Text2Story 2023
Sixth International Workshop on Narrative Extraction from Texts

Annotation and visualization of reporting events in textual narratives

Purificação Silvano, Evelin Amorim, António Leal, Inês Cantante,
Fátima Silva, Alípio Jorge, Ricardo Campos & Sérgio Nunes

fct Fundação para a Ciência e a Tecnologia
U.PORTO
clup Centro de Linguística da Universidade do Porto
INESCTEC
Ci2.ipt Smart Cities Research Center

1. The problem

News articles

narrative nature

recounting events

use of reported speech

(1) The police authority said that the suspect did not accept the money.

1. The problem

News articles

- Two intertwined levels of discourse:
  - the level of reported events: the story of the news
  - the level of reporting events: the information about the sources

- Two narratives:
  - the narrative of the story
  - the narrative of the report

1. The problem

News articles

Why is it important to separate the story's narrative from the report's narrative?

- to extract narratives
- to generate visualizations

BUT,

- majority of the computational models lack formalisms to represent the two levels and the relation between them (Gérvas, 2021).

2. Our objectives

- ✓ demonstrate how the reporting events that build the reporting narrative can be properly integrated into a multilayer annotation scheme designed initially to extract the story narrative from news articles;

- ✓ put forward a proposal of building a visualization format for the reporting narrative;

- ✓ describe some lexico-grammatical features of reporting events, which can contribute to their automatic detection.

3. Related work

3.1. Reporting events:

- in (NLP) tasks: **attribution** relations, that is, relations between pieces of information and the sources expressing them (Pareti, 2009).

source: the original producer of the content, (thematic role of experiencer or agent).	cue: the linguistic element that explicitly links source and content (a verb, noun,…).	content: the linguistic material that someone, usually other than the speaker, said.

(2) The police said that the suspect resisted arrest.

3. Related work

3.1. Reporting events:

- Attribution relations: annotated in dedicated schemes (Pareti, 2015) or merged into the annotation schemes of other linguistic aspects (temporal relations and discourse relations).

Annotation of reporting events (TimeBank, FactBank)	Annotation of DR attribution (RST, Mann & Thompson, 1987), PDTB (Prasal et al., 2008)	Annotation of DR Attribution in a separate layer (ISO-8).

- Attribution relations in the news: extensively studied for English, but not Portuguese (eg. Pardo et al. 2004).

3. Related work

3.2. Visual representation of events in narratives:

- to allow for a more accessible investigation of relations between the events that are represented in different types of discourse (eg. Sun et al. 2013).

- to speed up the skimming of texts since it summarizes the relevant information in one picture (Chongtay et al. 2013).

3. Related work

3.2. Visual representation of events in narratives:

Knowledge Graphs widely used, but difficult to read (Rospocher et al., 2016)

Message Sequence Charts useful for temporal timelines (Palshikar, et al., 2019)

Both lack mechanisms to represent events that belong to different layers, as is the case of the reporting and reported events in news articles.

4. The study

4.1. Data and annotation

- ✓ 120 news articles in European Portuguese from the Lusa News Agency;

