VANNOTATOR: a Gesture-driven Annotation Framework for Linguistic and Multimodal Annotation

Christian Spiekermann, Giuseppe Abrami, Alexander Mehler
Text Technology Lab
Goethe-University Frankfurt
s2717197@stud.uni-frankfurt.de, {abrami,amehler}@em.uni-frankfurt.de

Abstract

Annotation processes in the field of computational linguistics and digital humanities are usually carried out using two-dimensional tools, whether web-based or not. They allow users to add annotations on a desktop using the familiar keyboard and mouse interfaces. This imposes limitations on the way annotation objects are manipulated and interrelated. To overcome these limitations and to draw on gestures and body movements as triggering actions of the annotation process, we introduce VANNOTATOR, a virtual system for annotating linguistic and multimodal objects. Based on VR glasses and Unity3D, it allows for annotaing a wide range of homogeneous and heterogeneous relations. We exemplify VANNOTATOR by example of annotating propositional content and carry out a comparative study in which we evaluate VANNOTATOR in relation to WebAnno. Our evaluation shows that action-based annotations of textual and multimodal objects as an alternative to classic 2D tools are within reach.

Keywords: Virtual reality, gesture-driven annotation, multimodal annotation objects

1. Introduction

Annotation processes in the field of computational linguistics and digital humanities are usually carried out using two-dimensional tools, whether web-based or not. They allow users to add annotations on a desktop using the familiar keyboard and mouse interfaces. The visualization of annotations is limited to an annotation area which is delimited by a manageable number of windows. Within a single window, relationships of annotation objects are graphically visualized by connecting them to each other by means of lines as an add-on to the 2D surface. This diagnosis also includes tools for annotating multimodal objects (Cassidy and Schmidt, 2017). Further, most of these tools do not support collaboratively annotating the *same* document simultaneously – though there exist recent developments of collaborative web-based tools (Biemann et al., 2017). Popular frameworks for linguistic annotation such as *Atomic* (Druskat et al., 2014) or *ANNIS* (Chiarcos et al., 2008), respectively, *brat* (Stenetorp et al., 2012) and *WebAnno* (de Castilho et al., 2014) are partly sharing these limitations. *Brat*, for example, is a web-based annotation framework that allows different users to annotate a document simultaneously. All changes are made directly available to all annotators. In contrast, *WebAnno* based on *brat* concentrates on parallel annotations where annotators cannot see changes made by users sharing the same rights. Curators can then compare and verify annotations of different users.

In this paper, we introduce VANNOTATOR, a 3D tool for linguistic annotation to overcome these limits: (1) first and foremost, VANNOTATOR provides a three-dimensional annotation area that allows annotators to orient themselves within 3D scenes containing representations of natural objects (e.g., accessible buildings) and semiotic aggregates (texts, images, etc.) to be annotated or interrelated. (2) A basic principle of annotating by means of VANNOTATOR is to manifest, trigger and control annotations with gestures or body movements. In this way, natural actions (such as pointing or grasping) are evaluated to perform annotation subprocesses. (3) In addition, according to the strict 3D setting of VANNOTATOR, discourse referents are no longer implicitly represented. Thus, unlike WebAnno, where anaphora have to be linked to most recently preceding expressions of identical reference (leading to monomodal line graphs), discourse referents are now represented as manipulable 3D objects that are directly linked to any of their mentions (generating multimodal star graphs connecting textual manifestations and 3D representations of discourse referents). (4) VANNOTATOR allows for collaboratively annotating documents so that different annotators can interact within the same annotation space, whether remotely or not, though not yet simultaneously. (5) The third dimension allows for the simultaneous use of many different tools for annotating a wide variety of multimedia content without affecting clarity. In contrast, 2D interfaces that allow text passages to be linked simultaneously with video segments, positions in 3D models, etc. quickly become confusing. The reason for this is that in the latter case the third dimension cannot be used to represent relations of information objects. In other words, 3D interfaces are not subject to the same loss of information as 2D interfaces when representing relational information.

