are inclusion relations: 13% of cases involve the interval indicated by the second expression being included within the interval indicated by the first expression, while 14% of cases see the second expression's interval encompassing the first expression's interval.

The dominance of the *Before* relation likely stems from the structured format commonly used in news articles to convey information. Typically, a news article refers to events, which generally fall under the class of *Occurrence* (133 cases). These events are often described in sentences that identify the source of the information and include a *Reporting* event (43 cases) (Silvano et al., 2023a). In these instances, the *Occurrence* event is located before the *Reporting* event. Example 4.3 illustrates such cases.

Example 4.3 *Um homem [. . .] foi detido no concelho de Góis, [. . .] anunciou* ***hoje*** *a GNR. Segundo um comunicado [. . .], a detenção [..] ocorreu n****a terça-feira*** *(Lusa 43)*

A man was arrested in the municipality of Góis, the GNR announced today. According to a statement, the arrest took place on Tuesday

In this context, the first temporal expression, "hoje" (today), indicates the timing of the reporting event "anunciar" (to announce). The second temporal expression, "a terça-feira" (on Tuesday), specifies when the main event described in the news

(the arrest) took place. As a result, the second temporal expression occurs chronologically before the first expression.

Another reason for the recurrence of the *Before* relation is related to the structure of the news articles. Typically, a news text begins with a lead, which presents the central information necessary for understanding the story, followed by additional details that are less critical. This structure often includes references to previous events, which have earlier temporal contexts, helping to explain the causes of the main event. This is the case of Example 4.4, where the temporal expression "hoje" (today) locates the main event of the news - the arrest of the murder suspect-, indicating that it occurs after the murder event itself, which is situated in time by the expression "o domingo" (on Sunday).

Example 4.4 *A PJ deteve* ***hoje*** *o suspeito de matar um homem [. . .]* ***no domingo****, [. . .] em Cête. (Lusa, 5)*

The PJ arrested today the suspect of killing a man on Sunday in Cête.

The *Identity* relation is the second most frequent, as mentioned previously. This is because news articles often report on events that develop from the main event introduced in the lead, placing them within the same time interval. The fourth most frequent type of relation identified is temporal succession. This indicates that, in this genre of text, the chronological order of events does not always align with the linear narrative structure. Relations involving inclusion rank third and fifth. These are linked with expressions of type *Date*, which usually refer to time intervals represented by calendar terms, and expressions of type *Time*, which denote smaller segments of these intervals. For instance, in Example 4.1, the expression "19:00", categorized as *Time*, establishes an *isIncluded* relation with the expression "hoje" (today), which is classified as *Date*. These temporal relations are associated with a detailed breakdown of the information previously mentioned in the news lead.

Tables 2 and 3 in Appendix A detail all the attributes' statistics of time expressions and events, respectively.

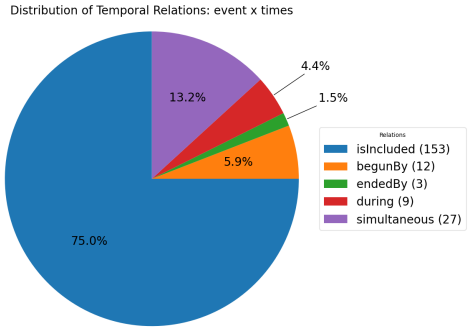
4.2 Visualization

Verifying the annotation using Vitra has led to several improvements both in the annotation process itself and in the overall framework. The proposed tool generates a much "cleaner" image, allowing the selection of only some elements that are part of the annotation's temporal level. This possibility of selection and simplification makes it much easier to identify (1) whether all the necessary relations have been made, i.e., whether important information is missing, and (2) whether the connections made are correct. This can be challenging in Brat, as the elements needing connection are often apart, and the sheer volume of annotations can create a "dense" visualization. Furthermore, the use of a color code enables us to quickly determine if temporal expressions have been annotated correctly with the appropriate attributes. In Vitra, the connecting lines provide an easy way to verify whether the relationships are correct and whether all temporal expressions and events are linked, allowing for the reconstruction of the event chronology. Examples of these advantages can be found in the scenarios presented in Appendix C.