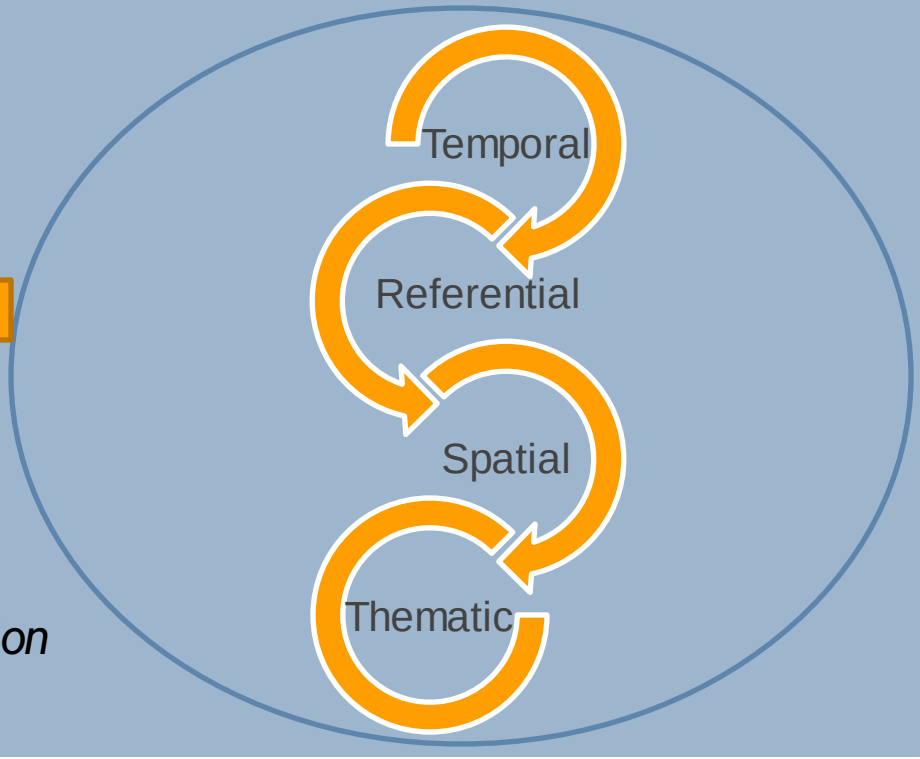
- ✓ fully annotated according to a multilayer annotation scheme (Silvano et al. (2021); Leal et al. (2022).

