

COMMUNITY ACTIVITY: TUTORIAL REPORT

Tutorial report: Understanding spatial thought through language use

Summary: The tutorial “Understanding spatial thought through language use” took place at the International Spatial Cognition Conference on August 31, 2012 at Kloster Seeon in Germany. This report outlines the main rationale for the tutorial along with central contributions by its participants, who considerably enhanced the success of the tutorial by sharing and discussing their own research experiences with respect to the analysis of language in spatial cognition contexts. The tutorial’s website is <http://knirb.net/TutorialSC2012.html>.

1 Introduction

The International Spatial Cognition Conference 2012 was located at the picturesque Kloster Seeon in Germany, and (like each of its instantiations) attracted spatial cognition researchers from across the world. This year it hosted the tutorial “Understanding spatial thought through language use,” which addressed questions around how language can be analyzed so as to reveal central aspects of spatial thought. Organized by Thora Tenbrink, the tutorial was considerably enriched by the contributions of its participants, who shared their own research experience and discussed further ideas concerning the systematization of language analysis. This report outlines the tutorial’s main contents, enhanced by the participants’ diverse research examples.

Language plays a role in many different areas of research in cognitive science, for example when giving verbal instructions to experimental participants, reporting participants’ comments as (unsystematic) anecdotal evidence, or eliciting verbal protocols that provide insights about problem solving processes [3]. Descriptions of scenes and events are verbalized representations of spatial (or spatiotemporal) perception [13]. Think-aloud protocols and retrospective reports can provide procedural information that complements behavioral performance results in spatial problem solving tasks (such as wayfinding [12]). Such data can be analyzed with respect to content as well as structure [11, 15]. Much research in cognitive linguistics, psychology, discourse analysis, and psycholinguistics indicates that patterns in language are systematically related to patterns of thought. These insights can be utilized for spatial cognition research.

Speakers producing language in relation to a spatial scene, event, or problem solving task may not be aware of the cognitive structures that are reflected in particular ways of framing a representation linguistically. Also, they may not be consciously aware of the underlying network of options that allows for a range of linguistic choices beside their own, which emerges more clearly by considering a larger data set collected under controlled circumstances. According to research in cognitive linguistics and discourse analysis, linguistic features indicate certain conceptual circumstances. Linguistic features, such as such as the verbal representation of semantic domains reflected in ideational networks, lexical omissions and elaboration, presuppositions, hesitation and discourse markers, and the like, are related to the current cognitive representations in ways that distinguish them from other options available in the network. Besides building on established insights about the significance of particular linguistic choices, validating evidence for the relationship between patterns of language use and the associated cognitive processes can be gained by triangulation, i.e., the combination of linguistic analysis with other types of evidence such as behavioral performance data.

The main goal of this tutorial was to raise awareness to the insights that linguistic data analysis can contribute to empirical research in spatial cognition, as well as to acquire and share practical expertise. The tutorial was highly interactive, taking into account the participants' relevant background knowledge and specific research goals. We started by discussing the role of language in spatial cognition research, followed by hands-on practices and discussions concerning data collection, transcription, and analysis procedures. In the following sections, we will retrace the tutorial's path by integrating the individual participants' contributions at relevant stages.

2 Motivation

Our tutorial started by considering how language data can serve as empirical evidence for research in spatial cognition. Following a general discussion of the issues and insights summarized in the introductory section above, two examples of ongoing research were presented to provide concrete ideas about the relevance of language analysis in this area.

First, Jinlong Yang presented his research on evaluating qualitative spatial calculi in collaboration with colleagues at the Human Factors in GIScience Lab at the Pennsylvania State University, USA. This research focuses on cognitive aspects of qualitative spatial calculi by taking a behavioral assessment approach. Category construction experiments were carried out to shed light on the conceptualization of spatial relations between two objects (e.g., a hurricane and an island, a lake and a house) in geographic events. A set of animated stimuli depicted movement patterns in geographic events such as hurricane and flooding. Participants were asked to sort these stimuli into groups, using criteria that they considered appropriate. In addition to the similarity ratings derived from participants' grouping behavior, linguistic descriptions were collected that elaborated on their reasons.