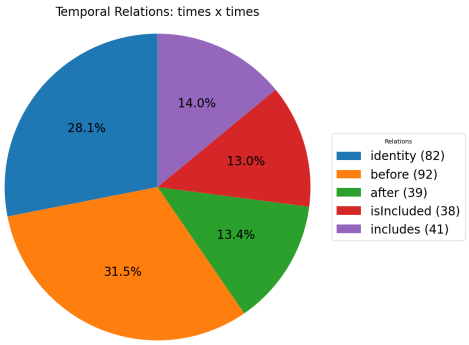
All in all, Vitra facilitates the comparison of a large number of annotated news articles, focusing on just one simplified annotation level. The visualization allows us to easily identify errors and inconsistencies in annotation across many news articles, thus contributing to improving the overall annotation quality of the entire corpus and the annotation manual itself.

To have a more independent assessment of the effectiveness of the visualization, we decided to develop a questionnaire with six claims related to the goals of this research, which are identifying annotation mistakes and recognizing patterns in the temporal structure of annotations. For each claim, we adopted a discrete Likert scale whose lowest number (1) in the score was associated with "Strongly Disagree", while the highest number (5) was associated with "Strongly Agree". Although this method has limitations, as pointed out by South et al. (2022), it is a standard quality evaluation method for visualizations. The list of all the claims is detailed in Appendix B. In addition to that, we left a text box for additional comments from the evaluators.

We invited three linguistics experts to complete a questionnaire designed to evaluate the proposed visualization. One of the experts had previously participated in the development of Vitra, therefore, we included two additional experts to eliminate any potential bias from the individual involved in the tool's development discussions. All three experts had similar profiles, were graduate students of lin-



(a) Temporal Relations between Events and Time Expressions



(b) Temporal Relations between Time Expressions

Figure 4: Comparison of temporal relations: (a) between events and time expressions, and (b) between time expressions.

guistics, already had experience with annotation tools like Brat, and knew the annotation schema used in the dataset that we employed in our experiments visualization. To compare with Brat visualization, four linguistics experts completed the same questionnaire designed to evaluate the Brat visual annotation tool. The only exception was the last question, which aimed to draw a comparison with Brat. The profiles of these four experts were similar to those of the linguists who answered the questionnaire for the evaluation of Vitra's visualization.

The first three claims of the questionnaire concern the interface, i.e., if the users were able to identify the events, times, and relations. Regarding this aspect, the users found Vitra's visualization mostly intuitive since the scores for the first three questions were between 4 and 5. The results were similar for Brat, where the scores also range from 4 to 5. The claims (4) and (5) of the survey were concerned with whether Vitra aids the process of identifying annotation mistakes and temporal patterns. Two of them scored 4 for the claim related to annotation errors (4), and one scored 5. Maybe this is related to the fact that Vitra's visualization does not present all the information of the annotation, for instance, the attributes of the events. For the Brat evaluation, claim 4, which is concerned with the identification of the errors, presented a great variation between the respondents. The scores for Brat in this issue were 2, 3, 4, and 5, showing that at least half of the linguistics experts think that Brat presents flaws in the inspection of annotations. In the assessment of Vitra in the discovery of temporal patterns (claim 5), two of them scored 5, while the other evaluator scored 4. Possibly, this is due to the arrangement of the temporal information, separated and combined with their relations, which aids in seeing all the temporal information as a whole. Regarding the Brat evaluation, three respondents scored 4 for the claim 5, while one scored 2. This suggests that the proposed visualization is competitive with respect to uncovering the temporal patterns with Brat.

