

Text2Story Annotation Manual

May 2024

Preface

This manual describes the process of developing a multilayer semantic annotation scheme designed for extracting information from a European Portuguese corpus of news articles, at four levels, temporal, referential, spatial and semantic role labelling. The framework was developed within the projects *Text2Story Extracting journalistic narratives from text and representing them in a narrative modeling language*, with reference PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185) and *StorySense: Uncovering deeper layers of Narrative Understanding*, with reference 2022.09312.PTDC ([DOI 10.54499/2022.09312.PTDC](https://doi.org/10.54499/2022.09312.PTDC)).

The proposed scheme harmonizes parts 1, 4, 7 and 9 of ISO 24617 *Language resource management - Semantic annotation framework*. This annotation framework includes a set of entity structures (participants, events, times, spatial relations and measures) and of links (temporal, aspectual, subordination, objectal, spatial, and semantic roles) with several tags and attribute values.

Sources

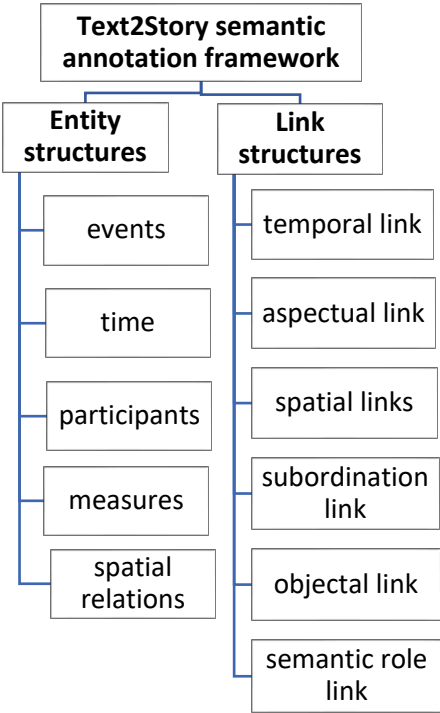
ISO24617-1:2012, Language resource management-Semantic annotation framework (SemAF)
- Part 1: Time and events (SemAF-Time, ISO-TimeML)

ISO-24617-4: 2014, Language resource management- Semantic annotation framework (SemAF) - Part 4: Semantic roles (SemAF-SR)

ISO 24617-9: 2019, Language resource management-Semantic annotation framework (SemAF)
- Part 9: Reference annotation framework (RAF)

ISO 24617-7: 2020 (second edition), Language resource management - Semantic annotation framework (SemAF) - Part 7: Spatial information

I. Multilayer semantic annotation framework



II. Stages of the annotation process

1 st stage	Annotation of events
	Annotation of times
	Annotation of temporal, aspectual and subordination links
2 nd stage	Annotation of participants
	Annotation of objectal relations
3 rd stage	Annotation of semantic role links
4 th stage	Annotation of spatial relations and measures
	Annotation of spatial and measure links

III. Annotation guidelines

A. Entity structures

A.1. Event

Definition

- eventuality that happens or occurs or state or circumstance that is temporally relevant, i.e., that is directly related to a temporal expression or change over the course of the text.

Tag span

- verb forms (only main verbs, not auxiliary verbs). Ex. *O polícia prende o ladrão. O João tem visitado a tia.*

Notes:

- in the case of aspectual verbs, they should be marked and linked to the main verb by an aspectual link.
- in case of a predicative construction, we leave the copular verb and mark the adjective/ noun/ prepositional complement, identifying the POS as an adjective or noun, but assigning the values of the tags TENSE, and ASPECT according to the information from the verb. Ex. *A polícia estava pronta.*

If the copula verb is followed by a prepositional or adverbial phrase with an interpretation of spatial or temporal location, mark only the verb. Ex. *O João está ali. O João ficou em Lisboa.*

- if the verb is included in a (semi) lexicalized expression, mark that expression as the predicator. Ex. *O suspeito será presente a tribunal.*
- in those cases of combinations of verb *ir* + V-infinitive, consider the following two scenarios:
 - the 2 verbs are in adjacency: consider only the main verb. Ex. *foi pedir satisfações a casa do vizinho.*
 - the 2 verbs are not in adjacency: consider 2 autonomous markables. Ex. *foi a casa do vizinho pedir satisfações)*

- nouns without specifiers, complements and modifiers. Ex. *O ataque à cidade ocorreu às 12.00.*

Notes:

- in the case of polysemous nominalizations when they do not convey an interpretation of event, they are not marked. Ex. *O relato da testemunha refere que o médico agiu de forma errada.* In this case, the meaning of “relato” that is relevant is content.

