Untangling a Web of Temporal Relations in News Articles

Purificação Silvano, António Leal, Evelin Amorim, Inês Cantante, Alípio Jorge, Ricardo Campos & Nana Yu

1. Motivation

Determining the temporal dimension of narratives

Key to understand them

Focus of several studies both in Linguistics and NLP

Still very complex

1. Motivation

Temporal information extraction

(i) recognition of events and temporal expressions;

(ii) recognition of temporal relations between them;

(iii) time-line constructions based on the temporal relations.

❑ Very challenging tasks, especially (ii) and (iii) (Leeuwenberg et al., 2018, Santana et al., 2023), even more, when dealing with texts with a non-chronological order, like news articles.

❑ Importance of the input of linguistic analysis.

1. Motivation

❖ **News articles as narratives**

➢ An intricate temporal structure compared to other types of narratives (Zahid et al., 2019).

➢ Cyclic or "zigzagging pattern, with the timeline repeatedly moving into the past and the future concerning the main event" (Chovanec et al., 2014).

➢ Two narratives: two intertwined levels of discourse (Silvano et al., 2023)
  ➢ Narrative of the story: the level of reported events: the story of the news
  ➢ Narrative of the report: the level of reporting events: the information about the sources

(1) The police chief said that the accident occurred at 4 pm. He added that one of the drivers was injured and taken to the hospital.

2. Objective and Research questions

➢   Conduct a comprehensive study of the temporal structure of news articles that can:
   ➢   determine the prototypical temporal structure of news
   ➢   contribute to the improvement of temporal information extraction

RQ1: How are the events temporally organised at the story level?

RQ2: What is the report level's role in the temporal organisation of the story?

RQ3: What is the association between the temporal relations and some of the events' grammatical features?

3. Dataset and annotation

❖ T2S Lusa Annotated Dataset[1] (a subset of the T2S Lusa dataset[2])

o 119 news annotated with **semantic information**

o Multilayer annotation scheme (Silvano et al, 2021; Leal et al., 2022):
   o Combining four parts of ISO-24617

      o Part 1 – Time and Events (ISO-24617-1, 2012)
      o Part 4 – Semantic Roles (ISO-24617-4, 2014)
      o Part 7 – Spatial Information (ISO-24617-7, 2020)
      o Part 9 – Reference Annotation (ISO-24617-9, 2019)

[1] https://doi.org/10.25747/ESFS-1P16

[2] https://doi.org/10.25747/ET95-BX90

EVENTS
- **class**: occurrence, state, reporting, perception, aspectual, i-action, i-state
- **type**: state, process, transition
- **pos**: verb, noun, adjective, preposition, other
- **tense**: present, past, future, imperfect, none

TEMPORAL LINK: Before, After, Includes, Is_included, During, Simultaneous, Identity, Begins, Ends, Begun_by, Ended_by

3. Dataset and annotation

❖ **Methodology of annotation**

- Training of an annotator (a PhD student in Linguistics, specialising in Semantics).

- Discussion of problematic cases with a team of linguists.

- Large-scale annotation.

- Validation of the annotation: annotation by a second annotator of a sample of 10% of the dataset (19 news articles) to conduct inter-annotator agreement.

3. Dataset and annotation

❖ **Methodology of annotation**

   ○ Inter-annotator agreement

- Events identification (computed as a pairwise f1): **0.77 (substantial agreement)**
- Attributes of events (computed as Cohen's Kappa score):
  - Class (**0.63**), Type (**0.51**), Part of Speech (**0.81**), and Tense (**0.74**) (**substantial the agreement**, except for Type)
- Temporal links between events (computed as Cohen's Kappa): **0.31 (fair agreement)**
- Attributes of temporal links (computed as Cohen's Kappa): 0.32 **(fair agreement)**

4. Analysis

❖ Extraction and visualisation of temporal information with **the package text2story**[3]:

❖ It is devoted to automatically extracting narratives programmatically and easily.

