User experimentation with terminological ontologies

Louise Pram Nielsen
Department of International Business Communication
Copenhagen Business School
Denmark
lpn.ibc@cbs.dk

Abstract

This paper outlines work-in-progress research suggesting that domain-specific knowledge in terminological resources can be transferred efficiently to end-users across different levels of expertise and by means of different information modes including articles (written mode) and concept diagrams (graph mode). An experimental approach is applied in an eye-tracking laboratory, where a natural user situation is replicated for Danish professional potential end-users of a terminology and knowledge bank in a chosen pilot domain (taxation).

1 Introduction

Modern lexicography and terminology are converging (Cabré, 1999). Traditionally, terminology and lexicography have been separate research fields with different approaches to compilation and presentation of data. However, modern technology offers unlimited opportunities to meet the needs for several target groups in one database by offering the possibility of choosing between different presentations, in theory, providing means for knowledge transfer across different information modes.

Madsen and Thomsen (2008) argue that systematic terminology work ensures consistency across the entries of a given database (ISO 704:2009). This improves the quality of the information tool considerably compared to other types of specialized reference works. Thus, the end-user is presented with consistent information in a *written mode* representing the terminological information such as definitions, synonyms, equivalents and sources. In practice, however, concept clarification usually takes a more graphical starting point, in particular, when terminology and knowledge banks follow the principles of terminological ontologies (concept systems) previously discussed by Madsen and Thomsen (2008). Here terminological ontologies are defined as domain-specific ontologies where certain aspects of terminology theory have been formalized: Characteristics are modeled by formal feature specifications (attribute-value pairs) and subdivision criteria that correspond to the attributes of the feature specifications. In other words, terminologists may structure their knowledge by means of concept systems, then develop consistent definitions, and subsequently process them into article-like entries using a knowledge management tool (Madsen, 1999). This implies that the graphical structuring is central to terminology method and theory, and end-users should have access into the underlying ontology or concept system (*graph mode*) as a complementary source of knowledge.

This short paper presents experiments with the primary purpose of exploring Danish end-users' understanding of concept systems. The preliminary results imply that domain-specific knowledge in terminological resources can be transferred efficiently to users across different levels of expertise and by means of different information modes. The paper outlines work-in-progress research and will focus on the information mode variable. The paper is organized as follows: in section 2 the method is outlined; in section 3 the eye-tracking experiments are described, and in section 4 the preliminary results suggested by the experiments are presented followed by a conclusion in section 5.

2 Method

An experimental approach is applied with Danish professional potential end-users participating in an eye-tracking experiment.

2.1 Information modes

Terminology work is concept-oriented (ISO 704:2009; Madsen, 1999), which means that synonyms are registered in one entry in the database, while lexicography is word-oriented, i.e. one dictionary entry comprises all meanings of an entry-word. With the use of databases, however, the possibilities for presentation do no longer depend on the structure of the data collection, and thus it is possible to present data from a terminological ontology with a concept-oriented structure in a word-oriented user interface. Terminology resources contain lexicographic (written) information such as definitions, sources, synonyms and equivalents, but the terminology work offers a complementary graphical information mode displaying the concept position and relations to other concepts (ISO 704:2009). Therefore, both information modes carry knowledge that can be transferred to end-users.

2.2 Experiment stimuli

Concept diagrams carry the same amount of information across eight blocks. Each block represents a taxation term: direct tax; land tax; middle-bracket tax; personal income tax; energy tax; excise duty; green tax; motor vehicles tax. The stimulus template is shown in Figure 1.

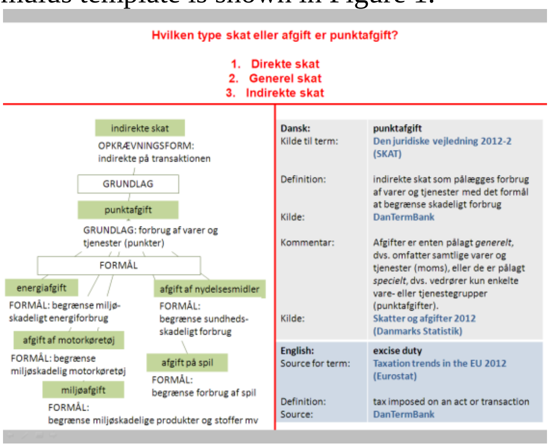

Figure 1. Stimulus template

Concept diagrams include 5-7 concepts structured in three levels filling out half of the screen. Articles are equally sized presenting more detailed information (term, definition, equivalent, and comment including sources of both the Danish source language and the English target language). The eye-tracking stimuli are thus constructed to provide participants with a double-mode design and each stimulus comprises three primary areas of interest (AOIs): the AOI-question (at the upper part of the screen); the AOI-diagram (placed below); the AOI-article (placed on the opposite side of the diagram). Figure 1 shows an example of excise duty and the question translated into English is "What type of tax or duty is excise duty? 1. Direct tax; 2. General tax; 3. Indirect tax."

