Transfer of ISOSpace into a 3D Environment for Annotations and Applications

Alexander Henlein, Giuseppe Abrami, Attila Kett, Alexander Mehler
Goethe University
Frankfurt am Main 60323, Germany
{henlein, abrami, mehler}@em.uni-frankfurt.de, attila.kett@stud.uni-frankfurt.de

Abstract

People's visual perception is very pronounced and therefore it is usually no problem for them to describe the space around them in words. Conversely, people also have no problems imagining a concept of a described space. In recent years many efforts have been made to develop a linguistic scheme for spatial and spatial-temporal relations. However, the systems have not really caught on so far, which in our opinion is due to the complex models on which they are based and the lack of available training data and automated taggers. In this paper we describe a project to support spatial annotation, which could facilitate annotation by its many functions, but also enrich it with many more information. This is to be achieved by an extension by means of a VR environment, with which spatial relations can be better visualized and connected with real objects. And we want to use the available data to develop a new state-of-the-art tagger and thus lay the foundation for future systems such as improved text understanding for Text2Scene Generation.

Keywords: ISOSpace, ISOTimeML, Unity3D, Annotation, Virtual Reality

1. Introduction

Humans have a strong spatial perception. This is reflected not only in how well people can adapt to new spatial environments, but also in their language (Haun et al., 2011).

In recent years there have been increased efforts to create a linguistic model for these spatial references. This led to new linguistic models, like ISOSpace (ISO, 2014a) and SceneML (Gaizauskas and Alrashid, 2019) and new tasks, such as Spatial Role Labeling (Kordjamshidi et al., 2010) or SpaceEval (Pustejovsky et al., 2015). Nevertheless, these annotation schemes have not really been able to establish themselves in applications so far. This could be due to the models' complexity, the availability of annotated training data and the lack of automated taggers. There were indeed approaches to apply such models to image descriptions (Pustejovsky and Yocum, 2014), but to our knowledge there were no efforts to transfer the corresponding annotation schemes into three-dimensionality. For the latter, the language model would be particularly interesting, for example, to reconstruct scenes from speech and text three-dimensionally.

In this paper we present our project plan on a 3D VR framework that addresses the problems mentioned above and offers a direct application. In Section 2 we describe the models and systems we refer to in our project, and in Section 3 we explain how we build on these models to create a framework that supports both annotation and application of these language models.

2. Related Work

In recent years, much work has been spent on the development of linguistic models for the semantic understanding of language. The largest of these is probably the Semantic Annotation Framework (SemAF), published under ISO/TC 37/SC 4/WG 2 Semantic Annotation. This consists of individual modules that relate to specific semantic units and are compatible with each other (Ide and Pustejovsky, 2017, Chapter 4). The most widespread model of SemAF is ISO-TimeML (Pustejovsky et al., 2010; ISO, 2012a), a scheme for the annotation of time and time dependencies of events based on TimeML (Pustejovsky et al., 2005). Such dependencies are important for text understanding, because without them text contents can hardly be fully understood (Ide and Pustejovsky, 2017, p. 942).

There is also a model that focuses more on spatial and spatial-temporal structures, the ISOSpace (Pustejovsky et al., 2011; ISO, 2014a). The focus is on spatial and spatial-temporal relations between (spatial) entities and the connection via motion events. Spatial Entities are marked and connected to each other via different spatial connections. QSLinks (Qualitative Spatial Links) are for topological relations, OLinks (Orientation Links) for non-topological relations and MoveLinks for movements of entities in space. This scheme was the basis of SpaceEval (Pustejovsky et al., 2015) and was successfully applied to image descriptions to differentiate between content and structural statements (Pustejovsky and Yocum, 2014). ISOSpace in particular is being further improved (ISO, 2019) and serves as a basis for more specialized models, such as SceneML (Gaizauskas and Alrashid, 2019) for scene descriptions. In addition, SemAF contains schemata such as Semantic Roles (ISO, 2014b), Dialog Acts (ISO, 2012b) and other modules are under development, e.g. QuantML (Bunt et al., 2018).

As the requirements for the annotation of text contexts are constantly changing, flexible and dynamic annotation environments are required to enable the efficient annotation of complex situations. This challenge is addressed by TEXTANNOTATOR (Abrami et al., 2019), a browser-based and therefore platform-independent annotation tool for collaborative multi-modal annotation of texts. Using TEXTANNOTATOR, NER annotations can be created in texts in a short execution time as well as the annotation of rhetorical (Helfrich et al., 2018), time, propositional and even argument structures can be graphically visualised and executed. Furthermore, texts can be linked to ontological resources (e.g.