❖ It offers three types of visualisation for narratives: Knowledge Graph (KG), Message Sequence Chart (MSC), and Bubble Diagram (BM).

[3]https://pypi.org/project/text2story/

4. Results

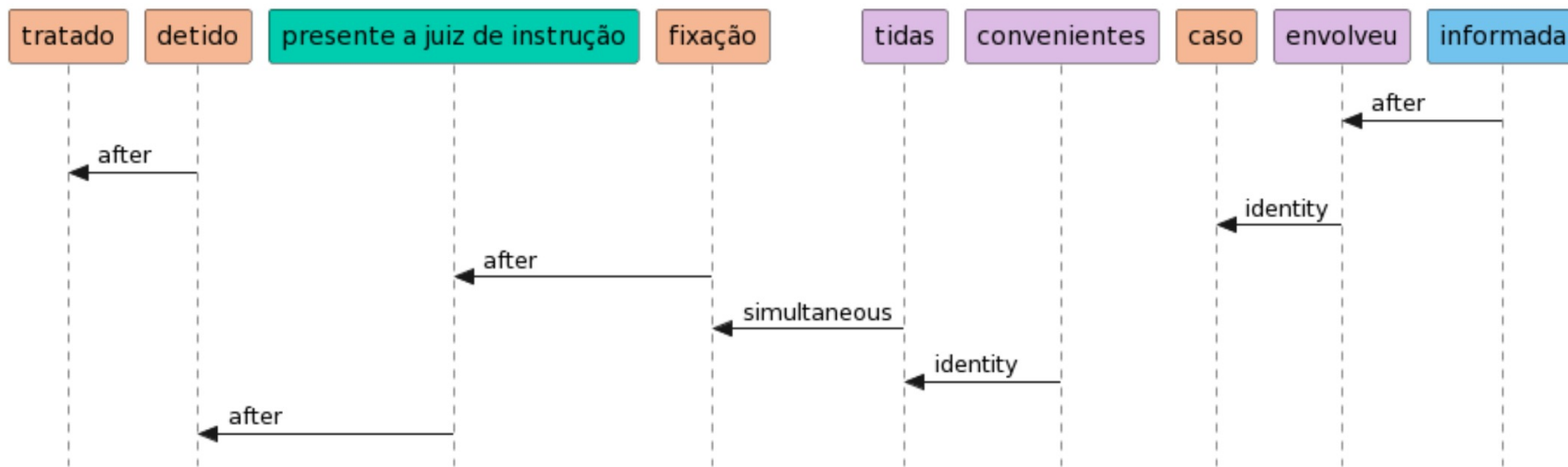❖ **General observations:**

○ <u>Total of events</u>: 3068 events were identified at the story and report levels.

○ <u>Total of events of the the story</u>: 2651 events (an average of 22.27 events per news).

○ <u>Total of reporting events</u>: 417 (an average of 3.50 reporting events per news).

4. Results

RQ1: How are the events temporally organised at the story level?