In this paper, we demonstrate the basic functionality of VANNOTATOR by focusing on its underlying data model, its gestural interface and also present a comparative evaluation in the area of anaphora resolution. The paper is organized as follows: Section 2. gives a short overview of related work in the area of VR (Virtual Reality) based systems. In Section 3. we briefly sketch the architecture of VANNOTATOR and its gestural interface. Section 4. provides a comparative evaluation. Finally, Section 5. gives a conclusion and an outlook on future work.

2. Related Work

Virtual environments have long been popular for visualizing and annotating objects, but not primarily in the NLP

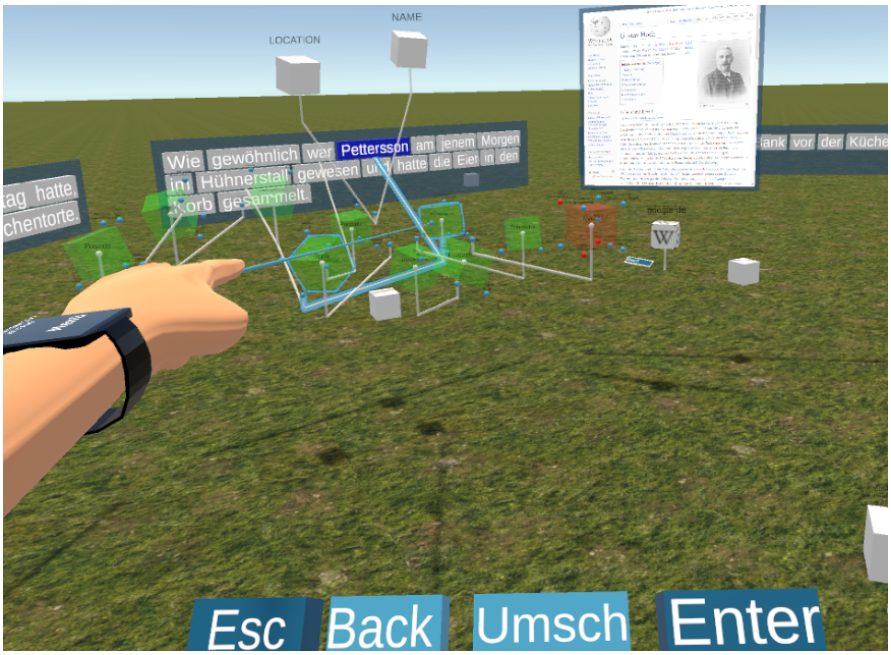

Figure 1: Sentences (blue boxes), tokens (grey), annotation cubes (green: complete annotations, red: incomplete ones, grey: stored annotations) and lines representing relations between annotations. A simple keyboard is visualized at the bottom.

domain. (Bellgardt et al., 2017) describe general usage scenarios of VR systems addressing actions of sitting, standing or walking. (Cliquet et al., 2017) even envision scenarios in which textual aggregates are accompanied with shareable experiences in the virtual reality – a scenario also addressed by VANNOTATOR. Older projects are, for example, *Empire 3D*, a collaborative semantic annotation tool for virtual environments with a focus on architectural history (Abbott et al., 2011). Based on *OpenScreneGraph*, Empire 3D visualizes database-supported information about buildings and locations. Another tool is *Croquet* (Kadobayashi et al., 2005); it allows for modeling and annotating scenes that are finally represented as 3D wikis. *Croquet* is followed by *Open Cobalt*.[1] Closer to the area of NLP is the annotation system of (Clergeaud and Guitton, 2017), a virtual environment that allows for annotating documents using a virtual notepad. Inserting multimedia content is also possible with this system.

To the best of our knowledge, there is currently no framework of linguistic or even multimodal annotation in virtual reality that meets the scenario of VANNOTATOR as described in Section 1.