The language data reveals the major rationales employed by participants in the category construction phase. This provides guidelines when different types of visualization methods are applied to analyze the data. For instance, in the cluster analysis of the similarity ratings of stimuli, the rationales gleaned from the linguistic descriptions can help to decide the optimal cut of the dendrograms derived from different types of clustering algorithms. Furthermore, the linguistic descriptions can be utilized to validate patterns found in the



analyses. In some cases, patterns found from visualization and statistical analyses might be artificial or random effects, rather than reflecting actual rationales employed by participants. Validating patterns on the basis of linguistic descriptions can help avoid such situations, minimizing the possibility of over-interpreting the data.

Second, Christoph Hertzberg reported on current work in the area of urban search and rescue (USAR). One of the main fields of application in this area is the localization of persons in collapsed buildings. For that purpose, endoscope-like devices are often employed, which allow USAR workers to look from outside into currently inaccessible areas of the structure. A major challenge here is to keep track of the camera's position and orientation, as it moves out of the operator's sight. Furthermore, it is hard to estimate distances inside camera images, especially if too few reference objects can be seen. The technical goal of this project is to provide software which, using camera images and other sensor data, calculates a virtual 3D model of the structure [4]. This should help USAR workers in localizing persons and identifying good locations for support or rescue drillings.

Along with benchmarking the accuracy of this software, its actual usage value also needs to be evaluated. For this purpose, an initial exploratory study was conducted in collaboration with Thora Tenbrink, Carsten Gondorf, and Evelyn Bergmann at the University of Bremen. The aim of this study was to investigate users' perception of the complex images provided by state-of-the-art endoscopes, based on their linguistic descriptions. A Styrofoam mock-up mimicking a collapsed building was built, with hidden objects inside it. A camera-head was moved through this structure, simulating typical endoscope movements (mostly forward, combined with large rotations and turnings of the camera). Participants in this study could freely navigate through the recorded images, i.e., virtually move the camera back and forth. They were asked to mark objects inside the images by clicking on them, and to think aloud while doing so. The linguistic descriptions collected in this way reveal the extent to which the participants could make sense of the distorted and frequently rotated images shown on the screen, and highlight search and identification processes. Following completion of the 3D software, comparative studies are planned so as to establish its usability in practice.

3 Data collection methods

The next issue discussed in the tutorial concerned the methods of data collection that could be useful in spatial cognition research. Various kinds of empirical designs and language elicitation methods were considered with respect to their advantages and disadvantages in light of concrete research purposes. Options range from verbal descriptions of scenes and events, via think-aloud protocols and retrospective reports, to interviews and dialogs.

Güzin Mazman presented the intriguing methodology of cued retrospective think aloud, which in their ongoing study is combined with eye tracking [6]. The purpose of this study is to examine individuals' computer-based complex task performance, processes, and strategies in order to determine reasons for failure. Five senior students were confronted with a complex computer-based task that included a logical reasoning process. Their eye movements were tracked during the problem solving process. Afterwards the participants were asked to think aloud while they were shown a gaze video replay of their task performance.

The rationale for using this particular method was as follows. As is generally acknowledged, the elicitation of concurrent think-aloud data may potentially influence performance

particularly in cognitively demanding tasks, by distracting participants' attention and increasing their cognitive load. By asking participants to think aloud only retrospectively, these limitations of concurrent verbal protocols can be overcome. Since participants may forget important steps of their performance and start fabricating some aspects, it is useful to present visual cues that reactivate the task process. Playback videos of the task session facilitate the retrieval information from memory and provide veridicality. Since eye movements provide another objective measurement of cognitive processes, they should facilitate reporting thoughts and elicit comments from participants when they are used as a cue. Both kinds of data, verbal protocols and eye movements, reveal valuable information about cognitive processes and thus allow for a triangulation of data to enhance validity of the findings.

