Information in Virtual Spaces

Susanne Riehemann

2010/06/13

sriehemann@gmail.com
Comments welcome

1 Introduction

This paper describes a potential new interdisciplinary research field that could emerge as a result of developments in immersive interactive virtual 3D spaces, geographic information systems, information visualization, pattern recognition, linguistics, knowledge representation and sharing, and other related areas.

This research field has many dimensions and is therefore difficult to describe in a paper that is mostly linear. The space under discussion is briefly described and visualized on a map in the first section. The next section discusses some of the key 3D concepts that are referred to throughout the rest of the paper. Then, the paper discusses the representation and communication of concrete spatiotemporal information, such as trips, life events, and sporting events, which make it clear that neither a virtual reproduction nor a natural language description alone are sufficiently complete for all purposes, and suggests ways of integrating them. Finally, applications of this type of approach to the visualization of language and abstract information are discussed.

The journey through this space will show how visualizations differ from and complement natural language. The journey has two main goals. One aim is to argue that significant linguistic insight can be gained by researching how to best integrate language elements with these visualizations and by studying automatic verbalization of virtual events and automatic creation of virtual events from natural language descriptions.

A more general goal is to show where these new technologies have the potential to aid with communication and the sharing of knowledge, lead to deeper understanding and novel insight, and enable people to deal with a larger amount of complex abstract information.

2 Putting the paper on a map

To orient the reader, the information representation space is depicted as a 'map' based on the two most important dimensions, with the range of concrete to abstract information from bottom to top, and symbolic to analog representation left to right, as in Figure 1.

Written language can be placed in the center of that map, covering the area from bottom to top, with more formally structured texts to the left of the center, and more evocative narratives to the right.[1]

Children start out with concrete experiential data in the bottom right corner and move via language acquisition and formal education towards the top left corner, where abstract relationships are represented in symbolic ways, such as mathematical and logical formalisms. This paper starts in the same place, at the bottom right, with concrete spatiotemporal information, but instead of moving diagonally across the map, it moves up towards the top right corner, where abstract patterns are represented visually, as in the field of information visualization.[2]

[1] The map is not intended to be complete – items like 'Google Earth' and 'Semantic Web' are merely examples of information representation in particular parts of the space.

[2] The term 'information visualization' is conventionally used only for abstract information that is not inherently visual in nature. However, it is useful to consider in what ways 'visualization' techniques can help clarify and

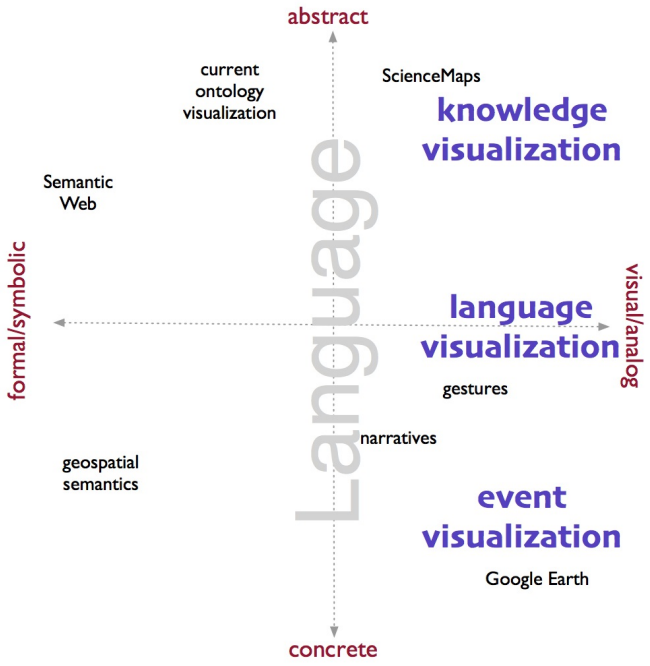

Figure 1: The space of concrete to abstract information, and symbolic to analog representation

Note that this map only depicts the two dimensions that are the most central to the paper. Other relevant dimensions are 'static/dynamic', 'passive/interactive', 'distant/immersive', 'individual/collaborative', etc. In this paper these dimensions are mostly discussed in conjunction with the visual representations on the right side of the map. In this context, 'dynamic' is the ability to show changes over time, which can be conceived as a fourth dimension, and 'interactive' means that the user has control over viewpoints and can manipulate data or affect events. The range of 'immersive' starts with 3D spaces displayed on regular 2D computer screens that appear three-dimensional because of the ability to move through the space and change viewing angles (also called 2.5D), and goes via 3D screens and projectors to IMAX domes and stereoscopic head-mounted displays (HMDs).

The reason for journeying through this space is to convince the reader that natural language is not always sufficient for a complete understanding of a complex event or system, and that visualization technologies have various advantages for understanding and communicating such information.

Linguists already know about language, but may not have thought about some of its limitations when compared to visualization. Language excels at summarizing information, focusing on particular aspects, framing an event from a particular point of view, and expressing abstract information like emotions, intentions, properties, and relationships. Stories help listeners make sense of information and provide a structure that is easy to remember and integrate into a visual model. However, natural language descriptions are by nature sequential, and when the information is very complex and detailed, listeners may have forgotten some of the earlier information by the time they hear about all the aspects of the picture. It is also impossible to control what exactly each listener will 'see'. Section 5 discusses these properties of language in more detail.

In contrast, virtual representations are more analog and can map directly onto the spatio-temporal domain, or project abstract data into that domain, taking advantage of the human abilities to integrate complex visual information, systematically explore a space and use spatial inference, and detect temporal changes. They overcome the limitation of sequential language and make it possible to include more detail than one can hold in memory, and to examine the

communicate information that has a visual aspect to begin with, and explore the space between these two extremes.

space from different perspectives to gain a better understanding and see new connections.