- in the case of nouns that represent events and are followed by adjectival modifiers, or are followed by an adjective and form a (semi)lexicalized expression, mark the whole expression. Ex. *fiscalização rodoviária, prisão efetiva.*
- nouns in prepositional phrases: in PPs only mark the noun phrase head, that is, the noun. Ex. *O Prof. Silva, em sabática na universidade do Porto, fez uma descoberta importante.*
- adjectives without complements and modifiers Ex. *O vulcão Kilauea, dormente até ontem, começou a expelir lava.*

Attributes

- class:
 - OCCURRENCE – a situation that happens or occurs. Ex. *O polícia prendeu o ladrão.*
 - STATE – a situation in which something obtains or holds true, and that is temporally relevant, that is,
 - a state that changes over the text, or
 - related to temporal expressions, or
 - a state that is introduced by I-action, I-state or reporting class, or
 - predicative states that are still valid taking into consideration the document creation time.

Ex. *A vítima estava sozinha em casa.*
 - REPORTING – an action of a person/ organization declaring something, narrating a situation, informing about a situation,...Ex. *declarar, dizer, explicar, relatar*
 - PERCEPTION – a situation involving physical perception of another situation. Ex. *ver, ouvir*
 - ASPECTUAL – A situation that focuses on a different part of another situation. Ex. *começar, continuar, terminar,...*
 - I-ACTION (Intensional action) – an event that introduces another event as argument in an intensional context. Ex. *tentar, adiar, evitar, cancelar, pedir, ordenar, autorizar, prometer, oferecer, propor, decidir, concordar, nomear, sugerir, alegar*
 - I-STATE (Intensional state) – a state that introduces events as argument in an intensional context. Ex. *acreditar, pensar, imaginar, querer, esperar, odiar, preocupar-se, exigir, estar preparado/ pronto/ ansioso*

- type:
 - STATE – a situation in which something obtains or holds true. Ex. *A vítima estava presa.*
 - PROCESS – a situation that is dynamic and atelic, with duration. Ex. *O João nadou.*
 - TRANSITION – a situation that is dynamic, with duration and consequent state. It subsumes accomplishments and achievements. Ex. *O João leu um livro./ O João ganhou a corrida.*

- pos:
 - VERB
 - NOUN
 - ADJECTIVE
 - PREPOSITION

- tense:
 - PRESENT – locates a situation at the present time, or in a time interval that starts in the past and still holds in the present. It can also depict habitual situations in the present.
 - PAST – locates situations in a past time.
 - FUTURE – locates situations in the future.
 - IMPERFECT – locates an ongoing situation in the past.

- aspect:
 - PROGRESSIVE – ongoing or continuous situation. Ex. *O soldado estava a sorrir.*
 - PERFECTIVE – situation that ended. Ex. *O polícia prendeu o assaltante.*
 - IMPERFECTIVE – ongoing, habitual, or repeated situation in the past. Ex. *O homem procurava uma saída do banco.*
 - PERFECTIVE-PROGRESSIVE – combination of the meanings of progressive and perfective. Ex. *O João esteve a ler um livro.*
 - IMPERFECTIVE-PROGRESSIVE – combination of the meanings of progressive and imperfective. Ex. *O João estava a ler um livro quando a Maria entrou.*
 - NONE – no aspectual value for non-verbal event-denoting expressions (nouns, adjectives)

- vform:
 - NONE- when the verb form is finite, or when it does not apply (noun, adjective)
 - GERUNDIVE
 - INFINITIVE
 - PARTICIPLE

- mood:
 - NONE – verb forms in the Indicative or no mood value. Ex. *O João comprou um carro.*
 - SUBJUNCTIVE – it signals the subjunctive mood. Ex. *O João pediu que vás ao mercado.*
 - CONDITIONAL - it signals the conditional mood. Ex. *O João viajaria se pudesse.*
 - FUTURE - it signals the future mood. Ex. *O João já estará em casa agora.*
 - IMPERATIVE - it signals the imperative mood. Ex. *Traz-me um copo de água!*

