Proceedings of the Second IASTED International Conference
HUMAN-COMPUTER INTERACTION
March 14–16, 2007, Chamonix, France
ISBN Hardcopy: 978-0-88986-654-6 / CD: 978-0-88986-655-3

EVALUATION OF INFORMATION VISUALIZATIONS VS. LANGUAGE PROFICIENCY

Olaf Thiele
Department of Information Systems
University of Mannheim
Germany
email: thiele@uni-mannheim.de

Dieter Thoma
Department of English Linguistics
University of Mannheim
Germany
email: thomad@uni-mannheim.de

ABSTRACT
In this paper we will ask the question of what proficiency in a textual language (e.g. English) has in common with evaluating information visualizations. We admit that this question might not seem completely straightforward. Therefore we will first present how experimental practice led us to this question. We hypothesize that the ability to use a visualization (to speak the visual language) affects both objective and subjective evaluation methods. We propose to transfer the use of language testing and grammar description from linguistics to human computer interaction evaluation.

KEY WORDS
information visualization, evaluation, language, usability, user interface, mobile computing

1. Introduction

In our contribution we want to emphasize the importance of looking at information visualizations as visual languages that have many similarities with traditional textual languages. We argue that the evaluation of human computer interfaces requires (1) an explicit description of the grammar of the visual language used. And once the system has been described, it is necessary (2) to assess to what extent users have acquired the system. Our line of research originates from a 132-participant, laboratory-based study which consisted of a complex sequence of experimental tasks. The study started out as a cooperation between an information systems researcher and a linguist who joined forces to recruit and finance a relatively large participant sample. The information systems part was concerned with the evaluation of an information visualization on the grounds of the Technology Acceptance Model (TAM [8]) in combination with the Technology Readiness Model (TRI [17]). The aim was to simulate a visually elaborated presentation of news texts in English using Treemaps. The linguistic part engaged in the validation of novel test methods for business English vocabulary knowledge of second language learners. When we analyzed our respective data which showed only a very weak relationship between the two studies, lengthy discussions evolved on how and why users' acceptance of information visualization technology might be influenced by their language proficiency.

A persistent problem of information visualization technology is that it is, despite its demonstrably higher efficiency [5], not widely used in today's world. More recently, evaluation studies of information visualization systems aim at exploring more subjective factors like technology acceptance. Still, textual representations are by far more popular. A textual language is an abstract system 'invented' for communication, but it is not the only and not always the most efficient tool. After all, language symbols are arbitrary, increasing the cognitive load in information transmission substantially as compared to more direct visual impressions. As compared to spoken language, written language is a further abstraction. The processing of written language can, however, become a highly automatic cognitive activity. On the other hand, we search for information for the purpose of enhancing our knowledge. Mental models that represent knowledge are generally understood as concepts [21]. Thus, both textual and visual languages serve the same purpose: to convey knowledge. Some prominent differences and similarities between both are discussed in this paper. Next we present the experiment that led to our ideas. The third section presents the concept of visual languages. The impact of visual languages on information visualization evaluation is discussed in the fourth section. We conclude with final thoughts on our work.

2. Background Experiment

2.1 Method

Participants were 132 university students (85 women and 47 men) aged between 18 and 31 years (M = 23.5, SD = 2.4). All participants had learned English as a second language for at least six years on secondary school level. All six tasks were conducted online at three separated desks each supporting an Intel Pentium IV computer, a 19-inch monitor, a keyboard, a mouse, and a set of headphones. As an incentive they were told that they received 5 Euro in any case, but that they could earn up to 10 Euro depending on their performance on the tasks. The two main studies were a language proficiency test and the information visualization test. The following two subsections describe them in more detail.

2.2 Language Proficiency Test

The language proficiency test we report on here is a test for vocabulary size in business English. The test contains 150 items and the participants are to indicate whether or not they know the meaning of each expression. As some testees tend to overestimate their vocabulary knowledge, 50 of the items are pseudowords. If a pseudoword is claimed to be known, the total score is adjusted downwards. Vocabulary size is the most important measure of lexical knowledge which is a good predictor of reading abilities [2]. The test is highly reliable (Cronbach's alpha$= .94$) and correlates highly with standardized language measures ($r = .752$ with Llex 10K [15]), and we consider it as a representative measure of English language proficiency for present purposes.

