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 \textit{room}, a proper \textit{room} for a human being, only somewhat too small, lay quietly \textit{between} the four well-known \textit{walls}, \textit{[Above]} the \textit{table} \textit{on} which he had \textit{cut out} which he had \textit{picture} was spread out, hung the \textit{picture} which he had \textit{cut out} \textit{m} of an illustrated \textit{magazine} \textit{a little} while ago and \textit{[set in]} \textit{a pretty gilt} \textit{frame}.

\text{QSLINK(p1, se1, ss1, between)}
\text{QSLINK(se3, se2, ss3, EC)}
\text{OLINK(se3, se2, ss3, above)}
\text{OLINK(se4, se2, ss2, above)}
\text{MOVELINK(m1, se4, se6, se4)}
\text{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.

\text{p: place, se: spatial entity, ss: spatial signal, m: move event.}
\text{QSLINK(f: figure, g: ground, s: signal, r: relation). MOVELINK(m: move, s: source, g: goal).}

Wikipedia, Wikidata, Wiktionary) and the annotations are managed in different annotation views based on user and group-based permissions \cite{Gleim2012}. As a result, \textsc{textannotator} is capable of creating a real-time calculation of an inter-annotator agreement based on classes defined in the annotation task \cite{Abrami2020b}. Since humans are spatially anchored not only in their actions and perception but also in their linguistic behavior \cite{Bateman2010, Bateman2010a}, this led to new efforts to spatially translate annotations by means of virtual reality. One of these projects is \textsc{vannotator} \cite{Spiekermann2018}, a system for the annotation of linguistic and multi-modal information units, implemented in Unity3D. \textsc{Vannotator} is a platform for use in various scenarios such as visualization and interaction with historical information \cite{Abrami2020a} or the annotation of texts and the linking of texts with 3D objects \cite{Mehler2018}. Since \textsc{Vannotator} integrates \textsc{textannotator} and thus makes the annotation spectrum of the latter available in VR, annotations in \textsc{Vannotator} can be performed collaboratively (in workgroups) as well as simultaneously.

### 3. Our Current Project

\textsc{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\% \cite{Pustejovsky2015} 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.

To this end, our aim is to integrate \textsc{ISOSpace} and other \textsc{SemAF} models such as \textsc{ISOTimeML} into \textsc{Textannotator}. Since \textsc{Textannotator} is based on UIMA (\textit{Unstructured Information Management Applications}) \cite{Ferrucci2004}, its annotation schemes are defined as UIMA \textsc{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 \textsc{Textannotator} embedded into \textsc{V-annotator}. 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 \cite{Chang2013}, the slightly deeper annotated objects from ShapeNetSem \cite{Savva2015}, objects annotated using VoxML notation \cite{Pustejovsky2016} (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 band-width 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 \cite{Hueflermann2016}. 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 cases...
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: A flexible framework for semantic annotations. In Proc. of ISA-15, May.

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 UIIMa based tool for simultaneous and collaborative annotation of texts. In Proc. of LREC 2020, LREC 2020. accepted.

Ahmed, S., Stoeckel, M., Driller, C., Fachzelt, 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.