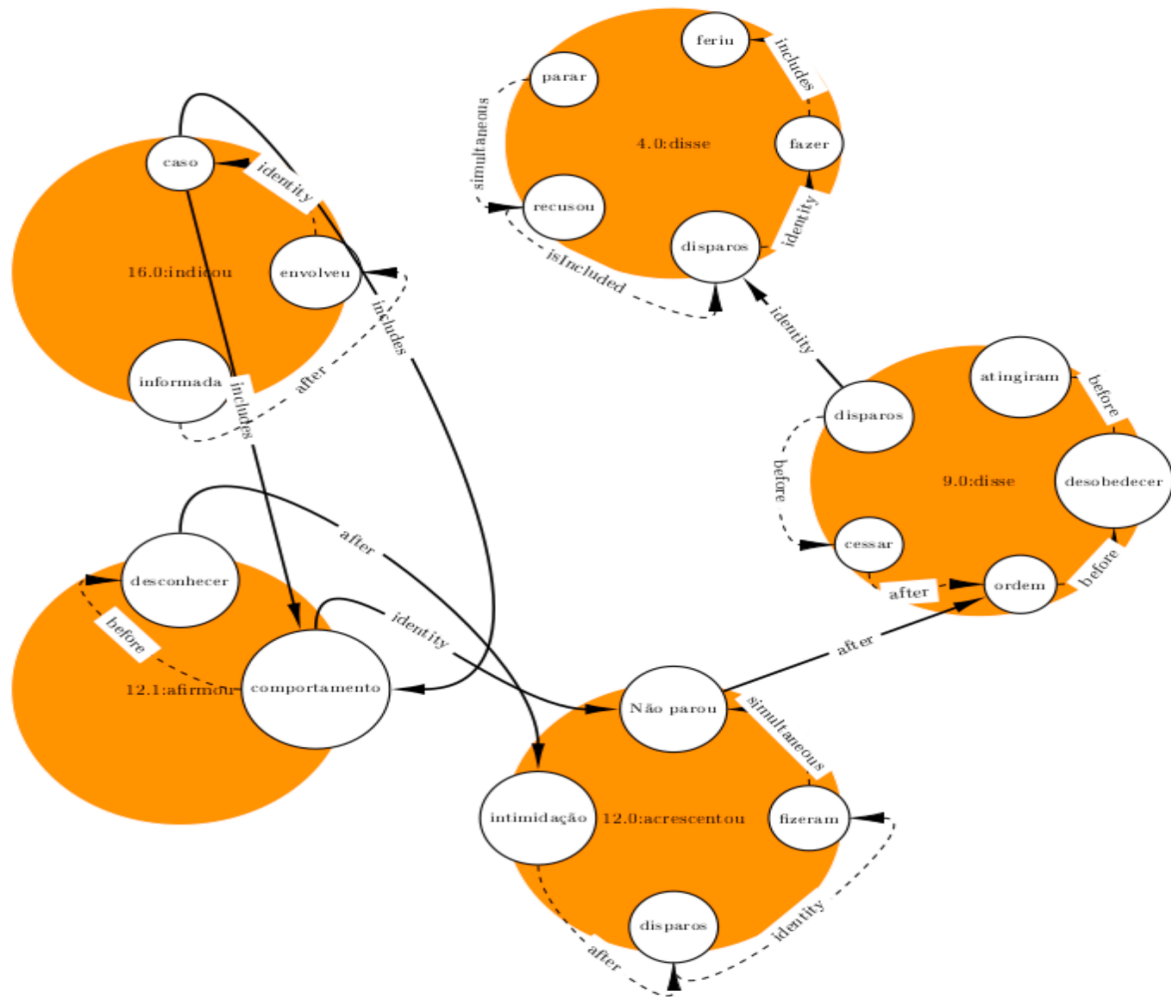| Relation	Between events...		
	(A)....of the story	(B)....embedded in the reports	(C)....across the reports
Identity	587	255	16
After	546	173	15
Before	420	151	7
Simultaneous	413	176	3
Includes	278	125	6
Is_included	232	109	5
During	39	23	0
Begun_by	4	3	0
Begins	3	1	0
Ends	2	0	0
Ended_by	2	1	0

4. Results

(2) After being **treated** for minor injuries at a hospital, he was **detained** at the GNR post in Cabeceiras. He will be **brought** before an investigating judge today for the **setting** coercive measures **established** as **appropriate**. As the **case involved** firearms, the Judicial Police will be **informed** of the details, the source **said**.

4. Results

RQ2: What is the report level's role in the temporal organisation of the story?

Relation	(A)....of the story	Between events...	
		(B)....embedded in the reports	(C)....across the reports
Identity	587	255	16
After	546	173	15
Before	420	151	7
Simultaneous	413	176	3
Includes	278	125	6
Is_included	232	109	5
During	39	23	0
Begun_by	4	3	0
Begins	3	1	0
Ends	2	0	0
Ended_by	2	1	0

4. RESULTS

4. Results

RQ3: What is the association between the temporal relations and some of the events' grammatical features?