2.3 Information Visualization Test

The experiment evaluates a visual representation of Google News [11] using Treemaps [22]. It is meant as a mobile adaptation for the well-known Newsmap visualization [24]. Visualizations are especially useful for mobile applications because of their restrictions like limited screen size or means of interaction. Participants were asked to solve three tasks with the given visualization. The first task was meant to familiarize the user with the interface for at least two minutes. No further instructions were given except for the help to the right side of the news map. The second task involved finding one news item from each of the four thematic news areas. The third task showed the actual news map of the respective day. Participants had no time limit for exploring today's news.

Figure 1 shows a screenshot of the user interface (the visualization is available online at [1]). The visualization was displayed in the middle of the screen with a PDA-like display resolution of 300x300 pixel. Instructions for the task were shown to the left, helpful information about the interface to the right. Generally, the visualization shows four thematic news areas: business, technology, world and entertainment news. The more online resources report on a news item, the bigger it is. And the older messages are, the more shaded they are presented in the visualization.

2.4 Results

Contrary to our expectations, we could find no systematic relationship between textual and visual language proficiency in our experimental setting. Results show only a very weak relationship between the proficiency measured by the test and the time the participants needed to complete the information visualization task ($r = -.209, p < .05$). This objective measure indicates that participants with a high level of language proficiency perform only slightly faster in the visualization task. Subjective measures of the relationship between language ability and the acceptance of the new technology pointed into the same direction. Although the visualization task eventually required knowledge of written English language, users with poor language proficiency did not prefer or reject the visualization task any more than participants highly proficient in textual language use. In the visualization task, the access to information is only possible via language. Put differently, the acquisition of conceptual knowledge is here mediated by linguistic knowledge. The crucial point is that the information access requires two kinds of language knowledge. The access to textual language, which eventually encodes the conceptual information, is mediated by knowledge of the 'language' of the information visualization - the visual language. In practice, users could only find and read the texts if they understood the 'grammar' of the visual language. Based on this reasoning, we think that it is plausible to hypothesize about a language-based distinction: Users' proficiency in the respective visual language can affect evaluation methods for information visualization because it can explain variation in objective measures such as time and errors and subjective measures like adoption. For example, we hypothesize that users untrained in the comprehension of the visual language will be more likely to refrain from actually using the technology. This is because they find it awkward to 'read' in a language they do not know properly, irrespective of the fact that it might actually take much longer to do the reading in a textual language they know well.

We argue that information visualizations can be seen as visual languages. Therefore, the field of human computer interaction evaluation might benefit from linguistic research. Within this work, we focus on two main ideas: (1) the possible influence of our assumptions on existing evaluations, and (2) how evaluations might be improved. Basic properties of both textual and visual linguistics are shown in the following section.

3. Textual vs. visual languages

From a Chomskian perspective, knowing a language means that we have a mentally represented grammar of it. This grammar is basically composed of words (lexicon) and of rules for combining these words into sentences (syntax) [6]. Both individual words (or parts of words) and their systematic combination convey meaning, i.e. they are used to label concepts. The grammar is by its nature implicit and therefore linguists attempt to describe and explain it explicitly, for example, in order analyze how languages are learned. Applied to information visualizations we could argue that these technologies make use of elements which are structurally similar to those of languages: They employ minimal units that have a meaning on their own and rules for combining these units into larger chunks of meaning. These elements form the grammar of a visual language. A dif□culty in the evaluation of information visualizations is that visual grammars are usually implicit. They are surely guided by various principles, but these principles are often fairly ig-