His [**room**]$_{p1}$, a proper [**room**]$_{p1}$ for a human being, only somewhat too small, lay quietly [**between**]$_{ss1}$ the four well-known [**walls**]$_{se1}$. [**Above**]$_{ss2}$ the [**table**]$_{se2}$, [**on**]$_{ss3}$ which an unpacked collection of [**sample cloth goods**]$_{se3}$ was spread out, hung the [**picture**]$_{se4}$ which he had [**cut out**]$_{m1}$ of an illustrated [**magazine**]$_{se6}$ a little while ago and [**set in**]$_{m2}$ a pretty gilt [**frame**]$_{se7}$.

QSLINK(p1, se1, ss1, between)
QSLINK(se3, se2, ss3, EC)
OLINK(se3, se2, ss3, above)
OLINK(se4, se2, ss2, above)
MOVELINK(m1, se4, se6, se4)
MOVELINK(m2, se4, se4, se7)

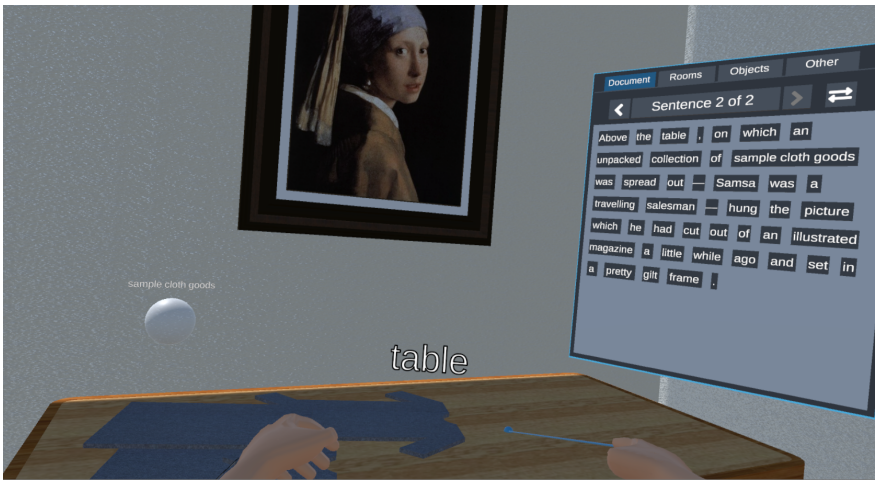

Figure 1: On the left side a (simplified) annotation of an abridged section of Kafka's: The Metamorphosis according to the ISOSpace (2014) scheme. On the right side a 3D representation. Each entity in the text is linked to the corresponding 3D object from ShapeNetSem and we linked the two clothing to one object group. The relationship between the table and the room is not explicitly mentioned, but is implied by the placement of the table in the room.
p: place, *se*: spatial entity, *ss*: spatial signal, *m*: move event.
QS/OLINK(*figure, ground, signal, relation*). MOVELINK(*move, mover, source, goal*).

Wikipedia, Wikidata, Wiktionary) and the annotations are managed in different annotation views based on user and group-based permissions (Gleim et al., 2012). As a result, TEXTANNOTATOR is capable of creating a real-time calculation of an inter-annotator agreement based on classes defined in the annotation task (Abrami et al., 2020b).
Since humans are spatially anchored not only in their actions and perception but also in their linguistic behavior (Bateman, 2010; Bateman et al., 2010), this led to new efforts to spatially translate annotations by means of virtual reality. One of these projects is VANNOTATOR (Spiekermann et al., 2018), a system for the annotation of linguistic and multi-modal information units, implemented in Unity3D[1]. VANNOTATOR is a platform for use in various scenarios such as visualization and interaction with historical information (Abrami et al., 2020a) or the annotation of texts and the linking of texts and images with 3D objects (Mehler et al., 2018). Since VANNOTATOR integrates TEXTANNOTATOR and thus makes the annotation spectrum of the latter available in VR, annotations in VANNOTATOR can be performed collaboratively (in workgroups) as well as simultaneously.