The advantages and disadvantages of natural language and virtual representations are such that they complement each other very well. Visualizations of abstract information usually need to be explained verbally, and even spatio-temporal data has important abstract properties that need to be communicated. Imagine taking a book about fictional events, including characters' thoughts and narrator's explanations, and combining it with a complete interactive 3D movie version of the book in which one can explore the events from different perspectives, see movements on a map, and see highlights condensed in space and time. This would lead to a more complete understanding than either the book or the 3D movie alone could accomplish. The remainder of this paper tries to show how virtual representations can provide valuable information, including insights into language.

3 Background: 3D concepts

This section discusses some of the key concepts in 3D technology that are presupposed by the rest of the paper and that may be unfamiliar to the reader: levels of detail (LODs), overlays, perspective, and devices for 'flying' in virtual spaces and for creating an immersive experience.[3]

One important concept in the area of 3D virtual reality is that of levels of detail (LODs). This technique was developed so as to stay within the limitations of current computer hardware. It takes advantage of the fact that objects at a distance can be approximated by simpler shapes requiring far less geometry, without the difference being noticeable.

For example, from a distance a tree can be represented by two 'billboards' intersecting at right angles, and still look like a tree because the textures on those billboards are photographs of a tree with enough resolution to look good from that distance. This tree looks unrealistic when viewed from nearby, especially from above, so more complexity is required at that level of detail. Similarly, when looking at the entire planet in Google Earth[4], one cannot perceive elevation differences or see an individual car, so this information does not need to be represented at that level. There are several intermediate levels between the 'whole planet' level and the highest fidelity level. Similarly, one can specify that from nearby, a certain avatar is a full-fledged model of a particular unique individual or a male of a certain height wearing a particular type of uniform, at a medium distance he is made up of two billboards, and from afar simply represented as a colored dot, or not at all, depending on the purpose of the visualization.

The concept of LODs can be useful when applied to more abstract domains, e.g. one can think of 1) paper, 2) excerpts, 3) summary, 4) topic as four levels of detail of a research paper.

The second important concept is that of overlays and layers of overlays. Hillshaded or shadow hachured (Imhof, 2007) topographic maps use shading to indicate the steepness of slopes and to simulate light sources. Overlaying such a map on the 3D terrain view in Google Earth provides information that is not as easily extracted from either source alone. In particular, aerial imagery does not look very three-dimensional when viewed from straight above, and it is not always possible to see where the highest or lowest points are. When moving through the terrain and viewing it at an angle, the elevation information can be perceived even without an overlay, but at those angles features obscure each other, and not all significant features like ridge lines and creeks are highlighted.

In contrast, a map alone does provide good overview information, but users who are not experienced with such maps can confuse ridges with valleys, may have a poor sense of the steepness of the terrain, and are often unable to form a mental picture of the actual 3D terrain.

The combined 3D-and-overlay representation minimizes these problems, making the spatial features of a larger area clear at first glance, while also being viewable from a ground level or from an oblique perspective.[5]

It is possible to make semi-transparent overlays, or only overlay hiking trails and creeks, so that the satellite imagery remains visible at the same time. The route of an actual hiking

[3]This paper was written in late 2009, and technology in this area is likely to change quickly. However, these general concepts should remain important, and changing technology will only make the research outlined in this paper more feasible.

[4]http://earth.google.com/. All URLs were current in late 2009.

[5]The value of an oblique perspective will be clear to anyone who has seen Pictometry's Bird's Eye imagery at http://maps.bing.com/

trip tracked with GPS can be overlaid on top of that. In this way, many different types of information can be combined and visualized, for example erosion, population movements, etc.

Another important aspect is the interface for navigating these 3D environments. It is possible to automatically follow a particular 'tour' as it unfolds over time. For example, in Google Earth it is possible to follow a GPS-tracked hiking trip at an accelerated pace, 'looking' in the direction of the tracked movement. The pace could be slowed down for important sub-events. This 'time lapse' method also makes the passage of time more perceptible, could be used to visualize daily and seasonal cycles, and clarifies temporal relationships. When there are simulated or tracked people in the environment, it is possible to see the events as if through their eyes.

More interestingly, the user can interact with the environment, see places and events from different perspectives, and have an experience as close as possible to 'being there'. For this purpose, computer games typically use keyboard-and-mouse approaches, special game controllers, or UI elements for zooming in, panning, rotating, and pitching the view. But these methods can be challenging to learn for older novices, especially when fine-grained control over movement in a 3D space is required.[6] There is no intuitive way to control the relative speeds of translation versus rotation.

However, there are now affordable USB-based six degrees of freedom (6DoF) controllers, e.g. the SpaceNavigator, that are fairly intuitive to learn and create the feeling of being able to 'fly' anywhere one wants to go. To navigate, one pushes, pulls, and twists the device in the desired directions – and these actions can be combined into one smooth complex motion. For example, with these devices it is easy to move forward quickly, while gently pitching downward, and slightly curving to the left, like flying an airplane on a landing approach.

HMDs or 'VR goggles' make the virtual environment even more immersive. But for most of the purposes described in this paper, a flat screen actually has the advantage of more readily allowing for inset/side windows to display related information. When 3D screens and projectors become available, they are likely to be a good compromise.

4 Visualizing concrete information

For some concrete information it is very easy to see how the various types of visual representations help with understanding and communicating information. This section starts at the bottom right corner of the map in Figure 1, and later sections build on the concrete examples to move towards the more abstract ones.