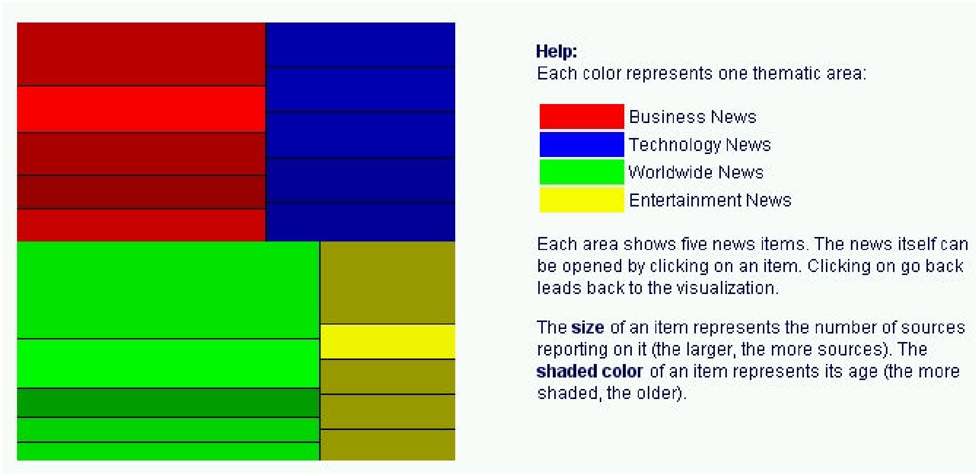

Figure 1. Information visualization test

norant of the ultimate language user. If visual grammars were made explicit, this would improve the visualization and could help evaluating it.

From the perspective of causality, there is a common view in cognitive psychology and linguistics that concepts come first and language merely labels them and that, somewhat as a consequence, language and general cognition are separate, modular systems (see [3], [10], [13] or [19]). This position is relatively commonsensical and deeply rooted in the empiricist tradition. Strong versions of the opposite position of 'linguistic relativity', i.e. that knowing a (particular) language determines and limits conceptual structures (see [26], [25], [18]) encounter many difficulties, most of them related to the fact that language apparently underspecifies the conceptual contents it labels. This becomes explicit in cases of linguistic ambiguity, paraphrase, and deictic reference where language users need to employ extralinguistic inferential processes to retrieve the conceptual information. For the present discussion this means that a visual language is just as any other language one relatively independent way of accessing conceptual knowledge.

It is very likely that humans are genetically predisposed to learn languages in childhood [7]. Healthy children can learn any language in only about 2.5 years equally well at an age where they cannot handle anything comparable to language in terms of its mental complexity, and they can do this despite the insufficient and erroneous input they get. Language learning in adults is different. Although it is not yet clear in how far they are supported by innate principles, it is clear that they start off with different preconditions. For a start, they already know one language, they are cognitively mature, and they are influenced by motivation and other socio-psychological factors. This means that if they are confronted with a new language, in our case with a visual language, theoretically they will look for similarities between this language and the language(s) they already know. They have a good deal of experience in how to learn and will be able to internalize the grammar of a new language explicitly. Of course, this requires an explicit representation of the grammar. If, for example, we traveled to China without any prior knowledge of the language, we would still assume that when a local talks to us the "stream of noise " we hear can be separated into words and sentences. Applied to our context, news items which are represented by bright and big shapes are recognized as more important because we know this is true for traffic signs.

Some open questions in information visualization evaluation might be explained by our approach. Chen, for example, states that [4, p. 210]: "Much of the existing empirical studies can be divided into ones that deal with lower-level elementary perceptual tasks or higher-level application-related tasks." Chen suggests to intensify research on lower-level tasks to begin with. This corresponds with typical language learning. A large vocabulary and a fluent application of syntactical knowledge aid in speaking a language well. There are, however, different types of language learners: some with a greater reliance on explicit rule learning and others with more implicit strategies. For the first group language learning is more effective if the work through a grammar book, while the second group will benefit more from a stay abroad. This phenomenon is typical for learning textual languages, but as visual languages are usually designed and implemented by technicians it is often not thought of. In our experiment, participants had the choice to study the structure of the information visualization by reading the detailed description next to the interface (first group) or to simply explore the workings of the visualization (second group).

As for grammar, Cleveland and McGill tried as early as 1984 to identify a possible vocabulary set for information visualizations [4, p. 186]. They proposed rules for using the vocabulary (e.g. positions along a common scale or length and height) and thus created a grammar. Nevertheless, only few other research exists on a possible visual grammar. The issue becomes even more difficult because many problems are generally solved through individualized visual languages. Still, only a managable number of visualizations exists and language classifications are in our view achievable. Morse and Lewis already tried to define a plain taxonomy in their study [16]. Plaisant asks for the development of repositories that could serve for other studies [20].

Yet another approach attempts to create a universal visual grammar from well-known visualizations. Apart from scatter plots or visualized trees this line of research tries to tell stories by more simple visual means. Arrows and comic-like elements are used together with textual elements to form a visual language [12]. Horn tries to combine charts (similar to mind maps) to produce coherent maps of knowledge. Similar to Horn's approach McCloud describes how comics are told and what grammar is used to convey information [14].