- modality – attributes values of epistemic and deontic modality to situations represented by the following modal verbs:
 - DEVER
 - PODER
 - TER DE
 - SER CAPAZ DE
 - NONE

- polarity:
 - NEG
 - POS

- movement¹:
 - MOTION_LITERAL: the participant changes her/his physical location during the course of the event (throughout a time interval). Ex. *O rapaz subiu até ao topo da colina.*
 - MOTION_FICTIVE: the eventuality describes a spatial property of a participant (no change throughout a time interval - neither in its location, nor in its spatial properties). Ex. *A montanha eleva-se acima do vale.*

¹ All tags “MOTION_” require a markable that is a verb of movement.

- MOTION_INTRINSIC-CHANGE: the eventuality describes a change in an intrinsic spatial characteristic of the participant) (height, width, length or shape). Ex. *O rio subiu acima do dique.*
- NON_MOTION (non-movement verbs)

A.3. Times

Definition

- Temporal expressions that represent time units classification: t<day (*duas horas, manhã,...*), t=day (*quarta-feira, 2 de janeiro, dia de Natal*), dia<t<year (*mês, semestre, Janeiro, verão,...*), t=year (*2021, próximo ano*), t>year (*século passado*)

Tag span

- noun phrase (*o verão passado, tarde, ...*).

Notes:

- whenever the temporal expression is preceded by a preposition, it will not be part of the tag.
- whenever there is a clausal modifier after the temporal clause, the tag will only include the temporal expression (ex. *cinco dias depois de ter saído*: <time>*cinco dias* <event> *ter saído*)

- adverbial phrase (*ontem, recentemente,..*)

Notes:

- two temporal expressions in a specification relation if (i) they belong to the same class, that is, the same time unit type (*terça-feira, dia 21*), and (ii) they belong to the same syntactic constituent.
- two temporal expressions in an anchoring relation only if denoting daytime (*dez minutos para as 15 horas*). But expressions like *duas semanas a partir da próxima semana* should be marked up with tags <time>*duas semanas*<time>*próxima semana* and a temporal link is established between them.
- two temporal expressions in a conjunction/ disjunction relation are always marked with two different tags. Ex. *seis meses ou um ano*

Non-markables

- Prepositions and subordinating conjunctions (*quando, enquanto, durante*)
- Sequencing and ordering expressions (*entretanto, antes*)
- Expressions of non-quantifiable duration (*permanentemente, temporariamente*)
- References to non-existent times (*nenhum data, nenhum tempo*)

- Time when it means eventuality or occasion (*Este tempo é complicado*)
- Temporal expressions used as proper names (*25 de Abril*)
- Bare frequency (*uma vez, frequente, raramente, nunca, usualmente*)

Attributes

- type:
 - DATE – calendar time (*ontem, 21 de março, o verão de 2000*)
 - TIME – time of the day (*cinco horas, a manhã de 21 de março, ontem à noite*)
 - DURATION – duration (*dois meses, 3 horas, toda a noite*)
 - SET – set of times (*duas vezes por semana, cada dois dias*)
- value: it specifies the value according to the type. It must be written in the notepad. Ex. *4 p.m. value=T16:00; durante dois meses value=P2M*
 Note: Use freq and quant to specify sets. Ex. *Duas vezes por semana value=PIW freq=2X; diariamente value=1D quant=EVERY*
- FunctonInDocument: indicates the function of a time expression in providing a temporal anchor for other temporal expressions in the document.
 - PUBLICATION_TIME – the time the text is published.
- anchortime: it specifies the time that serves as reference point. Ex. *O ladrão assaltou o banco no dia 23 de março. Nesse dia, agrediu ainda uma funcionária de um café.* (o dia 23 must be marked as the AnchorTime of *nesse dia*)

B. Link structures

B.2. Temporal link

Definition

- establishes different temporal relationships (simultaneous, after, before) between two events, two times, or between an event and a time.
- Directionally:
 - o Two events/ two temporal expressions: from the last to the previous one in the linear order of discourse.
 - o Event + temporal expression: from event to temporal expression (irrespective of linear order).