4. The study

4.1. Data and annotation

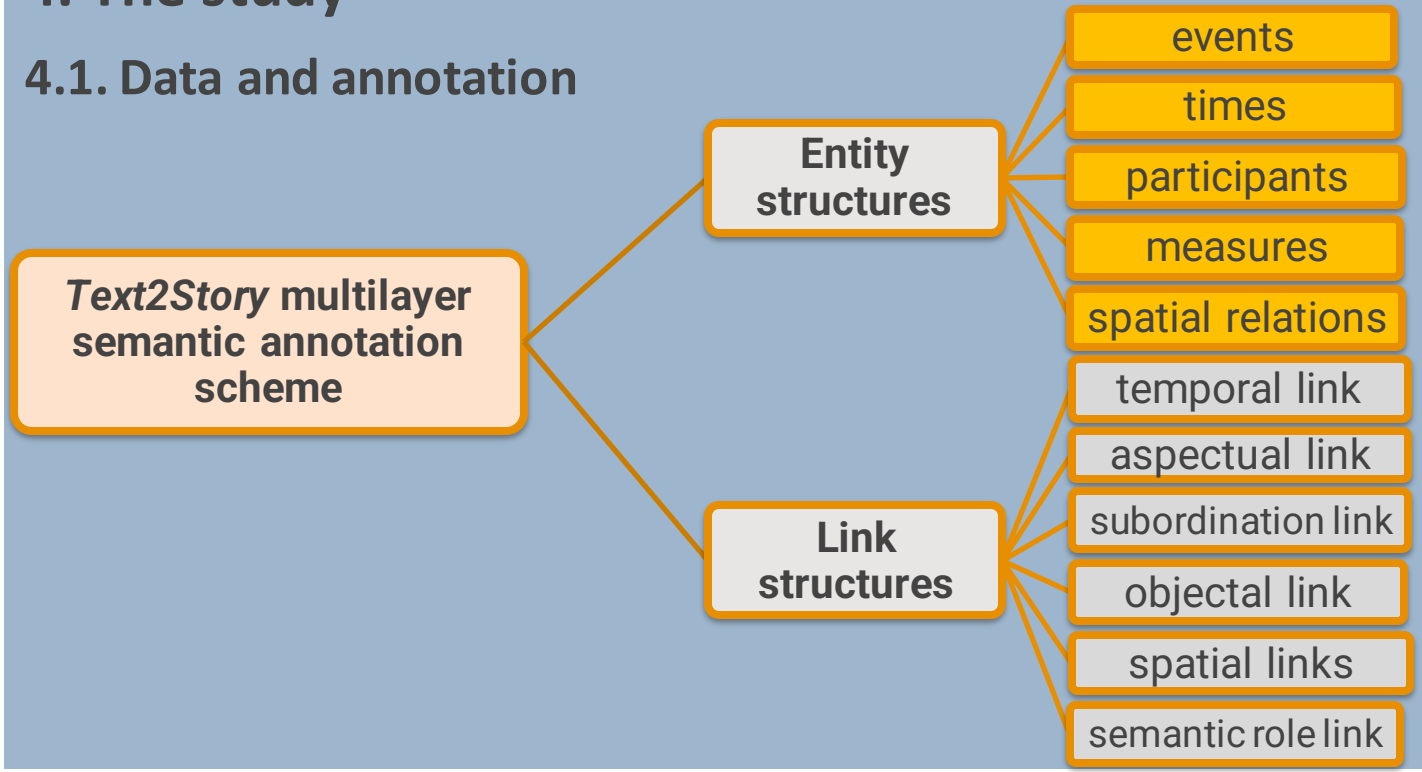
- Coherent
- Interoperable

ISO 24617 *Language resource management - Semantic annotation framework* (Parts 1, 4, 7, 9)



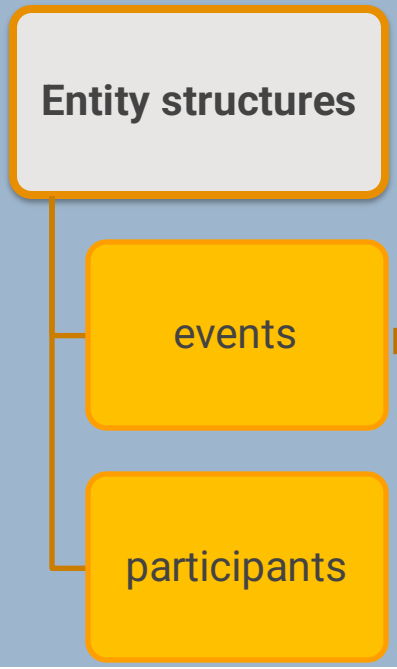
4. The study

4.1. Data and annotation



4. The study

4.1. Data and annotation

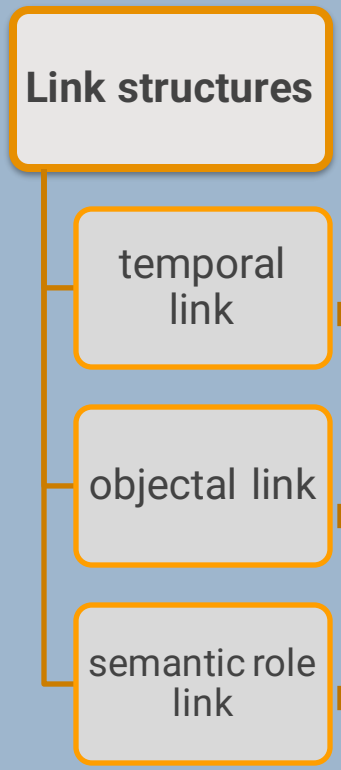


Reporting event: an action of a person/ organization declaring something, narrating a situation, or informing about a situation (eg. markables *informou, disse* ('informed', 'said'); *segundo/ de acordo* ('according to')+ noun phrase.

Source: markables identifying the reporting source annotated in the participant layer.

4. The study

4.1. Data and annotation



reporting events linked together by temporal link Identity (impossible to determine a temporal ordering between the reporting events)

establishes the connection between referential expressions (eg. the same person)

participants linked to events by semantic role links of agent or medium

4. The study

4.2. The visual representation construction

- A new visualization scheme that provides a mechanism to represent two different narrative layers: the reported narrative and the reporting narrative.