	Transitions (e.g., 'read a book', 'win a race')	Processes (e.g., 'sleep', 'swim')	States (e.g., 'be tall', 'like', 'live')
Telicity	✓	✗	✗
Duration	✓/ ✗	✓	✓

Successivity

Simultaneity

4. Results

RQ3: What is the association between the temporal relations and some of the events' grammatical features?

Type of Events		After	Before	Simultaneous	Includes	Is_Included
'Transition', 'Transition'	Count	422	284	132	20	44
	Percentage	61.26%	50.27%	36.16%	6.51%	16.3%
'Transition', 'State'	Count	104	79	39	9	88
	Percentage	15.09%	13.98%	10.68%	2.93%	32.59%
'Transition', 'Process'	Count	59	29	21	3	30
	Percentage	8.56%	5.13%	5.75%	0.98%	11.11%
'State', 'Transition'	Count	44	66	73	180	16
	Percentage	6.39%	11.77%	20%	58.55%	5.93%
'Process', 'Transition'	Count	43	40	27	21	8
	Percentage	6.24%	7.08%	7.40%	6.84%	2.96%

5. Conclusions

❖ Our study: a comprehensive understanding of the narrative temporal structure of news articles, contributing to the field of narrative analysis.

❖ Linguistic analysis: valuable insights into how news articles convey events.

▪ to be used to improve models to retrieve or predict temporal information.

References

A. Leeuwenberg, M.-F. Moens, Temporal information extraction by predicting relative time-lines, in: Proceedings of the EMNLP'18, ACL, Brussels, Belgium, 2018, pp. 1237–1246.

B. Santana, R. Campos, E. Amorim, A. Jorge, P. Silvano, S. Nunes, A survey on narrative extraction from textual data, Artif Intell Rev (2023).

I. Zahid, H. Zhang, F. Boons, R. Batista-Navarro, Towards the automatic analysis of the structure of news stories, CEUR Workshop Proceedings 2342 (2019) 71–79. 2nd International Workshop on Narrative Extraction From Texts, Text2Story 2019 ; Conference date: 14-04-2019.

J. Chovanec, Pragmatics of Tense and Time in News, Pragmatics & Beyond New Series, John Benjamins Publishing Company, Amsterdam, 2014.

P. Silvano, E. Amorim, A. Leal, I. Cantante, F. Silva, A. Jorge, R. Campos, S. Nunes, Annotation and Visualisation of Reporting Events in Textual Narratives, in: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (Eds.), CEUR Workshop Proceedings, volume 3370 of CEUR Workshop Proceedings, 2023, pp. 47–62.

Thank You!

Obrigada!

4. Analysis

RQ1: How are the events temporally organised at the story level?

Focus on the temporal relations between all the events that comprise the story, excluding the reporting events belonging to the report level, linked by the TLink Identity.

(2) A source from the GNR's Territorial Command in Braga told Lusa news agency that the officers only hit the man after he disobeyed the order to stop firing.

"He didn't even stop when the soldiers fired intimidating shots into the air," added the source, who said he didn't know the cause of the detainee's behaviour.

4. Analysis

RQ2: What is the report level's role in the temporal organisation of the story?

Analysis of two types of information: (i) TLinks between all the events within each report and (ii) the TLinks between the first story event of report block B and any story event from report block A.

(2)

Report A: A source from the GNR's Territorial Command in Braga told Lusa news agency that the officers only hit the man after he disobeyed the order to stop firing.

Report B: "He didn't even stop when the soldiers fired intimidating shots into the air," added the source, who said he didn't know the cause of the detainee's behaviour.

4. Analysis

RQ3: What is the association between the temporal relations and some of the events' grammatical features?

Analysis of the features of the events that form the story, information from the tags Class, Type, Part of Speech (PoS), and Tense.

(3) The boys played football. The girls swam in the pool. – simultaneity relation

(4) The boys broke the window glass. The girls tore the curtain. - successivity