3. VANNOTATOR

3.1. Annotation Space

Based on Stolperwege (Mehler et al., 2017), which aims to transform processes of documenting historical processes into virtual environments, VANNOTATOR has been designed for desktop systems and therefore supports the most common VR glasses[2] in conjunction with their motion controllers. The underlying environment is Unity3D, which allows for instantiating VANNOTATOR on different platforms.

Initially, VANNOTATOR gives annotators access to empty virtual spaces (work environments) providing flexible areas for visualizing and annotating linguistic and multimedia objects. Figure 1 illustrates the annotation of a text segment (sentence), its tokenization, specification of discourse referents and their relations forming a graphical representation of (phoric) discourse structure. In this example, the annotator has extracted several text segments from the VANNOTATOR browser (in our example displaying a Wikipedia article) and arranged them in circular order. In this way, she or he can move between the segments to annotate them.

The major instrument for interacting with annotation objects are virtual hands (see Figure 1) currently realized by means of the motion controllers. Walking or moving is also performed by means of the controllers. In this way, VANNOTATOR enables teleportation as well as stepless and real movements.

3.2. Data Model, Annotation Scheme and UIMA Database Interface

The integrity of VANNOTATOR-based annotations is evaluated with respect to the data model (see Figure 3) of the *Stolperwege* project. This joint project of historians and computer scientists aims at semi-automatically documenting the biographies of victims of Nazism. To this end, it includes a data model for modeling propositional text content: currently, propositions are modeled as logical expressions of predicate argument structures where arguments manifest semantic roles in the sense of role labeling systems. Arguments (see Figure 3) form a superclass of discourse referents (DR) modeled as virtual representations of persons, times, places or positions and events (being defined as sets of propositions in the sense of situation semantics) as well as multimedia objects (e.g., accessible animations of buildings or images). Beyond that, a DR can

[1] https://sites.google.com/site/opencobaltproject/

[2] Oculus Rift and HTC Vive.

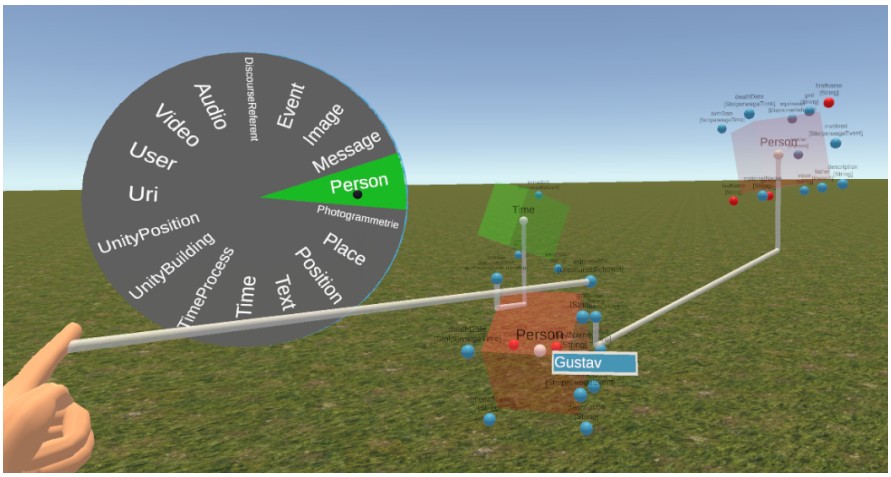

Figure 2: Incompletely annotated DR (red). The menu allows for generating a new DR using the touch gesture and to connect it to other DRs regarding the focal attribute.