In Mazman and Altun's study, none of the participants completed the task successfully within the given amount of time (ten minutes). Following transcription of the retrospective protocols, a coding schema was developed iteratively from segments. Results revealed seven distinct cognitive strategies, with trial and error as the most employed strategy, employed without reasoning. The general thought process could be modeled by defining sequences and the relations between actions and cognitive strategies. Based on these results derived from cued retrospective reports, it is suggested that participants could become more successful if they were provided with strategy instructions, raising awareness of their current procedures.

4 Data analysis

Prior to attempting any systematic analysis, a thorough understanding of the contents is essential so as to understand what people say and mean, and to develop intuitions and gain inspiration from the verbal data. Furthermore, those utterances need to be identified that are relevant to the research purpose at hand, and the relations between specific utterances and the task (or specific steps of it) should be established. The content of verbal utterances can then help to reconstruct the cognitive path(s), including individual differences, and to identify crucial cognitive processes. These cognitive processes may include false leads and dead ends, insights, causal relations, logical considerations, decisions, explicit reasoning processes, action plans, reasons for actions or non-actions, and many other conceptual aspects relevant to the research question at hand. To derive meaningful insights from the data, it is essential to employ systematic content annotation procedures such as those suggested by Ericsson & Simon [3] and Krippendorff [5].

As a next step, content analysis can be supported and substantiated by a closer examination of the linguistic features of the verbal data. Systematic patterns in language support the operationalization of relevant content distinctions and categories. Furthermore, specific linguistic structures can highlight cognitive structures, such as current focus of attention, conceptual perspectives, granularity levels, and specific types of activated or retrieved concepts as revealed in the lexical choices that speakers make from the network of linguistic options available to them [14].

To illustrate these issues for a spatial cognition context, one method commonly used is to ask people to report spatial information by free recall. After hearing a description of an environment, people can expound the spatial content in different ways. Such a description might be as follows [7]: "At the corner of the holiday farm, there is an entrance gate.

Once through the gate, you will find a water well used to irrigate the field on your left. Go straight on and you will find a nice restaurant in front of you. Then turn left, leaving the restaurant in the corner behind you." When asked to recall this kind of information, people generally use the same reference frame as in the description (which is intrinsic in this case), while some give a list of non-spatial and/or landmark-based details ("there is a water well") and rarely use extrinsic reference frames ("the gate is at the south-western corner of the holiday farm") [8].

However, there is considerable inter-individual variability in the descriptions. Some people strictly follow a sequence of landmarks ("I will find the water well on the left, then I will see the restaurant on the other corner"), while others locate landmarks accurately, but in a different order ("I will see the restaurant in front of me in the corner after I have walked past a water well on my left"). There are also verbal outputs that combine different sequences of information with more or less accurate landmark locations (e.g., "I pass the water well" vs "I will find the water well on the left"). One method often used to score such verbal feedback is to award one point for each landmark that appears to have been correctly located (whatever the format used to provide this information). Usually two judges score test protocols independently and their scores are then correlated. If they correlate well, but not perfectly (e.g., $r = 0.60-0.70$), a third judge can resolve the discrepancy between the scores awarded by the previous two. The inter-rater agreement can also be calculated using Krippendorff's Alpha [5]. In the light of the discussion above, it is easy to imagine that the methods used to analyze verbal protocols could be improved with a view to enabling a better detection of individual language patterns. The spatial language domain can provide new tools for a better approach to studying verbal production on spatial descriptions, thereby broadening and enriching the criteria to consider when assessing such verbal production.

In the tutorial, participants contributed two concrete research examples. First, Nina Reshöft reported on the analysis of spatial scene descriptions, with the aim of understanding how languages structure space and how different linguistic structures may point to cross-linguistic differences in spatial thought. In particular, research on the encoding of motion events shows that languages differ in the way in which the path of motion is lexicalized. Studies in this domain are typically based on a two-way typology proposed by Talmy [10]. He distinguishes two language types according to where the path is expressed in the surface structure. *Verb-framed languages* such as Spanish tend to express path in the main verb, whereas *satellite-framed languages* like English show a tendency to express the path of motion in a satellite, expressing manner of motion in the main verb, as in "He ran out of the house." By contrast, verb-framed languages express manner of motion optionally in an adjunct, as in "Salió de la casa (corriendo)—He exited the house (running)." Although much research follows this two-way distinction, it has led to considerable controversy in the field, with many scholars proposing some kind of revision or modification [9].