4.1 A spectrum from map to virtual reality

This section covers the space from simple 2D representations to virtual reality – that is, along the dimensions that are not directly depicted on the map, such as dynamics, immersiveness, and interactivity.

Starting with geographical maps, whether paper or digital ones, it is clear that they have advantages over natural language descriptions for many purposes. For example, showing the locations of multiple travel destinations on a map can be very helpful in identifying a sensible route, especially if the addressee is not familiar with the landmarks in the area. The map can contain more locations than one can easily keep in memory, including all the spatial relationships between them. There are already existing tools for this purpose, e.g. we can share such information using pins and descriptions on a customized Google map.[7] In this example one can also see that additional linguistic information is valuable and can help clarify the semantic relationships between the locations: *I included the bus stops near your hotel and near the department because it is very hard to find parking on campus.*

Combined with GPS, 2D maps are also valuable for the purpose of seeing where one is located on a map while actually moving through real space, e.g. driving on a highway in an unfamiliar

[6] It can be particularly confusing to mentally translate the effects of a particular control element when looking at an area from above, compared to the effects of that same control element when standing at ground level, i.e. with a viewing angle that differs by 90 degrees. Different applications also vary in using 'object pull/push' vs. 'eyepoint movement' metaphors.

[7] http://bit.ly/5FjjeL

city. This makes it possible to establish a much more direct and detailed link between places in the real world and the location of one's current position on a map, including information about what exactly one is seeing from those places, especially if a compass heading is also provided.

Moving from 2D to 3D, a Google Earth KML (Keyhole Markup Language) file with custom locations can show elevation and building models, and it is possible to define 'levels of detail' such that from a distance, only the most important locations are displayed, and further detail is added when getting closer.

Spatiotemporal data like pictures of life events and their locations can be visualized by 'touring': automatically zooming into and out of the locations in sequence. The pictures can be oriented so that the viewpoint matches, where that adds value. To add more of a 'story' element, one could control the timing, and integrate verbal narration and video elements. While the animation is 'playing' one could show the current location for each event on a small inset map, and the current date on a time line inset for easy orientation and context. This is already possible with the Google Earth plugin API, as used e.g. by the driving simulator.[8]

Ideally one could define levels of detail and topic tags for each element, so that for example only a few of the most important events in a person's life ones are shown for a very brief version, and more detail could be added for viewers with various backgrounds, interests, and goals.

Immersive 3D environments are closer to an actual 'experience' and provide the additional benefit of the users being able to see an environment or an event from multiple perspectives, and to be able to use pointing and deictic language to refer to things. For example, when trying to understand a historical battle outflanking maneuver in hilly terrain, it is important to have a bird's eye view of the scene and who is located where, and it is also necessary to understand what is visible from where, and what areas are within range of the available types of weapons. We can visit the location of the battle in Google Earth, see the elevation and get an idea of the vegetation from the satellite imagery and images taken in the area. Vegetation and rocky outcroppings can be automatically modeled based on photogrammetry and lidar scanning, and modified manually if necessary. The ranges of various weapons can be visualized as semitransparent overlays.

Currently, Google Earth is not set up for the purpose of displaying simulated or tracked movement with many human avatars, but its engine is in principle capable of doing so, and other software, e.g. MetaVR's VRSG, is available for that purpose. At a distance the avatars for the opposing forces could be represented by symbols in different colors, and when they are closer, their height could be taken into account for the purpose of judging line of sight. Speeding up the playback of the event can clarify the temporal relationships.

If we are dealing with a current event in which this technology is used for training purposes, the individual participants in the event can be equipped with GPS devices and compasses, allowing the playback of the event and viewing from any perspective. (See for example SRI's JTEP[9] project.) Video footage can be integrated, either by showing the locations of the cameras in the 3D environment, and allowing users to 'fly into' the live video camera feeds, or by automatically projecting the video images as textures onto the 3D models (see e.g. Sentinal AVE[10]). A 3D model can also be used to identify from which location videos and photographs were taken, and integrate that information into one spatiotemporally coherent picture, which can be important for example in the case of a crime that happened in a crowd.

In the case of movement by car around populated areas, the tracking does not need to be very accurate to provide meaningful data. The accuracy of the GPS in current cell phones is usually sufficient to know what street someone is on. Most routes can be simplified and described as discrete segments between turns at known intersections, which in turn can be grouped into segments between landmarks or other salient features. These partially abstract routes can be represented visually as in Mapblast's LineDrive[11] option, which are based on the types of simplifications found in human-generated route sketches (Agrawala and Stolte, 2001). These higher level abstractions from the analog data correspond well to natural language route descriptions (Dale, Geldof, and Prost, 2005).

These tools can be used to visualize and communicate a wide variety of information that is spatiotemporal in nature, for example weather patterns, urban sprawl, the history of the

[8] http://earth-api-samples.googlecode.com/svn/trunk/demos/drive-simulator/index.html
[9] http://www.jtepforguard.com/
[10] http://www.sentinelave.com/ave.html
[11] http://www.mapblast.com/directionsfind.aspx

Tour de France, or rent patterns changing over time in different neighborhoods. Note that presenting information in a geospatial context can add value even for some relatively abstract information like e.g. properties of fictional characters. For some examples, see the Google Earth gallery.[12] Changes over time, such as building activity or deforestation, can also be visualized using historical aerial imagery.[13]

A particularly good example of concrete information that is useful to visualize is space. A good 3D show, like in the Planetarium at the California Academy of Sciences, allows one to virtually leave the planet, see how thin the atmosphere really is, and explore space, traveling faster than the speed of light. In this example, that experience is as close as the human mind can get to being there. It is possible to incorporate multiple scales and allow users to fly straight from outer space into subatomic levels.