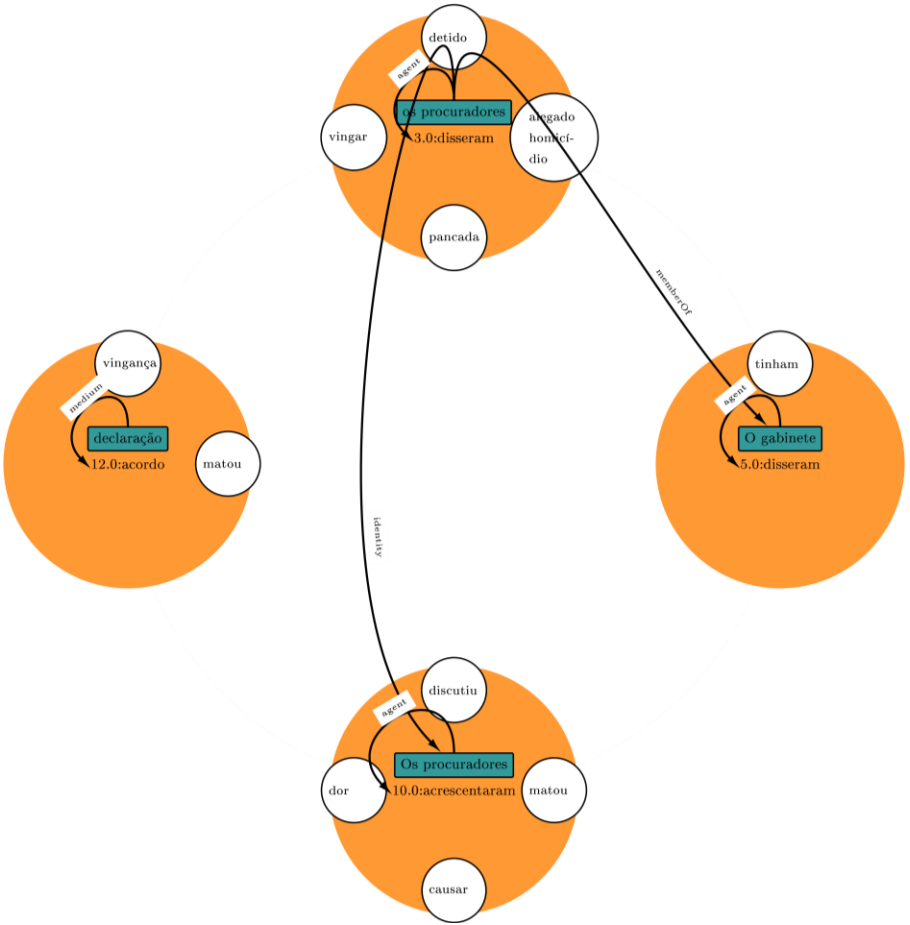
  - Enables annotators and other users to have a bird's eye view of the interplay of the two layers to:
    - detect inconsistencies in the annotation;
    - assess the main lexico-grammatical features of reporting events that can be used to develop models for automatic extraction.

4. The study

4.2.1. The Bubble data structure

- Bubble: a data structure with a given event object and a set of events' relations.
- Big Bubble: pre-defined events that can include other events represented by bubbles, in our case, the reporting events linked by the temporal link Identity. It includes:
  - information about the participants that intervene in the reporting events annotated with the semantic role of agent and of medium;
  - objectal links between those participants.

(4) Um homem foi detido no norte do México pelo alegado homicídio dos três filhos à pancada para se vingar da mãe das crianças, disseram os procuradores no domingo. O gabinete dos procuradores no estado de Sonora (norte) disseram que os rapazes tinham 3,7 e 8 anos. Os procuradores acrescentaram que o homem discutiu com a mãe das crianças e matou os filhos "de modo a causar-lhe uma grande dor". "Aparentemente, numa vingança contra a mulher, matou os filhos", de acordo com uma declaração publicada pelo gabinete.