In the experiment, the article entry and concept diagram of the stimuli are static images. In the real system, users should be allowed functionalities to unfold concept diagrams and articles further. The stimuli of the eye-tracking experiment can be seen as design artefacts closely resembling articles and concept diagrams of an existing knowledge management tool *i-Term* (DANTERMcentre). This approach constitutes a usual starting point of human work interaction design (Clemmensen, 2011). In addition, it is being assumed that users have entered the terminology and knowledge bank correctly and found the concept represented in the relevant diagram (graph mode) or article (written mode) necessary for concept clarification.

2.3 Question types

The experiment begins with a reading task using a specialized text in participants' first language (Danish). Then 48 multiple-choice concept-clarifying questions (trials) resulting from six types of questions about concept clarification pertaining to each of the eight chosen domain-specific terms (blocks) are posed. Questions can be answered by consulting information in either one of the information modes, or the answer lies in both: The six question types include: sub-ordinates (First diagram-based question); sub-division criteria (Second diagram-based question); equivalents (First article-based question); comments (Second article-based question); super-ordinate (First diagram- and article-based question); characteristics (Second diagram- and article-based question). The six question types are randomly distributed across the eight blocks. The eight blocks are also randomized, and so is the display-side of the information modes (diagram to the right and article to the left side of the screen or vice versa).

3 Eye-tracking experimentation

The sample comprises 40 Danish professional potential end-users of the terminology and knowledge bank in the taxation domain.

3.1 Experimental design

An experimental approach is applied in an eye-tracking laboratory, where a natural user situation is replicated. The approach is guided by the triangulation principle resulting in both quantitative and qualitative data (Holmquist et al., 2011) that will contribute to the understanding of professional end-users' performance and perception which contribute to the subsequent interface design process, in particular, for the development of personas and scenarios (Nielsen, 2002).

Prior to the experiment, participants' domain-specific expertise is measured in a combined assessment comprising self-assessment and a test revealing their declarative knowledge in the taxation domain. In particular, participants are asked to fill out a background questionnaire comprising a declaration of consent, background information (age, gender, education, industry, typical tasks during their professional working day), introduction to concept clarification and a terminology warm-up exercise.

During the experiment, participants are asked multiple-choice questions pertaining to concept clarification in the taxation domain, while they are presented with the double-mode stimuli and their eye-movements are being recorded. A remote SensoMotoric Instrument (*SMI*) eye-tracker, which supports gaze sampling rates of 50 Hz, is used for the recordings of participants' on-screen eye-movements. The experiment is built in the psychology software *E-prime*, which facilitates randomization, records user responses, and informs participants whether they answered correctly or not.

After the experiment, a retrospective interview is conducted with the participants. Here they evaluate their performance, preference and needs pertaining to concept clarification, including their use of taxation texts, in their work. In total, the experiment lasts about one hour.

3.2 Sampling across expertise

When compiling specialized dictionaries, it is necessary to distinguish between different types of users, i.e. experts, semi-experts and laymen (Gouws, 2009). Therefore, it has been crucial in the sampling of participants for the eye-tracking experiment that they represent different levels of expertise ranging from high expert to low non-expert level. In the sample, half of participants are staff members from *The Central Customs and Tax Administration (SKAT)* working in the taxation domain as e.g. legal advisers, economists, software developers, business analysts, communicators, translators or generalists. The remaining participants are professional staff members from e.g. private companies, universities and other government organizations. All participants have Danish as their first language, and all questions and concept diagrams are in Danish.

The background questionnaire primarily assesses declarative knowledge skills in the taxation domain, whereas the eye-tracking experiments also require procedural knowledge or logical reasoning skills. Expertise variables should reflect the expertise needed in the experiments. In order to overcome any discrepancy between the declarative nature of the assessment and the procedural nature of the expertise needed, the expertise assessment also comprise participants' information seeking skills and their weekly number of electronic searches in search engines, terminological resources such as encyclopedia, dictionary or term banks.