4.2 Sporting events

We will examine sporting events as a good potential testbed for natural language technology because they provide a rich and relatively complete set of information to work with, and show that neither analog reproduction nor natural language description alone are sufficient to fully understand what is happening and why.

Current game technology (e.g. Madden NFL[14]) can produce very realistic looking football game simulations. Because real life sporting events such as football happen on playing fields surrounded by cameras, they can be recorded in such a way that they can be re-visited in 3D and viewed from any angle. By extrapolating the players' fields of view from their head and eye positions it is possible to see the game through their eyes.

Coaches can use this as a tool to understand why their players behaved in certain ways, to show what a particular play will look like from their point of view, and to visualize the automatically detected weaknesses of opposing players. When available in real time, it can benefit referees trying to decide what a player was actually looking at, and whether an action was intentional.

However, even the most perfect analog reproduction of a football game is meaningless if not accompanied by an understanding of the goals of the game, and the particular situation of the teams and players. If a foreigner has never seen a football game and does not know the rules, they can learn more from a verbal description than from watching a game without any commentary. Even people who do know the game still benefit from listening to an expert, because the commentators' knowledge of the teams and leagues and the full complexity of the rules can add valuable information.

Analog reproduction and natural language description need to be combined to be able to understand the patterns of activity and the motivations behind the actions at all levels. An interactive virtual representation is better at clarifying the spatiotemporal properties and exact details of the event, and letting people form an impression of it, while language is required to clarify the more abstract relations, and add information about intentions, thoughts, and emotions.

At the most general level of detail, a football game can be distilled down to the question of which team won. At the next level, information might include the final score, whether or not the outcome was a surprise given the recent records of the teams, and what implications the outcome has on the teams' chances to advance. Language is both necessary and sufficient at these levels.

But at a more detailed level, it becomes helpful to be able to see the actual event rather than just descriptions of it. In spite of dividing up the playing field into various zones, language is not ideally suited to describing the positions and trajectories of multiple players. On the other hand, those positions and trajectories form patterns that can be usefully characterized as a certain type of play, adding information beyond what is directly observable to an untrained eye. A skilled human commentator will also be able to verbalize the dynamics of the events.

[12] http://earth.google.com/gallery

[13] At the time of writing, the Stanford Campus is a good example in Google Earth, with imagery going back to 1949.

[14] http://www.ea.com/games/madden-nfl-09

The Madden NFL game already provides game analysis, summarization, and automatically generated commentary, but it is assembled from relatively 'robotic' pre-recorded phrases such as *got the completion*, *got just past the first down marker*, and *that'll keep this drive alive*. Both the computer game and the virtual version of the real event make a very good domain for natural language generation software, because so much knowledge of the events is available for reference, including what the intended actions were. The domain is restricted enough to be tractable, and the rules of the game provide relatively clear categories and labels for each role, action, and location.

In addition to generating natural language descriptions, a natural language interface is valuable for the purpose of revisiting parts of the events. People are not likely to remember the exact time at which each event happened, but a unique verbal description is easy to produce, e.g. *go back to the moment when our quarterback made a handoff after dodging three tackles*. Similarly, natural language could be used to ask for a visualization of possible alternative scenarios: *what could have happened if the wide receiver had run at his personal best during that play?*

However, not all of the relevant higher level patterns are best communicated verbally. For example the 'flow' of the game can be visualized by showing a time lapse version with a sequence of scoring points along with a visualization of the score – making it clear which team is ahead, and by how much, at what point. If various relevant properties of the players are indicated visually, that information forms a spatial pattern that shifts with their movements on the field and makes the combined information easily accessible.

4.3 Virtual lives

Virtual lives are another illuminating example. Automatically narrating real life events in addition to visualizing them would help identify the important abstract concepts and relations in the raw data and provide a summary for easy reference, facilitating learning from the past for the benefit of the future. This is currently not feasible even as a research project because of a lack of input data. People's lives are not being tracked with GPS and recorded on video. However, Bell and Gemmell (2009) argue that this will become common very soon. It is possible to start working on the technology now and seeing what its benefits and limitations are by doing this for 'life events' in virtual worlds like Second Life or World of Warcraft (WoW).

Because these environments are 'multiplayer', all the data about the movements and actions of each player are being transmitted in a compact symbolic form, and could be recorded. This would allow for a perfect playback of each event, viewable from any perspective desired, including the actual viewpoints each player had during the event. Several hours of playing could be summarized by showing the most significant events in detail (e.g. leveling or achievements in the case of WoW), and including snapshots of the somewhat smaller events if desired (finishing quests, entering significant areas, meeting people, etc.), while condensing a relatively long period of time into a statement like like 'after freeing the hostage, he took the boat from place A to place B'. Visualization of the track on a map could be one major aid not only to summarization but also to the future retrieval of these events.

Even though many events in WoW are 'discrete' – for example it is never unclear whether or not a player laughed – there is also much activity that is relatively analog, most notably in the movement of the characters around the landscape. With some natural restrictions (walls or steep slopes) the players can walk or ride or swim (and later fly) anywhere. Spoken languages are not good at describing complex spatial trajectories, and maps with tracks overlaid can communicate this information much better. But language is useful for describing the higher level patterns. In this domain there are many clues as to the intended destinations – what 'quests' a player has available or looked at before setting out, what the significant locations are, where the player ultimately ends up, and what they wrote in chats with fellow players. Goals can be described at various temporal horizons and levels of concreteness, e.g. *I am on my way to the store – I am going to do some engineering and will stop by the store to pick up supplies – I am making the explosives that I need for the dungeon later today – I am working on improving my contribution to group quests*, that are successively harder to identify automatically.