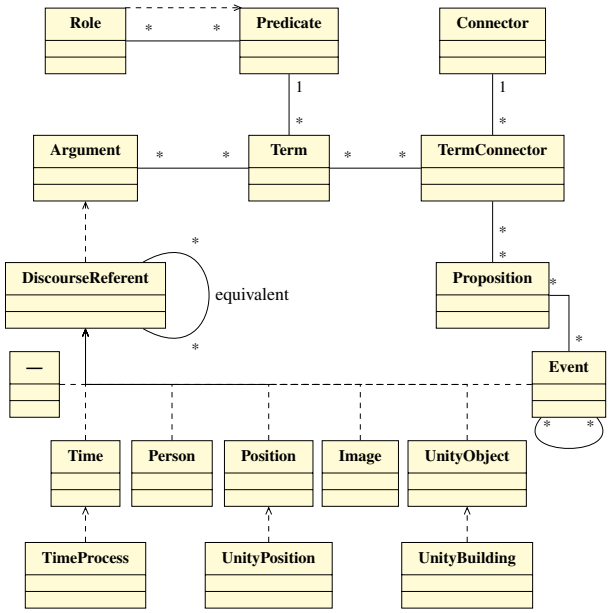

Figure 3: A subset of the data model of VANNOTATOR.

be introduced as an aggregation of more elementary DRs. In this way, for example, a group of persons can be defined as a candidate referent of an anaphoric plural expression. From a graph-theoretical point of view, homogeneous n-ary relations can be annotated as well as hyperedges manifesting heterogeneous relations. When the user introduces a new DR, the system visually represents it as a so-called annotation cube whose annotation slots are defined by the corresponding entity's (intra- or interrelational) attributes. VANNOTATOR supports the annotation process by providing visual feedback in terms of green (complete) and red (incomplete) cubes. In this way, VANNOTATOR can also be seen as virtual interface to relational databases.

We mapped the relational data model of VANNOTATOR onto *UIMA Type System Descriptor* so that the resulting annotation scheme and annotation objects can be managed by means of a UIMA-based database, that is, the so-called *UIMA Database Interface* of (Abrami and Mehler, 2018). The database is accessible through a RESTful web service. Any DR managed in this way can be linked to multimedia content or external information objects (extracted from Wikidata or Wikipedia). Further, DRs can be reused across multiple annotation scenarios including different texts. Each DR is uniquely identifiable via its URI being visualized as a corresponding cube. Any such cube can be manipulated using a range of different gestures.

3.3. Gestural Interface

The annotation process is driven by means of the following gestures:

Grab Pick up and move an element to any position.

Point Teleport to any position in the virtual environment or select a DR.

Touch Touching a DR with the point gesture either initiates the annotation process or establishes a relationship between this source node and a target node to be selected. As a result of this, a line is drawn between both DRs. Touching different tokens with both index fingers creates a text area between them.

Twist Grabbing and rotating a line manifesting a relation of DRs removes it.

Pull apart By means of this gesture, the characteristic action connected to a DR is executed. For a DR of type URI, this means, for example, that a window is opened in VANNOTATOR's browser to display the content of this resource.

Throw over the shoulder This action disables or resets the DR.

We now describe how to select, visualize and annotate text taken from VANNOTATOR's internal browser using these gestures. Note that this browser serves as an interface to introduce additional content, images or URI from outside of VANNOTATOR. To annotate a text, its tokens are typed by mapping them onto an appropriate class of the data model.

To this end, the touch gesture is used to select a corresponding data type using the so-called controller (see the circular menu in Figure 2). Then, a new DR is generated and visualized as a cube. Any such cube has blue slots indicating attributes to be set or relations to other DRs to be generated. Green cubes indicate DRs that can be stored in the database. After being stored, cubes change their color again (gray) to indicate their reusability as persistent database objects (see Figure 5).

4. Evaluation

A comparative evaluation was carried out to compare VANNOTATOR with *WebAnno* (Spiekermann, 2017) by example of anaphora resolution. The test group consisted of 14 subjects and was divided so that one half solved the test with *WebAnno* and the other with VANNOTATOR. Test persons hat to annotate two texts (Task 1 and 2). In task 1, a text was provided with predefined annotations which were to be reconstructed by the test persons. The idea was that they should get to know the respective framework and understand the meaning of the annotation process. For *WebAnno*, we provided the respective text on a large screen. In VANNOTATOR, the sample text was presented at another place within the annotation space. Thus, users had to move between the place displaying the sample and the one where it had to be re-annotated (see Figure 5). In the second task, users needed to annotate all anaphoric relations from scratch. Note that VANNOTATOR can represent anaphoric relations using hyperedges including a DR and all its mentions, while *WebAnno* generates sequences of reference-equal expressions.