We acknowledge that the number of respondents in our surveys evaluating Brat and Vitra, four and three participants, respectively, is not statistically significant. In most research, at least a sample size of 15 respondents is recommended; otherwise, the sample size is too small to draw a reliable conclusion (Sauro and Lewis, 2016). Consequently, the agreement scores derived from this sample size are also not statistically significant. However, our qualitative analysis, detailed with some examples in Appendix C, demonstrates the usefulness of this new visualization for analyzing the annotation schema and identifying errors.

5 Conclusion and Future Work

In this study, we investigate the application of visualization representation in the inspection of temporal relations involving events and time expressions. Narratives present dense information concerning events and time expressions. Hence, human annotators are presented with visually overloaded information in annotation tools when labeling narratives. In this investigation, we used a Portuguese dataset of news stories to answer the following research question.

RQ1) What insights can be derived from visualizing the events and temporal expressions and their temporal relationships

In our results, we showed that some unusual patterns in the temporal relations can be easily detected in the visualization of temporal expressions, events, and their relations. This is probably due to the nature of the proposed visualization, which sets aside temporal expressions, their related events, and their relations. In the Brat annotation tool, and usually in text annotation tools for documents, the raw text is presented along with all the relations and all entities. However, when facing a multilayer scheme, annotators can be challenged using a visualization like Brat since this is a more complex task.

Thus, answering the first research question, we conclude that unusual patterns in a multilayer annotation scheme are more salient in a visual representation that is devoted to the specific layers on which the focus of the investigation is. Some specific and relevant insights for our studied annotation scheme were observed, leading to adjustments in the annotation guidelines. A few use cases of insights are detailed in Appendix C.

RQ2) How effective is isolating time expressions and their related events, and arranging them on a timeline, for facilitating annotation inspection by expert audiences?

Isolating was beneficial for human annotators. In our questionnaire, the annotator experts positively assessed the identification of events, time expression, and their relations. Hence, all of them agreed that the proposed visualization facilitates the process of identifying errors or patterns.

We aim to advance the study of visual representations for human annotators as well as the quality of multilayer scheme annotations, which present complexities and challenges in development and assessment. In future work, we intend to add other annotation formats to Vitra, which can allow different types of arrangements, like the events in a timeline or even participants. Additionally, we plan to integrate Vitra into the Inception annotation tool (Klie et al., 2018), which is a more modern tool than BRAT. By doing this, we seek to stimulate human annotators to use our tool to aid the labeling process.

6 Limitations

The first limitation of our work is the small number of linguists who evaluate our tool, which could lead to a biased evaluation. The second limitation is that the tool still lacks interactive constraints. Currently, the annotator cannot correct annotation errors or move elements in the visual representation. These features could enhance the experience of the human annotator and broaden the functionalities of the representation. The third and last limitation that we can observe in this work is that the visualization is tied to the annotation scheme presented by Silvano et al. (2023a). However, we plan, as future work, to include other types of annotation schemes that include events and temporal expression as well.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020. DOI 10.54499/LA/P/0063/2020 | https://doi.org/10.54499/LA/P/0063/2020. The authors would also like to acknowledge the project StorySense, with reference 2022.09312.PTDC (DOI10.54499/2022.09312.PTDC).

References

Evelin Amorim, Ricardo Campos, Alípio Jorge, Pedro Mota, and Rúben Almeida. 2024. text2story: A python toolkit to extract and visualize story components of narrative text. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15761–15772.

Evelin Amorim, Alexandre Ribeiro, Inês Cantante, Alípio Jorge, Brenda Santana, Sérgio Nunes, Purificação Silvano, António Leal, and Ricardo Campos. 2021. Brat2viz: a tool and pipeline for visualizing narratives from annotated texts. In *Proceedings of Text2Story-Fourth Workshop on Narrative Extraction From Texts held in conjunction with the 43rd European Conference on Information Retrieval (ECIR 2021)*.

Allan Bell. 1997. *The Language of News Media*. Blackwell.

Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen. 2016. SemEval-2016 task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1052–1062, San Diego, California. Association for Computational Linguistics.