In order to understand and explain cross-linguistic differences in spatial thought, a semantic annotation scheme is needed that is flexible enough to account for the context dependency of spatial expressions. In Reshöft's study, natural language data was collected from speakers of English, German, and Spanish in a narrative elicitation task. The narratives were based on a wordless picture story book in order to elicit descriptions of motion events. The semantic elements of all motion expressions in the narratives were analyzed in terms of their spatial relational meaning. The annotation of semantic components was based on the linguistically-motivated ontology of the Generalized Upper Model spatial

extension (GUM-Space [1]). GUM-Space describes the semantics of spatial terms and the relation between the concepts underlying linguistic expressions of space. Given that spatial language exhibits extreme flexibility, the study shows how the problem of manner and path encoding can be addressed by analyzing motion expressions of different complexity on the basis of concepts and relations described in GUM-Space. This illustrates that a linguistic annotation scheme based on an ontology offers a valuable perspective on debates over the typology of motion events.

As a second example, relating to Reshöft's research in interesting ways, Tommaso D'Odorico presented his work on detecting events in video data using a formal ontology of motion verbs. The aim of this research is the formalization of an ontology for describing the physical world, with a particular focus on vague concepts and a specific set of motion verbs. Ultimately, automatic reasoning systems should be able on this basis to detect occurrences of such motion verbs in video sequences [2]. The input is a structured description of objects' type, shape, and position over time in the sequence (either manually annotated or automatically produced by vision trackers). These constitute the grounding of the ontology's lower-level predicates that allow mid- and high-level concepts to be logically inferred.

Some of the concepts in this ontology include verbs (e.g., *move*, *pick up*, *exchange*, *arrive*, *receive*, *walk*, *run*, *hold*), spatial prepositions (e.g., *near*, *far*, *behind*, *above*, *between*), and adjectives (e.g., *small*, *big*, *fast*, *vertical*). The problem researchers encounter when trying to formalize such expressions sourced from natural language is the issue of vagueness. For instance, consider nouns: given a spatial environment, how can one precisely define a mountain? In more formal words: how can one precisely set the boundaries of applicability of the concept mountain, or otherwise establish the spatial regions where the concept mountain holds and the regions where it does not? Alternatively, moving to adjectives or spatial prepositions: how can one precisely establish whether someone is tall, or whether two objects are near or far with respect to each other?

To meet these formalization challenges, linguistic evidence concerning the most salient characteristics of concepts is an important contribution. Such evidence is useful, for instance, in discriminating the smooth and ambiguous transition between groups of semantically close concepts. For example, given the verbs *walk* and *run*, we would like our system to infer whether a person moving in space is performing one action or the other. One of the possible strategies to tackle this issue is to organize concepts in a hierarchy, starting from the most general to the most specific. Linguistic data analysis is a key contributor in shaping up such a classification.

5 Conclusion

The exchange of practical experience between established researchers and postgraduates at the beginning of a promising scientific career can be extremely rewarding. The success of the current tutorial can be attributed to a number of factors: that the Spatial Cognition conference habitually brings together researchers with a clearly joined focus on related issues; that language is extremely prominent in cognitive science research; and that the tutorial's goals and issues were sufficiently open to be readily enhanced by the participants' personal viewpoints and established techniques. We conclude that such tutorials can offer an excellent opportunity to learn from each other across diverse research perspectives,

going beyond the traditional (uni-directional) teaching of one particular methodology or theory.

Acknowledgments

We thank all participants of this tutorial for their inspiring contributions, not all of which could be represented in this report. Funding by the DFG for the SFB/TR 8 Spatial Cognition is gratefully acknowledged.