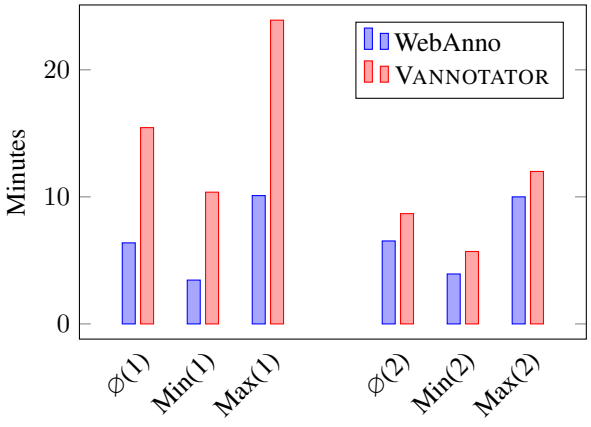

Figure 4: Minimum, maximum and average times (in minutes) for solving the tasks.

Figure 4 shows the average, minimum and maximum time taken by subjects to solve both tasks. It shows that test subjects using VANNOTATOR take on average more than twice as much time for the first text as the second one. However, the annotation time for the second text was almost halved, while it stagnated when using *WebAnno*. The average number of (in-)correctly annotated sections hardly differs between both frameworks.
The lower effort in using *WebAnno* is certainly due to the fact that the subjects used mouse and keyboard daily for years, in contrast to our new interface for which they lacked such experiences. The remaining time-related difference between both frameworks in executing Task 1 is probably due to the higher number of actions currently required by VANNOTATOR and the greater distance in the third dimension to be bridged by annotation actions. In any case of Task 2, the processing time is considerably shortened.
Finally, a UMUX (Finstad, 2010) survey was completed by the subjects. This produces a value in the range of 0 to 100, where 100 indicates an optimal result. *WebAnno* yields 66 points, VANNOTATOR 70. This shows that both frameworks have similarly good user ratings. Since some test persons had little experience in using 3D technologies, we also observed cases of motion sickness. In summary, our evaluation shows that VANNOTATOR provides comparable results to an established tool. VANNOTATOR performs slightly better in UMUX, which is not yet an optimal result, but indicates a potential of annotating in the third dimension.

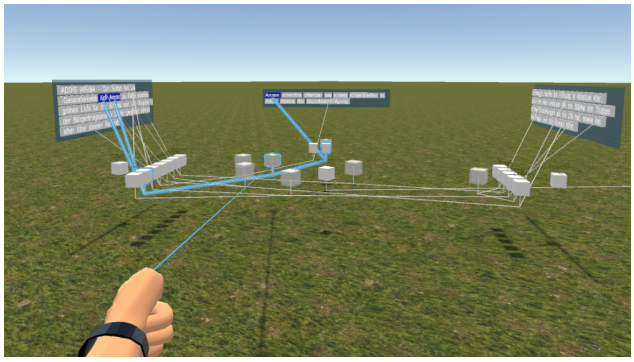

Figure 5: Visualization of an annotated text document.