3. Our Current Project

ISOSpace is a very expressive model, but its complexity makes it difficult to use it as a basis for annotation. Work is not made easier when 3D information is annotated on a 2D surface. This becomes particularly clear in the annotation of spatial relations between entities, where, e.g., in the case of SpaceEval data, the inter-annotator agreement was only 33% for QSLinks and 39% (Pustejovsky et al., 2015) for OLinks. These are hardly values that guarantee high data quality. Here an extended visualization, as our project aims at, could significantly support these annotation tasks.

[1] https://unity.com/

To this end, our aim is to integrate ISOSpace and other SemAF models such as ISOTimeML into TEXTANNOTATOR. Since TEXTANNOTATOR is based on UIMA (*Unstructured Information Management Applications*) (Ferrucci and Lally, 2004), its annotation schemes are defined as UIMA TYPE SYSTEM DESCRIPTORS (TSD). Before the ISO models can be used in UIMA, they have to be transferred to TSD. This is the first step towards collaborative annotation in a visually supporting interface. The annotation can then be enriched by TEXTANNOTATOR embedded into VANNOTATOR. This enables spatial annotations with a 3D interface in VR. In addition, spatial entities can be directly linked to 3D objects via a large number of categorized objects from ShapeNet (Chang et al., 2015), the slightly deeper annotated objects from ShapeNetSem (Savva et al., 2015), objects annotated using VoxML notation (Pustejovsky and Krishnaswamy, 2016) (under development) or via abstract representations (as exemplified in Figure **??**).

Simply by placing the objects in space, conclusions can be drawn about the relationships between them (and thus also about QSLinks and OLinks) because the information bandwidth of annotation acts in VR is much larger than with pure text annotation. For example, if a book is placed on the desk in VR, the corresponding QSLink and OLink can be set automatically with their relevant attributes. Such concrete pictorial representations are not always unambiguous, but in conjunction with the corresponding sentence, classifiers can be trained to solve this (Hürlimann and Bos, 2016). This can also be extended to MoveLinks, which are set automatically when, for example, the book is carried through the room and placed on a shelf. Or the annotator can follow a direction described in the text in the VR environment. Such actions are much more natural and easier for humans to perform than abstract annotations in a 2D display. Missing links can thus be more easily identified and in some

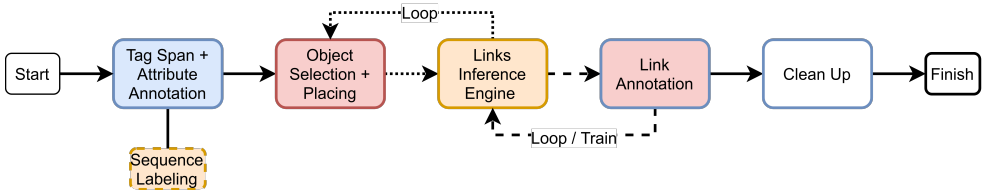

Figure 2: Workflow for ISOSpace Annotation. Blue borders stand for the original annotation steps (Pustejovsky et al., 2015). Red filled for VR support and orange for machine learning support. Span tagging can be supported with a sequence labeling system. And the link inference engine learns through annotations.

cases automatically predicted and attributed, e.g., by examining transitive relations. Such support has also been successfully applied to the annotation of the TimeML standard (Setzer et al., 2005; Verhagen et al., 2006; Verhagen, 2007). The underlying workflow is shown in Figure 2.

A central challenge will be the underspecification of scene descriptions. Related issues concern descriptions containing negations. Though we do not yet have a solution to solve the problems involved, we assume that by combining spatial experience in VR with annotation services provided by annotators, for example, underspecified reference relations can be annotated by exploring additional information with regard to the annotators' positions in relation to referred objects. In examples such as "There is no book on the table" a corresponding book object can be highlighted to indicate the negation (as done, e.g., in WordsEye (Coyne and Sproat, 2001)). In the case of underspecified relations, as expressed in examples of the sort of "The pencil is next to the book", there is the possibility of assigning relative or variable positions to objects (so that they take up tipping states in the visualization).

The next step is the stepwise extension of our annotation system by further (e.g. ISOTimeML) and future (e.g. QuantML (Bunt et al., 2018)) SemAF modules. In this way we create a multi-modal, virtualized annotation system capable of mapping text to abstract or concrete spatial representations of a very broad complexity.

The available ISOSpace data will then be used to develop and train taggers that automatically perform or largely support this annotation. The taggers can support annotators with annotation suggestions, which the annotators then only have to accept or minimally correct.