André Bittar. 2009. Annotation of events and temporal expressions in French texts. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 48–51, Suntec, Singapore. Association for Computational Linguistics.

Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 2011. Annotating events, temporal expressions and relations in Italian: the it-timeml experience for the ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 143–151, Portland, Oregon, USA. Association for Computational Linguistics.

Tommaso Caselli and Piek Vossen. 2017. The event storyline corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86.

Taylor Cassidy, Bill McDowell, Nathanael Chambers, and Steven Bethard. 2014. An annotation framework for dense event ordering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 501–506.

Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. 2023. How does automation shape the process of narrative visualization: A survey of tools. *IEEE Transactions on Visualization and Computer Graphics*.

Zhutian Chen, Yun Wang, Qianwen Wang, Yong Wang, and Huamin Qu. 2019. Towards automated infographic design: Deep learning-based auto-extraction of extensible timeline. *IEEE transactions on visualization and computer graphics*, 26(1):917–926.

Francisco Costa and António Branco. 2012. TimeBankPT: A TimeML annotated corpus of Portuguese. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3727–3734, Istanbul, Turkey. European Language Resources Association (ELRA).

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec*, volume 2, pages 837–840. Lisbon.

Catarina Justo dos Santos Fernandes. 2023. Visualizing news stories from annotated text. Master's thesis, Faculdade de Enegenharia da Universidade do Porto.

Elena Filatova and Eduard Hovy. 2001. Assigning timestamps to event-clauses. In *Proceedings of the Workshop on Temporal and Spatial Information Processing - Volume 13*, TASIP '01, USA. Association for Computational Linguistics.

Francisco Gonçalves, Ricardo Campos, and Alípio Jorge. 2023. Text2storyline: generating enriched storylines from text. In *European Conference on Information Retrieval*, pages 248–254. Springer.

Swapnil Hingmire, Nitin Ramrakhiyani, Avinash Kumar Singh, Sangameshwar Patil, Girish Palshikar, Pushpak Bhattacharyya, and Vasudeva Varma. 2020. Extracting message sequence charts from hindi narrative text. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pages 87–96.

Seohyun Im, Hyunjo You, Hayun Jang, Seungho Nam, and Hyopil Shin. 2009. KTimeML: Specification of temporal and event expressions in Korean text. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pages 115–122, Suntec, Singapore. Association for Computational Linguistics.

ISO-24617-1. 2012. Language resource management - semantic annotation framework (semaf) - part 1: Time and events (semaf-time, iso-timeml). Standard, Geneva, CH.

ISO-24617-4. 2014. Language resource management-semantic annotation framework (semaf) - part 4: Semantic roles (semaf-sr). Standard, Geneva, CH.

ISO-24617-7. 2020. Language resource management-semantic annotation framework (semaf) - part 7: Spatial information. Standard, Geneva, CH.

ISO-24617-9. 2019. Language resource management-semantic annotation framework (semaf) - - part 9: Reference annotation framework (raf). Standard, Geneva, CH.

Adam Jatowt, Antoine Doucet, and Ricardo Campos. 2022. Diachronic analysis of time references in news articles. pages 918–923.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

Ryan Ka Yau Lai. 2023. From annotation to analysis: Exploring conversational dynamics with rezonater. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation*, pages 303–313.

L. Lange, A. Iurshina, H. Adel, and J. Strötgen. 2020. Adversarial alignment of multilingual models for extracting temporal expressions from text. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 103–109.

António Leal, Purificação Silvano, Evelin Amorim, Inês Cantante, Fátima Silva, Alípio Mario Jorge, and Ricardo Campos. 2022. The place of ISO-space in Text2Story multilayer annotation scheme. In *Proceedings of the 18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation within LREC2022*, pages 61–70, Marseille, France. European Language Resources Association.

Anne-Lyse Minard, Manuela Speranza, Ruben Urizar, Begoña Altuna, Marieke van Erp, Anneleen Schoen, and Chantal van Son. 2016. MEANTIME, the NewsReader multilingual event and time corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4417–4422, Portorož, Slovenia. European Language Resources Association (ELRA).