4. The study

4.2.2. Building the visualization

- Built using the Tikz (Tantau, 2013) package that produces pictures for LATEX.
- The code for producing such visualization is available as a python package

https://colab.research.google.com/drive/1V2DCuP1qAlwUXThTKNUnZ98WxARZXC_v?usp=sharing

4. The study

4.2.3. Some findings about reporting events

- 120 news articles: 3040 events, 349 reporting events (average of 2.09083 reporting events per news).
- Reported events: verbs, nouns and discourse connectors.
  - verbs: 272 occurrences of 35 verbs (*dizer* 'to say', *referir* 'to refer', *informar* 'to inform', *anunciar* 'to announce', *acrescentar* 'to add' and *adiantar* 'to say');
  - nouns: 2 (*declarações* 'declarations' and *publicação* 'publication');
  - discourse connectors,: 75 (*segundo/ de acordo*... 'according to')

4. The study

4.2.3. Some findings about reporting events

- How are the reported events represented in the news?

- ✓ Three configurations.
- ✓ Common trait: only one main source of information.

➢ How are the reported events represented in the news?

1st configuration $R_0 \Longrightarrow$ story

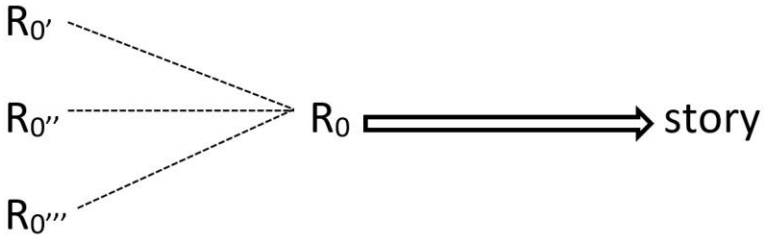
(5) A 39-year-old man was placed under electronic surveillance in Penafiel, Porto district, suspected of physical violence, injuries and death threats to his 42-year-old ex-partner, a GNR source reported today. In a statement, the police authority indicated that the suspect did not accept the separation...

2nd configuration (1st variant) $R_{0'}$ ---------- $R_0 \Longrightarrow$ story

(6) Dutch international Quincy Promes, Ajax footballer, was arrested for allegedly stabbing a family member in July, Dutch media reported today, citing police sources.

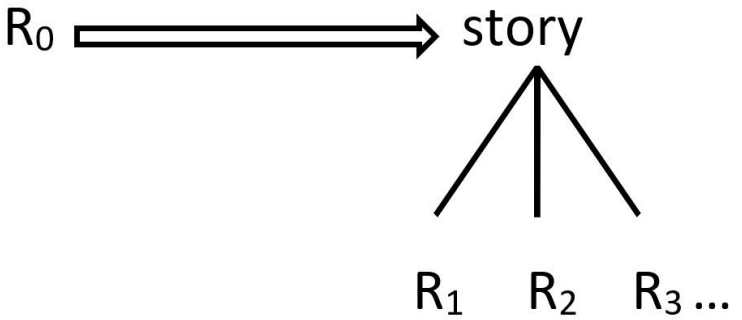
➢ How are the reported events represented in the news?

2nd configuration (2nd variant)



(7) Three people died in the municipality of Camacupa, in the Angolan province of Bié, following an attack by a hippopotamus, public radio reported today. (…) According to one of the survivors, the attack took place early Sunday night (…) The municipal administrator of Muinha, Baptista Sambango, said that the incident was reported to the municipal administrator…

The 3rd configuration



(8) The Sonora state attorney's office said the boys were ages 3, 7 and 8. The suspect called his father to tell him what he had done…

5. Conclusions and future work

- Our approach:
  - enables us to adequately include reporting events that constitute the reporting narrative in news articles in a multilayer annotation and portray them in visual representations.
  - has the same descriptive power of proposals that work with the concept of attribution, being more comprehensive and allowing a wide variety of annotations and representations.

