

The place of ISO-Space in Text2Story multilayer annotation scheme

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

Main objectives:

- present the process of integrating ISO-Space in the Text2Story annotation scheme, which uses ISO 24617-1/4/9;
- assess the compatibility of the different layers (temporal, referential, thematic and spatial);
- pinpoint the problems posed by the harmonization of all layers and by some specifications of ISO-Space.

Text2Story (T2S) project:

- aims to extract narratives from news, represent them in intermediate data structures, and make these available to subsequent media production processes;
- built a multilayer semantic annotation scheme that combines and harmonizes three parts of ISO 24617-1/4/9 (cf. Silvano *et al.*, 2021);
- enables the annotation of an intricate network of relationships, because the different entity structures can be related among them by the different types of link structures.

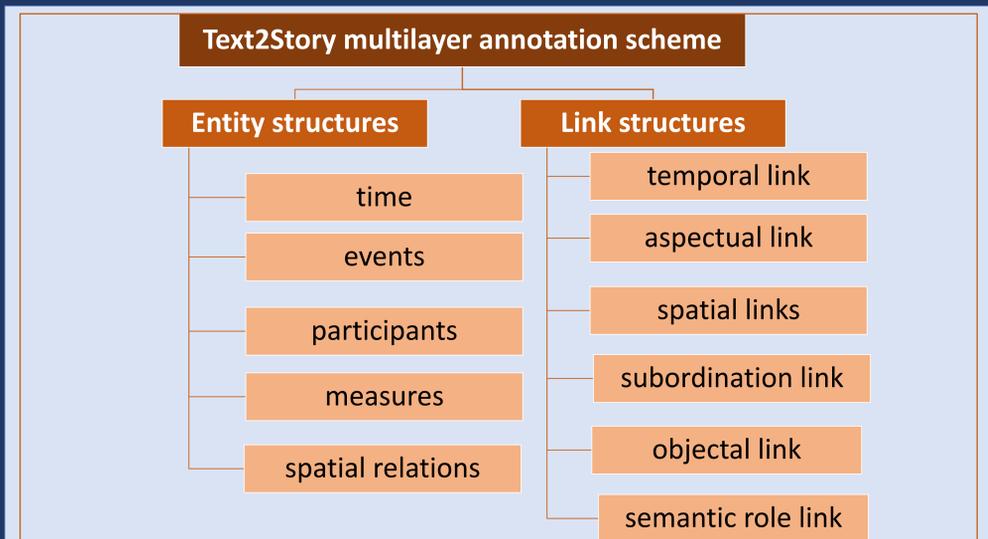
Recently, we moved forward and added a spatial layer using ISO 24617-7 to the annotation scheme.

ISO – Space:

- in accordance with the general principles of ISO 24617-6 (ISO-24617-6, 2016), proposes the use of a set of entity structures, and a set of link structures, with attributes and values, which are employed to annotate spatial information:
 - ✓ **entity structures:** participants, eventualities, and measures;
 - ✓ **link structures:** qualitative spatial links, orientational links, movement links, and measure links.

From ISO 24617-7 to Text2Story annotation scheme

- From ISO 24617-7, we chose the following tags, attributes and links:
 - ✓ **for entities:** place, path, nonlocational spatial entity, spatial relation, motion and non-motional eventualities and measure (leaving out event-path);
 - ✓ **for links:** qualitative spatial link, movement link, and measure link, leaving out orientational link.
- 3 ways of combining tags from ISO 24617-7 with Text2Story annotation scheme (keeping the annotation as simple as possible and fulfilling the project’s purposes):
 - adapting the pre-existing structures** (for participants and events);
 - adding new participant structures:** spatial relations and measures;
 - adding new link structures:** qualitative spatial link, movement link, and measure link.



Some benefits of adding spatial level to T2S annotation scheme:

- possibility of expressing different relations that were previously included in the semantic role of *finalLocation* as *end* or *goalDefining*, which enables different entailments concerning the Figure’s location at the end of the event;
- measure structures and links can now be used not only to link a participant to an event (replacing the semantic role link *amount*), but also to describe participants.

Major problem of adding spatial level to T2S annotation scheme:

- incompatibility between the existence of “non-consuming” tags in ISO 24617-7, namely *event-paths*, and our project’s principle of annotating only lexical material occurring in text (also a principle of parts 1, 4 and 9 of ISO 24617)



- Our solution - since event-paths triggers are motion verbs:
 - motion verbs are used as markables to build event structures;
 - movement links are associated to those event structures.

Future work:

- validation of T2S annotation scheme by using inter-annotator agreement and by generating different types of visualizations from narratives;
- insertion of more spatial information to enable the mapping of the locations referred in the texts to maps;
- study of the specificities of European Portuguese regarding the expression of spatial information and assessment to what extent ISO-Space can account for its representation;
- expansion of the initial corpus of T2S project to encompass texts that include a large array of spatial relations (e.g., texts depicting places, or travel descriptions), which will allow us to test all potentialities of ISO-Space.

References

- Amorim, E., Ribeiro, A., Cantante, I., Jorge, A., Santana, B., Nunes, S., Silvano, M. d. P., Leal, A., and Campos, R. (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). 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-24617-6. (2016). Language resource management- semantic annotation framework (semaf) - part 6: principles of semantic annotation (semaf principles). 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.

This research is funded by the ERDF – European Regional Development Fund through the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project PTDC/CCI-COM/31857/2017 (NORTE-01-0145-FEDER-03185) and within the project UIDB/00022/2020.