Phong H Nguyen, Kai Xu, Rick Walker, and BL William Wong. 2016. Timesets: Timeline visualization with set relations. *Information Visualization*, 15(3):253–269.

Sérgio Nunes, Alípio Mario Jorge, Evelin Amorim, Hugo Sousa, António Leal, Purificação Silvano, Inês Cantante, and Ricardo Campos. 2024. Text2Story lusa: A dataset for narrative analysis in European Portuguese news articles. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15773–15782, Torino, Italia. ELRA and ICCL.

Helene Olsen, Étienne Simon, Erik Velldal, and Lilja Øvrelid. 2024. Socio-political events of conflict and unrest: A survey of available datasets. In *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pages 40–53.

James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Rob Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. pages 28–34.

James Pustejovsky and Marc Verhagen. 2009. SemEval-2010 task 13: Evaluating events, time expressions, and temporal relations (TempEval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, pages 112–116, Boulder, Colorado. Association for Computational Linguistics.

Jeff Sauro and James R Lewis. 2016. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Christina Schokkenbroek. 1999. News stories: Structure, time and evaluation. *Time & Society*, 8(1):59–98.

Purificação Silvano, Evelin Amorim, António Leal, Inês Cantante, Maria de Fátima Henriques da Silva, Alípio Jorge, Ricardo Campos, and Sérgio Sobral Nunes. 2023a. Annotation and visualisation of reporting events in textual narratives. In *Proceedings of Text2Story 2023: Sixth Workshop on Narrative Extraction From Texts*.

Purificação Silvano, António Leal, Fátima Silva, Inês Cantante, Fatima Oliveira, and Alípio Mario Jorge. 2021. Developing a multilayer semantic annotation scheme based on ISO standards for the visualization of a newswire corpus. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 1–13, Groningen, The Netherlands (online). Association for Computational Linguistics.

Purificação Silvano, Evelin Amorim, António Leal, Inês Cantante, Alípio Jorge, Ricardo Campos, and Nana Yu. 2024. Untangling a web of temporal relations in news articles. In *Proceedings of Text2Story 2024 - Seventh Workshop on Narrative Extraction From Texts*, volume 3671 of *CEUR Workshop Proceedings*, pages 77–92.

Purificação Silvano, Alípio Jorge, António Leal, Evelin Amorim, Hugo Sousa, Inês Cantante, Ricardo Campos, and Sérgio Nunes. 2023b. Text2story lusa annotated corpus. Data set.

Hugo Sousa, Ricardo Campos, and Alípio Jorge. 2023. Tei2go: A multilingual approach for fast temporal expression identification. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, page 5401–5406, New York, NY, USA. Association for Computing Machinery.

Laura South, David Saffo, Olga Vitek, Cody Dunne, and Michelle A Borkin. 2022. Effective use of likert scales in visualization evaluations: A systematic review. In *Computer Graphics Forum*, volume 41, pages 43–55. Wiley Online Library.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association : JAMIA*, 20.

Tan Tang, Renzhong Li, Xinke Wu, Shuhan Liu, Johannes Knittel, Steffen Koch, Thomas Ertl, Lingyun Yu, Peiran Ren, and Yingcai Wu. 2020. Plotthread: Creating expressive storyline visualizations using reinforcement learning. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):294–303.

Tan Tang, Sadia Rubab, Jiewen Lai, Weiwei Cui, Lingyun Yu, and Yingcai Wu. 2018. istoryline: Effective convergence to hand-drawn storylines. *IEEE transactions on visualization and computer graphics*, 25(1):769–778.

Xiaozhi Wang, Yulin Chen, Ning Ding, Hao Peng, Zimu Wang, Yankai Lin, Xu Han, Lei Hou, Juanzi Li, Zhiyuan Liu, et al. 2022. Maven-ere: A unified large-scale dataset for event coreference, temporal, causal, and subevent relation extraction. *arXiv preprint arXiv:2211.07342*.