5. Conclusion & Future Work

We introduced VANNOTATOR, a tool for linguistic and multimodal annotation in the third dimension. VANNOTATOR is a first effort to show how annotations of linguistic objects can be transposed into three dimensional action spaces. To this end, we provided a virtualization of an interface to a relational database model currently managed as a UIMA database. In this way, relational entities as needed to annotate propositional content can be annotated using pointing gestures as well as iconic gestures. We also carried out a comparative study by comparing VANNOTATOR with WebAnno in the context of annotating anaphoric relations. We demonstrated that VANNOTATOR goes beyond its classical 2D competitor by not only allowing for annotating hyperedges. Rather, discourse referents are represented as 3D objects which can enter into recursive annotation actions and interactions with the user. Future work aims at enabling collaborative work of different annotators at the same time on the same document in the same space. In addition, we aim at extending the annotation of multimedia content in terms of image segmentation so that segments of images can serve as discourse referents. Finally, we will integrate *TextImager* (Hemati et al., 2016) into VANNOTATOR so that text to be annotated is mainly preprocessed.

6. Bibliographical References

Abbott, D., Bale, K., Gowigati, R., Pritchard, D., and Chapman, P. (2011). Empire 3D: a collaborative semantic annotation tool for virtual environments. In *Proc of WORLDCOMP 2011*, pages 121–128.

Abrami, G. and Mehler, A. (2018). A UIMA Database Interface for Managing NLP-related Text Annotations. In *Proc. of LREC 2018*, LREC 2018, Miyazaki, Japan. accepted.

Bellgardt, M., Pick, S., Zielasko, D., Vierjahn, T., Weyers, B., and Kuhlen, T. (2017). Utilizing Immersive Virtual Reality in Everyday Work. In *Proc. of WEVR*.

Biemann, C., Bontcheva, K., de Castilho, R. E., Gurevych, I., and Yimam, S. M. (2017). Collaborative web-based tools for multi-layer text annotation. In Nancy Ide et al., editors, *The Handbook of Linguistic Annotation*, pages 229–256. Springer, Dordrecht, 1 edition.

Cassidy, S. and Schmidt, T. (2017). Tools for multimodal annotation. In Nancy Ide et al., editors, *The Handbook of Linguistic Annotation*, pages 209–228. Springer, Dordrecht, 1 edition.

Chiarcos, C., Dipper, S., Götze, M., Leser, U., Lüdeling, A., Ritz, J., and Stede, M. (2008). A flexible framework for integrating annotations from different tools and tagsets. *Traitement Automatique des Langues*, 49.

Clergeaud, D. and Guitton, P. (2017). Design of an annotation system for taking notes in virtual reality. In *Proc. of 3DTV-CON 2017*.

Cliquet, G., Perreira, M., Picarougne, F., Prié, Y., and Vigier, T. (2017). Towards HMD-based Immersive Analytics. In *Immersive analytics Workshop, IEEE VIS*.

de Castilho, R. E., Biemann, C., Gurevych, I., and Yimam, S. M. (2014). WebAnno: a flexible, web-based annotation tool for CLARIN. In *Proc of CAC'14*, Utrecht, Netherlands.

Druskat, S., Bierkandt, L., Gast, V., Rzymski, C., and Zipser, F. (2014). Atomic: an open-source software platform for multi-layer corpus annotation. In *Proc. of KONVENS 2014*, pages 228–234.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5):323–327.

Hemati, W., Uslu, T., and Mehler, A. (2016). TextImager: a Distributed UIMA-based System for NLP. In *Proc. of COLING 2016 System Demonstrations*.

Kadobayashi, R., Lombardi, J., McCahill, M. P., Stearns, H., Tanaka, K., and Kay, A. (2005). Annotation authoring in collaborative 3d virtual environments. In *Proc. of ICAT '05*, pages 255–256, New York, NY, USA. ACM.

Mehler, A., Abrami, G., Bruendel, S., Felder, L., Ostertag, T., and Spiekermann, C. (2017). Stolperwege: an app for a digital public history of the Holocaust. In *Proc. of HT '17*, pages 319–320.

Spiekermann, C. (2017). Ein Text-Editor für die Texttechnologie in der dritten Dimension. Bachelor Thesis. Goethe University of Frankfurt.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). BRAT: a web-based tool for NLP-assisted text annotation. In *Proc. of EACL 2012*, pages 102–107, Avignon, France.