While we just tried to show what constitutes visual languages, we will now seek the connections to information visualization evaluations.

4. Influence of linguistics on new forms of evaluations

Evaluations of information visualization judge the performance of the software on some criteria of merit which are then compared to standards of performance. Most evaluations of information visualizations focus on objective criteria like task completion time or variables like recall and precision [5]. Some more recent studies aim at including subjective measures like user acceptance. The Technology Acceptance Model seems to be especially popular (see [23] for an overview). The intriguing problem here is that the measurement of the performance of the information visualization may be confounded by the participants' heterogeneous abilities. Following our language distinction, objective evaluations *also* measure how well participants know the visual language and subjective evaluations *also* measure whether they like using it.

Objective evaluations center around how well participants in an experiment know the visual language. But methodologically we cannot distinguish between an insufficient language ability and the ability to solve the task using the given interface. Just as it is dif□cult to understand an unnecessarily complex and badly formed English sentences even if you know the language well. In our experiment we cannot clearly distinguish if participants just do not know that bright and big shapes signal importance or if our representation of importance in the visualization is difficult to understand and unattractive. Therefore we propose to bring in language tests to identify participants on different proficiency levels. This would enable evaluators to unveil the artificial homogeneity of their participant sample. They could thus explain a greater proportion of the variation in their objective or subjective data. We want to illustrate this idea with three possible situations within an evaluation in figure 3 (curve progression follows general learning curves [9]).

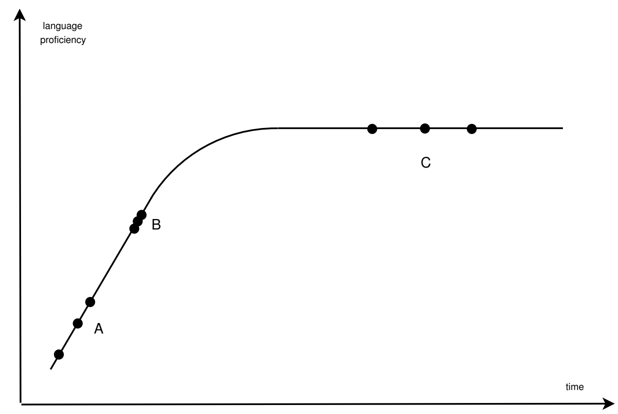

Figure 2. Possible language proficiency situations

Three participants are on different proficiency levels in situation A. An objective evaluation at this point might be interrelated with testees' respective language ability. Preferably, while still in the learning phase of the visualization, participants are all at the same proficiency level as displayed in situation B. This could be verified by a test. Ideally, all participants know the language fluently as depicted in situation C. This would be the best time to start the experiment. On the downside, testees might get bored if they need to attend unnecessary training prior to the experiment. This might influence their subjective evaluation results.

Subjective evaluations could in our view be part of information visualization evaluations. For example, if solely objective criteria led to car sales, most car manufacturers would be selling low fuel engines. But the opposite is the case. Transferred to information visualizations the Technology Acceptance Model aims at measuring factors like ease of use or usefulness. The model can easily be extended. Is an interface enjoyable (see for example [23]). Furthermore, whether a testee likes the method of learning the interface could affect the evaluation results.

Novel evaluation methods could be derived from classical language proficiency tests. Two classical test types can be distinguished: discrete and integrative tests. Discrete tests check whether the learner knows the elementary blocks of the visualization. We could check whether participants understand that larger rectangles symbolize more popular items in our news map visualization. The main advantage of this approach lies in the generalizability of its results, i.e. they are relatively independent of the specific test situation. Moving objects, for example, are naturally detected and followed by the eye. Whether this has the desired overall effect in a visualization technology cannot be evaluated properly by discrete test items. An appropriate alternative would be integrative tests that incorporate the whole language use situation. These methods attempt to make the test situation as authentic as possible by closely simulating the actual task. Integrative tests yield an overall picture of language proficiency, but it is more difficult to reach conclusions that are independent of the test situation. This test type is especially useful for measuring the fit of the visual language to solve a given problem.

Language proficiency and usage of certain vocabulary changes over time in textual languages. This probably holds true for visual languages as well. The only way to track and analyze such developments are longitudinal studies. In our case a potential longitudinal study would be to make the visual grammar explicit and to train participants in using it.