Yunchao Wang, Guodao Sun, Zihao Zhu, Tong Li, Ling Chen, and Ronghua Liang. 2024. E2 storyline: visualizing the relationship with triplet entities and event discovery. *ACM Transactions on Intelligent Systems and Technology*, 15(1):1–26.

Li Ye, Lei Wang, Shaolun Ruan, Yuwei Meng, Yigang Wang, Wei Chen, and Zhiguang Zhou. 2024. Storyexplorer: A visualization framework for storyline generation of textual narratives. *arXiv preprint arXiv:2411.05435*.

Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. 2021. Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1384–1399.

Xiaoshi Zhong and Erik Cambria. 2023. Time expression recognition and normalization: a survey. *Artificial Intelligence Review*, 56(9):9115–9140.

Xiaoshi Zhong, Chenyu Jin, Mengyu An, and Erik Cambria. 2024. Xtime: A general rule-based method for time expression recognition and normalization. *Knowledge-Based Systems*, 297:111921.

A Time Expressions and Events Attributes Statistics

Time Expressions		
Attribute Name	Attribute Value	N.
Type	Date	226
	Duration	16
	Time	43
Temporal Function	Publication Time	67
Total		294

Table 2: Time Expressions Statistics

Events		
Attribute Name	Attribute Value	N.
Class	Occurrence	133
	Reporting	43
	State	16
	I_State	8
	I_Action	1
	Aspectual	1
	Perception	1
Type	Transition	173
	State	24
	Process	7
Pos	Verb	177
	Noun	13
	Adjective	5
	Preposition	6
Tense	Past	133
	Imperfect	5
	Future	7
	Present	10
Aspect	Perfective	161
	Imperfective	5
	Progressive	1
Polarity	Positive	201
VForm	Participle	18
	Infinitive	9
	Gerundive	2
Movement	Motion Literal	4
Modality	*Poder* (may)	7
	Dever (should)	1
Mood	Future	5
	Conditional	2
	Subjunctive	1
Total	204	

Table 3: Events Expressions Statistics

B Questionnaire About the Visualization

1. The visualization is easy to navigate.
2. The narrative components - time expressions and events - are visually distinct.
3. The relationships between narrative components are clearly represented.
4. The visualization effectively supports the identification of annotation errors in time expressions, events, and their relations.
5. The visualization enables the identification of temporal patterns.
6. Compared to the BRAT annotation tool, the proposed visualization provided a better depiction of the relations between all the time expressions and their connected events

C Examples of Identified Error Annotation and Insights

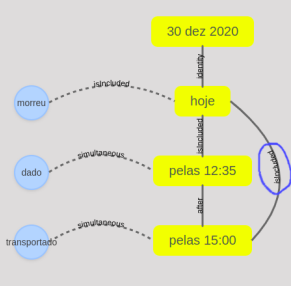


Figure 5: In the visualization of lusa_49, the connections were correct. However, these temporal connections did not capture the critical information about the temporal relation between the time interval denoted by "pelas 15 horas" (around 3 p.m.) and the time interval denoted by "hoje" (today). Thus, we added to the annotation manual that temporal expressions can connect to two or more temporal expressions, ensuring that the correct chronology is captured. This figure already represents that missing link.