TEXTANNOTATOR is already actively used for annotating historical text data in the BIOfid project[2]. These annotations (Ahmed et al., 2019) will be extended in the near future to include ISOSpace, ISOTimeML, SemAF-SR and probably also QuantML.

Such in-depth annotations could form the still missing basis for Text2Scene systems (Coyne and Sproat, 2001), which in turn should be able to provide a much deeper understanding of spatial language than previous systems that focus primarily on key words (e.g. (Chang et al., 2017; Ma et al., 2018). Application areas could be, for example: Reconstructing events from multiple texts (based on Twitter, news reports, etc.), visualizing descriptions of accidents (Johansson et al., 2005) or crime scenes or 3D visualizations of text content to clarify certain relations (e.g. intersections of biographical life paths).

This could also help to identify weaknesses of the ISOSpace model, such as missing information relevant for spatial annotation. A problem that could occur is that RCC (Region Connection Calculus) (Randell et al., 1992) for representing topological relations of regions is not sufficient to represent 3D spaces. One reason is that it does not refer to a specific dimension (Renz, 2002).

4. Conclusion

We argued that ISOSpace, despite its expressiveness, has not yet reached the application density that is essential to provide training data for tools for automatically annotating spatial language. To fill this gap, we plan to integrate ISOSpace into VANNOTATOR to enable 3D annotations of spatial language. This will also include other SemAF models in order to ultimately provide the data basis for the creation of Text2Scene systems.

5. Acknowledgements

Many thanks to all reviewers for their comments, suggestions, hints and references. These were very helpful and we will incorporate much of this in our future work.

6. Bibliographical References

Abrami, G., Mehler, A., Lücking, A., Rieb, E., and Helfrich, P. (2019). TextAnnotator: A flexible framework for semantic annotations. In *Proc. of ISA-15*, May.

Abrami, G., Mehler, A., Spiekermann, C., Kett, A., Lööck, S., and Schwarz, L. (2020a). Educational technologies in the area of ubiquitous historical computing in virtual reality: Finding new ways to teach in a transformed learning environment. In Linda Daniela, editor, *New Perspectives on Virtual and Augmented Reality*. Taylor & Francis. in press.

Abrami, G., Stoeckel, M., and Mehler, A. (2020b). TextAnnotator: A UIMA based tool for simultaneous and collaborative annotation of texts. In *Proc. of LREC 2020*, LREC 2020. accepted.

Ahmed, S., Stoeckel, M., Driller, C., Pachzelt, A., and Mehler, A. (2019). BIOfid Dataset: Publishing a german gold standard for named entity recognition in historical biodiversity literature. In *Proc. of CoNLL 2019*.

Bateman, J. A., Hois, J., Ross, R., and Tenbrink, T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14):1027–1071.

Bateman, J. A. (2010). Language and space: A two-level semantic approach based on principles of ontological engineering. *Int J Speech Tech*, (1):29–48.

[2] https://www.biofid.de/en/

Bunt, H., Pustejovsky, J., and Lee, K. (2018). Towards an ISO standard for the annotation of quantification. In *Proc. of LREC 2018*.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., and Yu, F. (2015). ShapeNet: An information-rich 3D model repository. Technical Report arXiv:1512.03012 [cs.GR].

Chang, A. X., Eric, M., Savva, M., and Manning, C. D. (2017). SceneSeer: 3D scene design with natural language. *arXiv preprint arXiv:1703.00050*.

Coyne, B. and Sproat, R. (2001). WordsEye: An automatic text-to-scene conversion system. In *Proc. of SIGGRAPH 2001*, pages 487–496.

Ferrucci, D. and Lally, A. (2004). UIMA: An architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, (3-4):327–348.

Gaizauskas, R. and Alrashid, T. (2019). SceneML: A proposal for annotating scenes in narrative text. In *Workshop on ISA-15*, page 13.

Gleim, R., Mehler, A., and Ernst, A. (2012). SOA implementation of the eHumanities Desktop. In *Proc. of the Workshop on SOAs for the Humanities: Solutions and Impacts, Digital Humanities 2012*.

Haun, D. B., Rapold, C. J., Janzen, G., and Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, (1):70–80.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., and Mehler, A. (2018). TreeAnnotator: Versatile visual annotation of hierarchical text relations. In *Proc. of LREC 2018*.

Hürlimann, M. and Bos, J. (2016). Combining lexical and spatial knowledge to predict spatial relations between objects in images. In *Proc. of CVPR 2016*, pages 10–18.