3.3 Eye-tracking measures

In eye-tracking research, the recorded eye movements are analyzed by means of detecting events, i.e. measures accounting for scan paths (where do participants look and do they revisit AOIs) and fixation duration (what do participants look at and do they fixate on AOIs) (Holmquist et al., 2011). In particular, the eye-to-mind-hypothesis (Just and Carpenter, 1980) uses eye-movements (fixations) to indicate the cognitive effort needed to process and understand stimuli.

4 Preliminary results

Preliminary results reveal a "learning effect" which reduces the response times of participants across the 48 trials without reducing the relative number of correct answers. Moreover, high relative average fixation duration per trial in the AOI-diagram and the AOI-article on diagram and article questions respectively, suggests that users "know" where to look for answers and can access information in the graph mode.

Potential interactions were observed during the experiments, but need further testing as part of the inferential statistical analyses:

Participants assessing their level of expertise to be high on the expertise measure (experts) are quite critical towards the stimuli. Experts have a high success rate, but they are sometimes confused by the simplified concept diagrams and article entries of this experiment and express verbally their disagreement. In addition, experts express a high preference for the detailed and precise articles compared to diagrams, which they might be confused by. Once they have learned to navigate the experiment, they start appreciating the advantages of diagrams, especially if they were new to a field, including the double-mode interface design.

Participants assessing their level of expertise to be low on the expertise measure (non-experts) have hardly any opinion on the taxation domain. Non-experts tend to be overwhelmed by the complex taxation domain and spend quite a long time understanding the questions and information modes. However, if the long response time is disregarded, non-experts perform quite well. The learning effect also applies to non-experts who learn to navigate the information modes across the experiment.

An inherent impatience is revealed during the experiments. It seems to lead participants to fuzzy scan paths and random guesses if they do not locate an answer inside the stimuli space. This impatience is due to the time pressure that participants feel they are performing under and the fact that the answer "do not know" is not available to them. Opposed to the inherent impatience, which tends to shorten the response time, an inherent insecurity tends to prolong the response time. The inherent insecurity makes participants search for answers they already know or have already found.

5 Conclusion

Despite the inherent drawbacks due to the experimental design, it can be concluded that domain-specific knowledge is transferred across written and graph modes to both experts and non-experts. It should be noted that complete descriptive and inferential statistical analyses are in progress. In addition, the eye-tracking experiments constitute a first step, which needs to be followed by future research on a dynamic system version offering participants the possibility of interacting with the article entries and concept diagrams of the terminology and knowledge bank.

Acknowledgments

The research is funded by the VELUX FOUNDATION and constitutes sub-project three of the DanTermBank project which aims at developing the foundations for the establishment of a terminology and knowledge base in Denmark.

References

Bodil Nistrup Madsen. 1999. Terminologi 1. Principper og metoder. Copenhagen, Gads Forlag.

Bodil Nistrup Madsen and Hanne Erdman Thomsen. 2008. Terminological Principles Used for Ontologies. In: Proceedings from the International Conference on Terminology and Knowledge Engineering. Managing Ontologies and Lexical Resources, pp. 107-122.

DANTERMcentre. DOI= http://www.iterm.dk/.

ISO 704 :2009. Terminology work. Principles and methods. International Organization for Standardization.

Kenneth Holmquist; Marcus Nyström; Richard Andersson; Richard Dewhurst; Halszka Jarodzka; Joost Van De Weijer. 2011. Eye tracking. A comprehensive guide to methods and measures. Oxford University Press.

Lene Nielsen. 2002. From user to character: an investigation into user-descriptions in scenarios. In Proc. DIS2002, ACM, 99-104.

M. A. Just and P. A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension, Psychological Review 87(4), 329-354.

Maria Teresa Cabré. 1999. Terminology. John Benjamins Publishing Company.

Rufus H. Gouws. 2009. Integrated dictionary use of specialised dictionaries for learners. Lexikos. 19:72-93.

Serge Verlinde, Patrick Leroyer, and Jean Binon. 2010. Search and You Will Find. From Stand-Alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. International Journal of Lexicography vol. 23 (1): pp. 1-17.

Thorkil Clemmensen. 2011. Designing a simple folder structure for a complex domain. Human Technology, 7(3), 216-249.