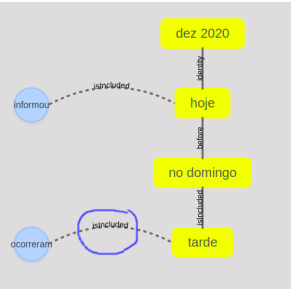
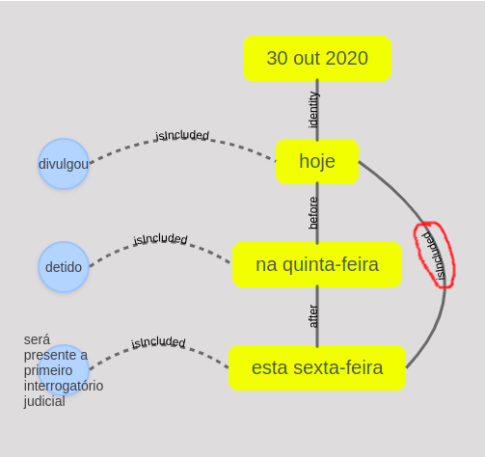
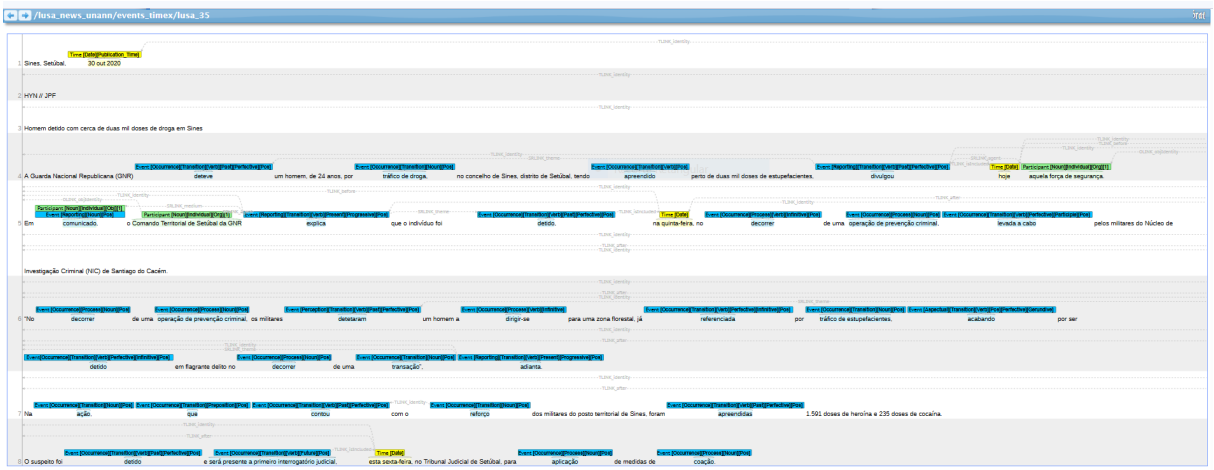


Figure 6: In the scenario of the lusa_82 news story, the event "ocorreram" (occurred) was initially linked to "no domingo" (on Sunday), which did not allow us to infer that the event took place during the time frame denoted by "tarde" (afternoon). We needed to modify the guidelines to connect "ocorreram" to "tarde." This adjustment enables us to infer a relationship between "ocorreram" and "no domingo," establishing a transitive relation. The figure already illustrates the correct temporal representation.



(a) The relation between “esta sexta” (this Friday) and “hoje” (today) is incorrect in the Lusa_35 news story. The IsIncluded link was annotated instead of the Identity link. This figure shows the representation that highlights this annotation error.



(b) BRAT visualization of anotations of news story lusa_35

Figure 7: Comparing the lusa_35 news story using the proposed visualization and BRAT.

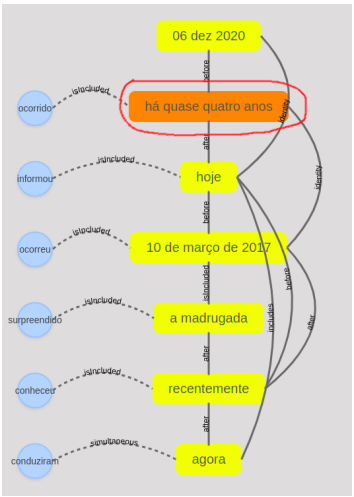


Figure 8: In the visualization of lusa_76, “há quase 4 anos” (almost 4 years ago) is not a temporal expression of type Duration, so it should be in yellow (not orange).