The typed and spoken communications among players would be a good corpus of natural language expressions with known referents and full contexts, and a good environment for tuning intention detection and other behavior analytics methods, and learning how to identify and

summarize an otherwise random seeming pattern of movement, e.g. *he was trying to take a shortcut to location X but couldn't find a way through the hills*, or *while looking for person X in area Y, she got attacked by a wolf*, or *she needed to make space in her bags and made a detour via the trading post to sell some items*. For complex events like a battle between many players, some of the mistakes and other factors that explain the outcome could be automatically identified.

This artificial environment makes it more tractable to study various aspects of life events and how to represent them: concrete geospatial 'who, when, where' information, fairly complex activities and relationships, and internal states like health (visualized in the game as a bar), emotions and goals. If automatically generated natural language summaries of each day's events and achievements were provided in a textual and/or audio-visual form, the researchers could probably get feedback on the quality of the summaries, or even corrections, from large numbers of players.

5 Visualizing linguistic information

This section discusses visualizing the information conveyed in natural language utterances, and how this can lead to a deeper understanding of the effects of natural language. On the map in Figure 1, language is positioned between concrete and abstract information. It can be about both, but is somewhere in the middle most of the time. Natural language descriptions of abstract objects are often based on analogies with more concrete ones. Conversely, natural language descriptions of very concrete events are usually abstractions that generalize, simplify and leave out much specific information.

The simplifying effect of language is particularly true for spatial expressions, at least for languages other than sign languages[15]. We do not usually specify exact distances, elevations, and dimensions. Worse, we do not have a conventional way of talking about orientation angles, at least in English and most other spoken languages. The most precise it seems to get is the 30° increments one gets from descriptions like *the bird is at 11 o'clock*, and even those are being used less frequently in the age of digital clocks. Phrases like *the bird is at 30 degrees* do not have a standard interpretation specifying from what axis and in which direction to rotate, and are not frequently used, even in increments that could be distinguished.

It gets worse when the third dimension is added, with no conventional way to describe pitch or roll angles, distinguishing between them, distinguishing them from yaw angles, describing combinations of them, or verbalizing any type of motion through 3D space. Some of this information can be communicated through gestures, but this is not helpful for written language.

In addition, the interpretation of prepositions is often dependent not only on the geometry of the objects that are being related but also their function in the context. For example, there are two interpretations for *The girl was in front of the car*: in front of the front of the car in the direction it is normally traveling, or in front of the car from the point of view of the speaker, i.e. not visually obscured by the car - see Figure 2. So for most natural language utterances it is not the case that there is exactly one correct visualization.

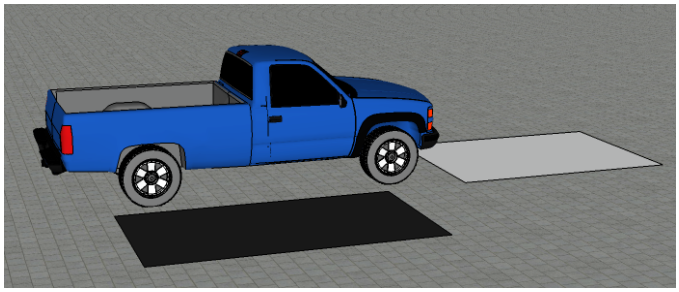

Figure 2: Two interpretations for *in front of* the pickup truck

[15]Sign languages make it possible to express many more categories and elements at the same time and in more flexible combinations (Talmy, 2001).

However, it can be highly instructive to depict the range of possible visualizations and clarify what the area of uncertainty is for each type of expression and to show the source of some misunderstandings.[16] And it can provide insight for linguists studying spatial language who do not always seem to consider how language maps onto the real world.[17]

There is already some related work in linguistics. Pustejovsky and Moszkowicz (2008) try to derive spatiotemporal data automatically from textual information. To that end, SpatialML markup was developed and used to annotate corpora. Bergen, Lindsay, Matlock, and Narayanan (2007) give evidence for the role of mental imagery and simulation in language comprehension. WordsEye[18] is an interesting attempt to automate the creation of 3D scenes from natural language descriptions about the sizes, locations, and orientations of objects (Coyne and Sproat, 2001). It analyzes the natural language inputs into sets of semantic elements and combines them graphically. The 3D scenes it generates are static, but do use 'poses' to depict actions. This work contrasts with formal linguistics, which has been mostly concerned with bridging the gap from language, in the middle of my map, to precise formal representations that computers can make use of, on the left. WordsEye instead tries to bridge the gap between language and what it is about in the real world, towards the right.

A WordsEye type of approach could be augmented to show the range of possible objects, properties, locations, and poses that are consistent with the language input. It could be used to visualize differences in individuals' mapping between language and the world, and when applied cross-linguistically, lexical gaps for categories at different levels of generality. It could also visualize the notion of 'focus', by highlighting the focused elements visually, e.g. with a halo of light.

Other aspects of language, such as generalized quantifiers, tenses, modals, negation, and embedded clauses of various types, are harder to visualize. One might think that it is not worth trying, because language seems clearly superior. It may well be the case that for practical applications it is best to use verbal elements for these purposes, but it is still instructive to think about how one would go about trying to visualize them. This could lead to novel linguistic insights and an appreciation for why they can be hard to learn. It might also become clearer when the state of the virtual model of the world in a listener's mind is such that a particular utterance is harder to process. For example, some aspects of natural language expressions require more than one virtual world, fast-forwarding through the changing state of one world over time, or even running simulations of various possibilities. Negations like *I did not do laundry this weekend* are particularly interesting. While it might be possible to come up with a convention to mark negated content visually, it is also interesting to consider what happens if someone's representation is the assumed state of affairs given that the action did not take place, such as a pile of dirty laundry in the example above. In most contexts there is probably an implicature that this state of affairs is true, assuming that the speaker was cooperative. But it is not necessarily true, and might result in a false belief that this information was actually directly communicated.