Attributes

- BEFORE – a situation/ time expression occurs before another one. Ex. *A polícia investigou os crimes que tinham ocorrido nos anos 90. (ocorrido – BEFORE – investigou).*
- AFTER – a situation/ time expression occurs after another one. Ex. *Se ele partir agora, chegará a tempo. (chegará – AFTER – partir).*
- INCLUDES – a situation/ time expression includes another one. Ex. *O João saiu quando a Maria dormia. (dormia – INCLUDES – saiu).*
- IS_INCLUDED – a situation/ time expression is included in another one. Ex. *O João chegou ontem. (chegou – IS_INCLUDED – ontem).*
- DURING – a situation/ time expression occurs during another one. Ex. *O João nadou durante 2 horas. (nadou – DURING – 2 horas).*
- SIMULTANEOUS – a situation/ time expression occurs at the same time as another one. Ex. *Enquanto a Maria cantava, o João tocava piano (tocava – SIMULTANEOUS – cantava).*
- IDENTITY – two situations/ time expressions are the same.
 - o *O João conduziu até Boston. Durante a sua viagem ele comeu um donut.*
 - o Causative constructions: triggered by verbs like *causar, induzir, produzir, conduzir*. Ex. *A chuva causou a inundaç o. chuva/causou-IDENTITY; causou/inundaç o-AFTER.*
 - o Light verb constructions: *dar um passeio – IDENTITY*
 - o Reporting events that introduce the story being told in the news. Ex. *O polícia disse que a v tima foi transportada para o hospital. O polícia adiantou que ela estava a recuperar bem dos ferimentos. (disse/adiantou-IDENTITY).*
- BEGINS – a situation/ time expression begins another one.
- ENDS – a situation/ time expression ends another one.

- BEGUN_BY – a situation/ time expression is begun by another one. Ex. *O João trabalhou na escola desde 1994.*
- ENDED_BY – a situation/ time expression is ended by another one. Ex. *O João trabalhou na escola até 2000.*

Publications using the annotation framework

Silvano, Purificação; Leal, António; Silva, Fátima; Cantante, Inês; Oliveira, Fátima & Jorge, Alípio (2021). Developing a multilayer semantic annotation scheme based on ISO standards for the visualization of a newswire corpus. In: *Proceedings of The Seventeenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation*. Groningen, Netherlands.

Leal, António; Silvano, Purificação; Amorim, Evelin; Cantante, Inês; Silva, Fátima; Jorge, Alípio & Campos, Ricardo (2022). "The place of ISO-Space in Text2Story multilayer annotation scheme". In *Proceedings of The Eighteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Marseille, France. ACL Anthology.

Silvano, Purificação; Amorim, Evelin; Leal, António; Cantante, Inês; Silva, Fátima; Jorge, Alípio; Campos, Ricardo & Nunes, Sérgio (2023). "Annotation and visualisation of reporting events in textual narratives". In Ricardo Campos, Alípio Jorge, Adam Jatowt, Sumit Bhatia, Marina Litvak (Eds.). *Proceedings of Text2Story'21 - Sixth Workshop on Narrative Extraction from Texts held in conjunction with the 45th European Conference on Information Retrieval* (47-62). CEUR-WS.org. Vol - 3370.

Silvano, Purificação, Jorge, Alípio, Leal, António, Amorim, Evelin, Sousa, Hugo, Cantante, Inês, Campos, Ricardo & Nunes, Sérgio (2023). "Text2Story Lusa Annotated Corpus [Data set]. INESC TEC. <https://doi.org/10.25747/ESFS-1P1621>

Sousa, Hugo, Guimarães, Nuno, Jorge, Alípio & Campos, Ricardo (2023). GPT Struct Me: Probing GPT Models on Narrative Entity Extraction. In *Proceedings of the 22nd IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WIC'23)*. Venice, Italy. October 26-29.

Silvano, Purificação; Leal, António; Amorim, Evelin; Cantante, Inês; Jorge, Alípio M.; Campos, Ricardo; Yu, Nana. (2024). "Untangling a web of temporal relations in news articles". In Ricardo Campos, Alípio Jorge, Adam Jatowt, Sumit Bhatia, Marina Litvak (Eds.). *Proceedings of Text2Story - Seventh Workshop on Narrative Extraction From Texts held in conjunction with the 46th European Conference on Information Retrieval (ECIR 2024)* (77-92). Glasgow, Scotland, UK.