5. Conclusions and future work

- In the future:
  - enrich the Bubble visualizations with more relevant information, namely temporal relations;
  - design an experiment with other alternatives to assert the degree of suitability of our proposal of visualization compared to others.

References

P. Gervás, A model of interpretation of embedded stories, in: Text2Story: 4th International Workshop on Narrative Extraction from Texts, CEUR Workshop Proceedings, CEUR Workshop Proceedings, Lucca, Tuscany, 2021.

S. Pareti, Towards a discourse resource for Italian: developing an annotation schema for attribution, Ph.D. thesis, Master's thesis, Faculty of Letters and Philosophy, University of Pavia, Italy, 2009.

S. Pareti, Attribution: a computational approach (2015).

R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. K. Joshi, B. L. Webber, The penn discourse treebank 2.0., in: LREC, 2008.

W. Mann, S. Thompson, Rethorical Structure Theory: Toward a functional theory of text organization, Text 8 (1988) 243–281. doi:10.1515/text.1.1988.8.3.243.

N. Asher, N. M. Asher, A. Lascarides, Logics of Conversation, Cambridge University Press,
2003.

ISO-24617-6, Language resource management- Semantic annotation framework (SemAF) - Part 6:Principles of semantic annotation (SemAF Principles), Standard, Geneva, CH, 2016.

T. A. S. Pardo, M. das Graças Volpe Nunes, L. H. M. Rino, Dizer: An automatic discourse analyzer for brazilian portuguese, in: Advances in Artificial Intelligence–SBIA 2004: 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil, September 29-Ocotber 1, 2004. Proceedings 17, Springer, 2004, pp. 224–234.

Y. Sun, Y. Tao, G. Yang, H. Lin, Visitpedia: Wiki article visit log visualization for event exploration, in: 2013 International Conference on Computer-Aided Design and Computer Graphics, IEEE, 2013, pp. 282–289.

R. Chongtay, M. Last, M. Verbeke, B. Berendt, Summarize to learn: summarization and visualization of text for ubiquitous learning, in: Proceedings of the 3rd IEEE Workshop on Interactive Visual Text Analytics at VIS 2013, The 3rd IEEE Workshop on Interactive Visual Text Analytics, Atlanta, GA, 2013, pp. 1–4.

G. Palshikar, S. Pawar, S. Patil, S. Hingmire, N. Ramrakhiyani, H. Bedi, P. Bhattacharyya, V. Varma, Extraction of Message Sequence Charts from Narrative History Text, in: Proceedings of the First Workshop on Narrative Understanding, 2019, pp. 28–36.

P. Silvano, A. Leal, F. Silva, I. Cantante, F. Oliveira, A. Mario Jorge, 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, Association for Computational Linguistics, Groningen, The Netherlands (online), 2021, pp. 1–13. URL: https://aclanthology.org/2021.isa-1.1.

A. Leal, P. Silvano, E. Amorim, I. Cantante, F. Silva, A. Mario Jorge, R. Campos, 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, European Language Resources Association, Marseille, France, 2022, pp. 61–70. URL: https://aclanthology.org/2022.isa-1.8.

T. Tantau, The TikZ and PGF Packages, 2013. URL: http://sourceforge.net/projects/pgf/.

Thank You!

Obrigada!

Any question?

Text2Story 2023
Sixth International Workshop on Narrative Extraction from Texts

Annotation and visualization of reporting events in textual narratives

Purificação Silvano, Evelin Amorim, António Leal, Inês Cantante,
Fátima Silva, Alípio Jorge, Ricardo Campos & Sérgio Nunes

fct Fundação para a Ciência e a Tecnologia
U.PORTO
clup Centro de Linguística da Universidade do Porto
INESCTEC
Ci2.ipt Smart Cities Research Center