A language visualization system, when combined with user feedback, could also be used to investigate what viewpoints, i.e. camera angles, are the most natural to distinguish active and passive, or describe the perspective differences between *he loaded hay onto the wagon* and *he loaded the wagon with hay*, and how these influence what is communicated. Conversely, this would lead to more insight into how language can 'frame' an event such that the listener 'sees' it from a certain perspective. One could study how a 3D movie can tell a different story when the same events are shown from different angles, or how certain perspectives result in a cognitive mismatch because they do not fit the story. One could also visualize how someone's knowledge and goals highlight certain elements in their environment and focus their attention, and how language can refocus them on other properties, objects, or patterns. And such a system makes it clearer what elements are not being communicated directly and need to be supplied by the listener's imagination unless the language is made more vivid. Combining WordsEye with

[16] Some of the same methods can also be used for memories of past events and thoughts about future events, which are incomplete in many of the same ways that natural language descriptions are, and often consist of separate threads or stories that are not combined into a coherent picture.

[17] At a recent conference discussion at CSLI, several semanticists were surprised about the fact that a road that *narrows* in one direction can be described as *widening* from the other perspective.

[18] http://www.wordseye.com/

Google Earth or Second Life would open up another large set of research opportunities.

Comparing a virtual reproduction of a real event with a visualization of the aspects that are being communicated verbally would yield insights into what exactly is gained and lost in the process. The details that are lost can be visualized as the uncertainty inherent to the underspecification in various aspects of the natural language description, and what is gained can be visualized as the focusing effect that language has to direct our attention to the important aspects of the event or its context, as well as the categorizing effect it has on seeing a variety of actual events as 'the same' at that level of abstraction.

6 Visualizing abstract information

This section gradually moves into the upper right quadrant of the map in Figure 1 and suggests how the same methods described above can be used for more abstract information. As a starting point, it is relatively straightforward to visualize some properties of concrete objects, relationships between them, and abstract information that is spatial in nature, like coordinate systems or geometry.

For example, one could make an overlay of latitude and longitude lines in Google Earth, and an animation that shows what effect projecting the 3D globe down onto a 2D computer screen has on, for example, a round lake that is far from the equator, or on the size of the landmass of Greenland, and compare the various distortions of commonly used map projections.

The octants in a three-dimensional Cartesian coordinate system, and vectors in that space, can also be represented much more clearly and unambiguously than in any graphical representation projected onto a two-dimensional page.

6.1 Metaphors

Many types of abstract information have structural similarities with more concrete knowledge. Metaphors exploit these analogies (Lakoff and Johnson, 1980), and make it possible to visualize aspects of the abstract concepts. For example, it is possible to visualize many aspects of relationships using the 'love is a journey' metaphor. For the purpose of this paper it is not important whether this metaphor is seen as directly influencing the 'love' concept, or whether it merely shows enough structural similarity with that concept for the mapping to be meaningful (Murphy, 1996), (Nunberg, Sag, and Wasow, 1994).

The high-dimensional similarity space of various life goals (learning, contributing, meaning, happiness, health, relationships, children, respect, honor, attention, legacy, pleasure, leisure, comfort, wealth, power, independence, security, peace of mind, ...) could be mapped down to 2D using information visualization methods. For example, when looking at a large corpus of data one might find that power overlaps with wealth but is more distant from peace of mind. Elevation in the terrain could be used to show the relative difficulty of getting to certain places, i.e. the amount of effort it usually takes to e.g. become a doctor.

For the purpose of elucidating relationship issues, it is helpful to be able to see the metaphorical locations of the life goals of the partners with respect to each other. For example, if one partner has exploration and learning from other cultures and *expanding their horizons* as one of their goals, then they might feel like things *aren't going anywhere* if the other partner prefers to maximize their comfort and relaxation and stay at home. If one partner values their career above all else, and to the other partner family relationships are the most important, then they may well reach a *crossroads* where they have to *go their separate ways*. Some roads, for example the one towards having children, cannot be traveled successfully by one partner alone, resulting in a *dead end* if the other partner does not share that common goal and does not wish to travel in that direction. Some regions can be visited without any effect on which other regions are reachable later, while other roads are *one-way streets* that cause other options to become inaccessible and make it impossible to *turn back*. Life's vicissitudes like diseases and accidents can be visualized as obstacles along the road.

If such a metaphorical life space were available to visit virtually, one could visit various possible future locations, see what some of the properties of that area are, and look at the journey to see how much energy it will take to get there. This might help people place more emphasis on

how and where the bulk of their time is spent instead of only chasing new destinations to get to. It would also make it clearer which goals are incompatible or hard to combine with each other. Comparing how different people visualize themselves and their goals in such a metaphorical space might also help elucidate the causes for misunderstandings, for example when a linguistic expression is being is being wrongly interpreted as an act of aggression when in fact the collision was accidental and due to crossed paths and lack of attention to where the other person was coming from and where they were trying to go.

6.2 Information Visualization

For other abstract information, it is somewhat less obvious that visualization, especially visualization in 3D, is beneficial. Visualizations that are trying to depict complex information and relationships require guidance to interpret the visualization. Many 'visualization' attempts on the web are more confusing than helpful because labels, captions, and accompanying explanatory text are absent or insufficient. Information visualization in 3D also faces the challenge that information can be obscured by other items, and text can be hard to handle.