5. Conclusion

A large-scale experiment in which we found only a very week relationship between language proficiency in a written language and the ability to use as well as the readiness to accept an information visualization task lead us to the question of the role of the knowledge of the 'grammar' of visual languages in the completion of such tasks. We argued that such knowledge can be a confound which affects the evaluation on the grounds of both objective and subjective measures. We do not imply that knowledge of visual languages is the only potential confound. Rather we propose that evaluations of information visualizations should increasingly head towards the direction of more molecular forms of causation. If we limit ourselves to the design features of the interface as independent variables and their influence on the performance the user achieves with the help of the interface or how much he or she enjoyed working with it, we might ignore considerable sources of variation caused by the language abilities of the user.

More practically, Chen states that "the empirical link between visual attributes and perceptual tasks is still missing" [4, p. 210]. The visual language point of view presented in this paper might bring both closer together. Finally, some other aspects of written textual languages might be transferred to information visualizations: Ambiguities and jokes, for example, are an integral part of textual languages - so why are there none in the languages of information visualization?

References

[1] *DataRug Visualization*. available online: http://www.olafthiele.de/research/, 2007.

[2] R. Anderson and R. Freebody. Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading Language Research 2*, pages 231–256, 1983.

[3] K. Bühler. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Orignal publication: Jena, Fischer, 1934. Stuttgart: Lucius und Lucius, 1999.

[4] C. Chen. *Information Visualization*. Springer, 2 edition, 2004.

[5] C. Chen and Y. Yu. Empirical studies of information visualization: a meta-analysis. *Int. J. Hum.-Comput. Stud.*, 53(5):851–866, 2000.

[6] N. Chomsky. *Reflections on language*. New York: Pantheon, 1975.

[7] V. Cook. *Inside Language*. London: Arnold, 1997.

[8] F. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quaterly*, 13:319 – 339, 1989.

[9] H. Ebbinghaus. *On memory*. New York: Teachers College Press, 1885.

[10] J. Fodor. *The modularity of mind: an essay on faculty psychology*. Cambridge MA: MIT Press, 1983.

[11] GoogleInc. *Google News*. available online: http://news.google.com/, 2007.

[12] R. E. Horn. Visual language and converging technologies in the next 10-15 years (and beyond). In *National Science Foundation Conference on Converging Technologies*, Dec. 3-4, 2001.

[13] W. J. Levelt. *Speaking: from intention to articulation*. Cambridge MA: MIT Press, 1989.

[14] S. McCloud. *Understanding Comics: The Invisible Art*. Northampton, MA, Kitchen Sink Press, 1993.

[15] P. M. Meara. English vocabulary tests 10k. llex 10k vocabulary tests (paper-and-pencil version). *Swansea: Centre for Applied Language Studies*, 1996.

[16] E. Morse, M. Lewis, and K. A. Olsen. Evaluating visualizations: using a taxonomic guide. *Int. J. Hum.-Comput. Stud.*, 53(5):637–662, 2000.

[17] A. Parasuraman. Technology readiness index (tri), a multiple-item scale to measure readiness to embrace new technologies. *J. Service Research*, 2(4):307–320, 2000.

[18] E. Pederson, E. Danzinger, D. Wilkins, S. Levinson, K. Stephen, and G. Senft. Semantic typology and spatial conceptualization. *Language*, 74(3):557–589, 1998.

[19] S. Pinker. *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press, 1989.

[20] C. Plaisant. The challenge of information visualization evaluation. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 109–116, New York, 2004. ACM Press.

[21] M. Schwarz. *Einführung in die kognitive Linguistik*. Tübingen: Francke, 1992.

[22] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11(1):92–99, 1992.

[23] O. Thiele and G. Mau. Acceptance of visual search interfaces for the web. In *ISVC 2005*, volume 3804 of *Springer LNCS*, pages 263–270.

[24] M. Weskamp. *newsmap visualization*. available online: http://www.marumushi.com/apps/newsmap/, 2007.

[25] B. L. Whorf. *Language, thought and reality*. Editor: John B. Carroll, Cambridge, MA: MIT Press, 1956.

[26] L. Wittgenstein. *Tractatus logico-philosophicus*. Editor: David Pears, London: Routledge, 1981.