References

- [1] BATEMAN, J. A., HOIS, J., ROSS, R. J., AND TENBRINK, T. A linguistic ontology of space for natural language processing. *Artificial Intelligence* 174, 14 (2010), 1027–1071. doi:10.1016/j.artint.2010.05.008.
- [2] D’ODORICO, T., AND BENNETT, B. Detecting events in video data using a formal ontology of motion verbs. In *Spatial Cognition (2012)*, C. Stachniss, K. Schill, and D. H. Uttal, Eds., vol. 7463 of *Lecture Notes in Computer Science*, Springer, pp. 338–357. 10doi:1007/978-3-642-32732-2_22.
- [3] ERICSSON, K. A., AND SIMON, H. *Protocol Analysis: Verbal Reports As Data*. MIT Press, Cambridge, MA, 1984.
- [4] HERTZBERG, C., WAGNER, R., BIRBACH, O., HAMMER, T., AND FRESE, U. Experiences in building a visual SLAM system from open source components. In *Proc. IEEE International Conference on Robotics and Automation (ICRA) (2011)*.
- [5] KRIPPENDORFF, K. *Content Analysis: An Introduction to Its Methodology*, 2nd ed. SAGE Publications, London, 2004.
- [6] MAZMAN, S., AND ALTUN, A. Modeling cognitive strategies during complex task performing process. *Turkish Online Journal of Qualitative Inquiry* 3, 4 (October 2012), 1–31.
- [7] MENEGHETTI, C., DE BENI, R., GYSELINCK, V., AND PAZZAGLIA, F. Working memory involvement in spatial text processing: what advantages are gained from extended learning and visuo-spatial strategies? *British Journal of Psychology* 102 (2011), 499–518. doi:10.1111/j.2044-8295.2010.02007.
- [8] MENEGHETTI, C., PAZZAGLIA, F., AND DE BENI, R. Spatial mental representations derived from survey and route descriptions: When individuals prefer extrinsic frames of reference. *Learning and Individual Differences* 21 (2011), 150–157. doi:10.1016/j.lindif.2010.12.003.
- [9] SLOBIN, D. The many ways to search for a frog: Linguistic typology and the expression of motion events. In *Relating Events in Narrative (2004)*, S. Strömquist and L. Verhoeven, Eds., vol. 2, Lawrence Erlbaum Associates, pp. 219–257.

- [10] TALMY, L. Path to realization: A typology of event conflation. In *Proc. Seventeenth Annual Meeting of the Berkeley Linguistics Society* (1991), pp. 480–520.
- [11] TENBRINK, T. CODA: Kognitive Diskursanalyse. In *Linguistik im Nordwesten: Beiträge zum 1. Nordwestdeutschen Linguistischen Kolloquium* (2010), E. Ruigendijk, T. Stolz, and J. Trabant, Eds., Bochum: Brockmeyer.
- [12] TENBRINK, T., BERGMANN, E., AND KONIECZNY, L. Wayfinding and description strategies in an unfamiliar complex building. In *Proc. 33rd Annual Conference of the Cognitive Science Society* (2011), L. Carlson, C. Hölscher, and T. Shipley, Eds., Cognitive Science Society.
- [13] TENBRINK, T., COVENTRY, K., AND ANDONOVA, E. Spatial strategies in the description of complex configurations. *Discourse Processes* 48 (2011), 237–266. doi:10.1080/0163853X.2010.549452.
- [14] TENBRINK, T., AND FREKSA, C. Contrast sets in spatial and temporal language. *Cognitive Processing* 10 Supplement 2 (2009), S322–S324. doi:10.1007/s10339-009-0309-4.
- [15] TENBRINK, T., AND GRALLA, L. Accessing complex cognitive processes via linguistic protocol analysis. In *Proc. KI 2009 Workshop Complex Cognition* (2009), U. Schmid, M. Ragni, and M. Knauff, Eds.

Thora Tenbrink

School of Linguistics & English Language, Bangor University, UK

Tommaso D'Odorico

School of Computing, University of Leeds, UK

Christoph Hertzberg

Faculty of Mathematics and Computer Science, University of Bremen, Germany

Güzin Mazman

Computers Education and Instructional Technology, Hacettepe University, Turkey

Chiara Meneghetti

General Psychology Department, Padova University, Italy

Nina Reshöft

Faculty of Linguistics and Literary Sciences, University of Bremen, Germany

Jinlong Yang

Department of Geography, GeoVISTA Center, The Pennsylvania State University, USA