Ide, N. and Pustejovsky, J. (2017). *Handbook of linguistic annotation*. Springer.

ISO. (2012a). Language resource management — Semantic annotation framework (SemAF) — Part 1: Time and events (SemAF-Time, ISO-TimeML). Standard ISO/IEC TR 24617-1:2012.

ISO. (2012b). Language resource management — Semantic annotation framework — Part 2: Dialogue acts. Standard ISO/IEC TR 24617-2:2012.

ISO. (2014a). Language resource management — Semantic annotation framework (SemAF) — Part 7: Spatial information (ISO-Space). Standard ISO/IEC TR 24617-7:2014.

ISO. (2014b). Language resource management — Semantic annotation framework — Part 4: Semantic roles (SemAF-SR). Standard ISO/IEC TR 24617-4:2014.

ISO. (2019). Language resource management — Semantic annotation framework (SemAF) — Part 7: Spatial information (ISO-Space). Standard ISO/IEC TR 24617-7:2019.

Johansson, R., Berglund, A., Danielsson, M., and Nugues, P. (2005). Automatic text-to-scene conversion in the traffic accident domain. In *IJCAI*, volume 5, pages 1073–1078.

Kordjamshidi, P., Moens, M.-F., and van Otterlo, M. (2010). Spatial Role Labeling: Task definition and annotation scheme. In *Proc. of LREC 2010*, pages 413–420.

Ma, R., Patil, A. G., Fisher, M., Li, M., Pirk, S., Hua, B.-S., Yeung, S.-K., Tong, X., Guibas, L., and Zhang, H. (2018). Language-driven synthesis of 3D scenes from scene databases. In *SIGGRAPH Asia 2018 Technical Papers*, page 212. ACM.

Mehler, A., Abrami, G., Spiekermann, C., and Jostock, M. (2018). VAnnotatoR: A framework for generating multimodal hypertexts. In *Proc. of HT 2018*.

Pustejovsky, J. and Krishnaswamy, N. (2016). VoxML: A visualization modeling language. *arXiv preprint arXiv:1610.01508*.

Pustejovsky, J. and Yocum, Z. (2014). Image annotation with ISO-Space: Distinguishing content from structure. In *Proc. of LREC 2014*, pages 426–431. ELRA.

Pustejovsky, J., Ingria, B., Sauri, R., Castano, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G., and Mani, I. (2005). The specification language TimeML. *The language of time: A reader*, pages 545–557.

Pustejovsky, J., Lee, K., Bunt, H., and Romary, L. (2010). ISO-TimeML: An international standard for semantic annotation. In *Proc. of LREC 2010*.

Pustejovsky, J., Moszkowicz, J. L., and Verhagen, M. (2011). ISO-Space: The annotation of spatial information in language. In *Proc. of the Sixth Joint ISO-ACL SIGSEM Workshop on ISA*, pages 1–9.

Pustejovsky, J., Kordjamshidi, P., Moens, M.-F., Levine, A., Dworman, S., and Yocum, Z. (2015). SemEval-2015 Task 8: SpaceEval. In *Proc. of SemEval 2015*, pages 884–894.

Randell, D. A., Cui, Z., and Cohn, A. G. (1992). A spatial logic based on regions and connection. *KR*, pages 165–176.

Renz, J. (2002). A canonical model of the region connection calculus. *Journal of Applied Non-Classical Logics*, (3-4):469–494.

Savva, M., Chang, A. X., and Hanrahan, P. (2015). Semantically-enriched 3D models for common-sense knowledge. *CVPR 2015 Workshop on Functionality, Physics, Intentionality and Causality*.

Setzer, A., Gaizauskas, R., and Hepple, M. (2005). The role of inference in the temporal annotation and analysis of text. *Language Resources and Evaluation*, 39(2-3):243–265.

Spiekermann, C., Abrami, G., and Mehler, A. (2018). VAnnotatoR: A gesture-driven annotation framework for linguistic and multimodal annotation. In *Proc. of AREA 2018 Workshop*, AREA.

Verhagen, M., Knippen, R., Mani, I., and Pustejovsky, J. (2006). Annotation of temporal relations with Tango. In *LREC*, pages 2249–2252.

Verhagen, M. (2007). Drawing TimeML relations with TBox. In *Annotating, Extracting and Reasoning about Time and Events*, pages 7–28. Springer.