One reason to believe that visualizations are helpful even on the most abstract end of the scale is the mounting evidence for the perceptual basis of reasoning (Levinson, 2003; Casasanto, 2005; Landy, Allen, and Zednik, 2009), and the fact that mathematical and logical concepts can benefit from visualization, cf. Tarski's World (Barwise and Etchemendy, 1993), Mathematica (Wolfram, 2003), Venn diagrams, and visual proofs, e.g. of Pythagoras' theorem (Nelsen, 1997). In some cases the visualizations for more abstract properties can be combined with representations of the concrete data they are describing.

The field of information visualization focuses on visualizing the most complex scientific and other phenomena we deal with today, highlighting the salient features and allowing researchers to detect patterns in very large data sets much faster than they otherwise could (Johnson, Moorhead, Munzner, Pfister, Rheingans, and Yoo, 2006). This is achieved with very sophisticated methods compared with the other 'visualization' methods discussed in this paper. Google Earth has made geospatial data visualization accessible to more than just geographic information systems (GIS) experts, but other tools are not as easily available or combinable. Connecting all the way to the top right corner of the map in Figure 1 would require interdisciplinary collaboration with visualization researchers. It is necessary to study what types of information are best communicated verbally, which are better suited to visualization, and how to combine the two modalities the most effectively.

As one example, the Science Map project (Boyack, Klavans, and Börner, 2005) used statistical similarity measures to map the space of academic publications. This resulted in a high-dimensional space which was simplified down to a 3D version and also projected on to 2D. This makes it possible to locate interdisciplinary topics on a map.

Using these maps in a manner analogous to Google Earth, it becomes possible to represent the areas of knowledge of individuals, and of locating the knowledge that someone is trying to share in a communication attempt. Then the directions to this communicative goal can be read off the map and explained with respect to known landmarks. By looking at the space from different perspectives, new and previously unexplored connections become apparent. Missing connections can be filled in by people with expertise in both subject areas. Someone's intellectual history can be visualized in a manner analogous to the life events described above. When people collaborate online, their background and their stated goals can be taken into account to make sure their contributions are seen from the right perspective.

6.3 Knowledge Models

For any complex knowledge to be acquired and fully mastered, it is necessary to combine the individual pieces of information into a coherent structure. Consider for example a relatively complex piece of software such as Photoshop. If one simply teaches someone how to perform a few fixed procedures, they may not get enough information to build a reasonably good working model of the software, and may not be able to solve new problems. The connections between some of the individual pieces of knowledge will be obscure to them, and new knowledge harder to remember because it is not integrated with anything else. They can also spend a lot of time

looking at irrelevant parts of the screen and may not be able to quickly scan the menus and toolbars to find what they need.

One way to build a good model of such a new domain of knowledge is to figure it out on one's own, which can result in a relatively deep understanding, good connections to other knowledge, and a decent likelihood of remembering the information. However, this can be a lengthy process which time and other factors can make infeasible, so that it becomes necessary to try to learn from someone who already knows it.

However, because the structure of the model is multidimensional, it is hard to share via documentation written in natural language. In addition, people who have complete knowledge of the domain have often forgotten what it was like not to have that knowledge, and have trouble identifying even basic sources of problems. They may not know which technical terms are confusing, and which pieces of information someone may be missing, especially if these are the most fundamental ones. In this example, that could be the concepts of pixels and resolution, information about the range of functionality to be expected from the software, or familiarity with certain types of UI elements, which might not even be identifiable from verbal descriptions.

This problem gets worse as the knowledge to be acquired becomes more complex and abstract. Even in the case of this type of software, 3D development or GIS software is far more complex than Photoshop. In a more abstract case, for example while doing research for a paper, the amount of information that needs to be integrated can be extremely complex, there is no documentation to consult, and there are no good tools to help with this process. The best available one appears to be Tinderbox,[19] which allows for notes to be organized visually. It is useful to be able to spatially arrange, group, color-code, and link ideas and concepts, especially in the brainstorming phase of a project. Tinderbox also allows for symbolic properties and parallel structured views of the same data. However, it requires a major time investment to master, does not include 3D visualization, and is not suitable for collaboration.

A good knowledge model tool needs an intuitive 3D UI allowing for the dynamic rearranging of items, saving multiple arrangements from multiple viewpoints for easy revisiting, and sharing these with others, perhaps even automatically comparing perspectives and identifying mismatches. Techniques from the information visualization field are required to solve the problems of representing textual information in 3D and avoiding issues with occlusion, i.e. information being obscured by other information. The system also needs LODs for abstractions, e.g. the ability to zoom or 'fly' with a 6DoF controller or another effortless method into a note that is a three-word description of an idea to see a more detailed summary, with each sentence of the summary in turn linking to a relevant section, down to the ability to fly to the location in a source document in which the information originated. These tools would help make sure that the most significant information is not lost when 'squinting' to see things at lower levels of detail, and accelerate the process of integrating diverse bits of symbolic information into a complex picture that is a closer model of the analog reality. This is analogous to speeding up the process of building a complete mental map of a complex new neighborhood by sketching a map during the exploration process, or having someone else's sketch available to start from. Such a virtual knowledge model is beneficial even to people who have internalized one already for a particular subject area. The computerized version can contain far more detail than one can hold in memory, allows for viewing from multiple perspectives, enables a more systematic exploration of the space, and utilizes the human capacity for spatial inference. It can also be shared with others and saved for the future without degrading.

But it also needs to be possible to see structured views of the same information, which should be linked to formal ontologies and the semantic web[20]. Someone who has a complex model of a domain in their head may be able to 'see' solutions to problems intuitively without being able to justify them to others verbally, because the symbolic complexities are hidden in a more analog representation and the solution was not derived by logical reasoning. An integrated visual and structural representation of the problem space can help explain the solution, and build a bridge between an analog state to a more symbolic one that can be shared more easily in words.

Sharing knowledge between two individuals is easier if their virtual mental models are similar. This is facilitated if both use the external world as the basis for their virtual model, and have

[19] http://www.eastgate.com/Tinderbox/
[20] http://semanticweb.org/

integrated a sufficiently similar range of experiences into that model, or managed to communicate the missing aspects. However, for most people there are too many differences between their internal virtual world and the real world, and consequently between two individuals' internal worlds. Jacobson (1991) argues that this is partly due to the fact that so much of our information is indirect and supplied by the media. Visualizing the most common 'TV tropes'[21] might help make it clear in what ways our perceptions of reality are skewed.

Technological tools can help bridge the gap between individuals' models by visualizing the differences, transforming the information so that it lines up, and helping people make better virtual models. They can help put new pieces of information in the right place in the virtual model analogous to the way a GPS enabled phone shows your location on a map. If information from many individuals needs to be combined, such as for online collaboration projects, it becomes essential that there is a shared reference model, so that the contributions from various different perspectives can be integrated appropriately.

Better virtual models and natural language interfaces to them will also become indispensable in the context of the 'e-memory revolution'. The more information an individual can store, the more important it will become to have intuitive ways of accessing that information, finding the relevant parts, and organizing it in such a way that one can derive insight from it. If the information is not integrated into an existing mental model, one is less likely to remember it and make use of it.

7 Conclusions

This paper has attempted to summarize currently available 3D virtual world and visualization technology and its benefits for gaining insight and sharing knowledge. The analysis of the strengths and weaknesses of these tools compared to language suggested that they complement each other well if the visual representations are supplemented with integrated natural language descriptions of the more abstract levels of the information. Two domains, virtual game environments and sporting events, were discussed in terms of their value as a testbed for natural language technology, because they include detailed information about events as they unfold over time. It was argued that much can be learned about spatial and other aspects of language by constructing 3D representations of natural language descriptions and visualizing what is gained and lost in the process.

This has been but one thread woven through the complex multidimensional space of information visualization and virtual reality, trying to use language to evoke the depth of that space in your imaginations. Once this interdisciplinary research field is more established and the tools are available, expect to be able to visit a model of this space directly, see it both from this author's perspective and many others, and hopefully contribute to it.

Acknowledgements

Thanks to Edward Zalta and John Pacheco for patiently listening to me talking about this subject for far too long. Thanks to Mike Beebe, Brett Heliker, James Steele, and Daniel Gruver for sharing their 3D and GIS wisdom with me, and a special thanks to Jesse Alama for exploring the virtual world with me. In addition to the above-mentioned people, Emily Bender and Chris Culy provided helpful comments on an earlier draft.

[21] http://www.tvtropes.org/

References

Agrawala, Maneesh and Chris Stolte. 2001. Rendering effective route maps: Improving usability through generalization. In *ACM SIGGRAPH Proceedings*, pages 241–249. Addison Wesley.

Barwise, Jon and John Etchemendy. 1993. *Tarski's World*. CSLI Publications, Stanford, CA.

Bell, Gordon and Jim Gemmell. 2009. *Total Recall: How the E-Memory Revolution Will Change Everything*. Dutton, New York.

Bergen, Benjamin, Shane Lindsay, Teenie Matlock, and Srini Narayanan. 2007. Spatial and Linguistic Aspects of Visual Imagery in Sentence Comprehension. *Cognitive Science* 31:733–764.

Boyack, Kevin, Richard Klavans, and Katy Börner. 2005. Mapping the backbone of science. *Scientometrics* 64:351–374.

Casasanto, Daniel. 2005. *Perceptual Foundations of Abstract Thought*. Ph.D. thesis, MIT.

Coyne, Bob and Richard Sproat. 2001. WordsEye: An automatic text-to-scene conversion system. In *ACM SIGGRAPH Proceedings*, page 487496. Addison Wesley.

Dale, Robert, Sabine Geldof, and Jean-Philippe Prost. 2005. Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology* 37:89–105.

Imhof, Eduard. 2007. *Cartographic Relief Presentation*. ESRI Press, Redlands, CA.

Jacobson, Robert. 1991. Virtual Worlds, Inside and Out. In D. Mark and A. Frank, eds., *Cognitive and Linguistic Aspects of Geographic Space: An Introduction*, pages 507–514. Kluwer Academic.

Johnson, Chris, Robert Moorhead, Tamara Munzner, Hanspeter Pfister, Penny Rheingans, and Terry Yoo. 2006. NIH/NSF Visualization Research Challenges Report Summary. In *IEEE Computer Graphics and Applications*, vol. 26(2), pages 20–24.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press.

Landy, David, Colin Allen, and Carlos Zednik. 2009. A perceptual account of symbolic reasoning. ms, University of Richmond.

Levinson, Stephen. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press, New York.

Murphy, Gregory. 1996. On metaphoric representation. *Cognition* 60:173–204.

Nelsen, Roger. 1997. *Proofs without Words: Exercises in Visual Thinking*. The Mathematical Association of America.

Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 1994. Idioms. *Language* 70:491–538.

Pustejovsky, James and Jessica Moszkowicz. 2008. Integrating Motion Predicate Classes with Spatial and Temporal Annotations. In *COLING 2008 Companion Volume - Posters and Demonstrations*, pages 95–98.

Talmy, Leonard. 2001. How spoken language and signed language structure space differently. In D. R. Montello, ed., *Spatial information theory: Proceedings of COSIT 2001*, pages 247–262. Springer, Berlin.

Wolfram, Stephen. 2003. *The Mathematica Book*. Wolfram Media.