

Linguistic and Cultural Differences in Content Management - INDEXING AND TITLING IN MULTILINGUAL AND MULTICULTURAL BLOGOSPHERE

The final paper reports part of the results of an on-going study, which topic is linguistic and cultural differences in multilingual and multicultural blogosphere. The preliminary results with the Finnish language material clearly show how the idea of indexing diverges partly fundamentally between information professionals and blog-writers.

In the abstract the theoretical background as well as motivations for the study and some general viewpoints are discussed.

1 Background and significance

To convey meaning we need to communicate, use language, and this is not always an easy task in a globally united yet linguistically and culturally separated world. Libraries and tools for information documentation and retrieval are not only living a new renaissance, but also encountering new problems. How to provide relevant information in today's global life? What are the obstacles? How do we face them?

In the World Wide Web (WWW) information is commonly published by other than professionals on knowledge management, and the content is in its character mostly so called social media and disseminated by others than traditional publishing channels. At the same time the border between writers and non-writers, and fiction and non-fiction has become blurred. In the information science Library 2.0 era is discussed (Widén-Wulff & al. 2008, 842).

The starting point for the definition of '**blogs**' (also known as *weblogs*) is often form instead of content (see e.g. Blood 2000; Gillmor 2004; Ojala 2005). In this study the emphasis is specifically on **subjective and text-form diary-type blogs**, which are further here referred to as blogs. As a text type the material represents web prose, which is creative, dialogical and oral like but stylized writing in weblogs, chats and discussion forums (see the term's origin in Niemi-Pynttari 2007).

English, Finnish and Russian belong to different language groups, but more important difference is that Finnish is culturally very Finnish and Russian very Russian, whereas English is globally very common *lingua franca*, i.e. used in communication between persons not sharing a mother tongue.

Traditionally language (cf. Bühler 1965) and texts (cf. Reiss 1977) are considered to serve three **basic functions** – to represent, express and appeal. The functions are bound to communication. In informative function (to represent) language is used to refer to the reality, and the topic is in the foreground of the communicative intention. In practice it is involved in every message. For example encyclopedias and lectures represent informative text type. In expressive function (to express) the emphasis is on the sender of the message, as is case in expressive text types like in poems and in drama plays. In vocative function (to appeal) the emphasis is on the receiver. For example commercials and election speeches represent vocative text type. Nevertheless, in practice texts often include characteristics of several text

types, and for instance, memoirs are close to expressive text type, but in their nature also informative. (Vehmas-Lehto 1999, 71-73) Jakobson (1981) distinguishes six functions of language, which are referential, emotive, conative, phatic, poetic and metalingual. In his opinion the functions impact in parallel resulting on the simultaneous existence of different functions in different situations and language use manners. (Ibid, 22-27)

The Jakobson's model has been especially popular in the study of reception of cultural products, and it has been used as a framework as well as a tool for analysis. It takes into consideration different kinds of contexts, is multifunctional (it is not necessary to reduce discourse into one function at a time) and with the help of it the communicative action or result of an action can be analyzed in detail. (Kovala 2005) We can assume web prose to include mainly attributes of expressive and informative text types, but its indexing may be also closest vocative or emotive, and thus not primarily to pass on informative content metalingually. In this study the focus is on emphasis of different (dominant and secondary) functions and on their tension (cf. Kovala 2005).

In the study **indexing** (documentation) is understood in coherence with the standards of the field (ISO 5963-1985, 2) as the "act of describing or identifying a document in terms of its subject content".

In this study different discourses of information storage are explored. At a general level **discourse** can be defined to refer to language use in social context (Pälli 2003, 22). It is also noteworthy to realize, that although discourse is a substantive, it means also doing something and active processes. In the center of the analysis is thus human action. (Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003).

As a research method is used **discourse analysis**, which is context-bound study of human activity (Fairclough 1992; Olsson 2004, 2). It is not reasonable to characterize discourse analysis as a distinct research method, but more as a loose theoretical framework (Potter & Wetherell 1990; Suoninen 1992, 125), in which different emphasizes in focus and methodological applications are allowed (Jokinen & al. 1993, 17-18). Discourse analytical methods have been scarce in the field of librarianship and information science research, although the interest has been growing (Hedemark, Hedman & Sundin 2005, 17). The starting points of discourse analysis have been discussed in several articles, but it has seldom been used as a method for empirical studies (Talja 1998, 18). However, discourse analysis is a natural choice, when we think of the key element in our field – information. Talja (1997, 70; 1999, 460) describes information as being concerned with what people do with language and what language does to people.

In this study blogs are analyzed as **cultural products**, since languages and the thinking they reflect mainly stem from cultural needs for expression (see Suojanen 1993; Slater 1998). Language is not static (see e.g. Varantola 1990; Aitchison 1991; Wierzbicka 1997; Lehtonen 2000; Katan 2004) and therefore also their documentation (indexing) is tied with time and the surrounding culture (Nykyri 2010).

2 Aims, material and methods

This study analyses different ways how bloggers provide textual tools for information retrieval and enables finding information from their blogs. The four language and culture pairs analyzed are American English, English as *lingua franca*, Russian and Finnish. On blogs

written in Russia and Finland even samples written in mother tongue and English are selected. The total sample is in all fifty blogs, from which ten most recent updates are collected. The analyzed material covers thus 500 text samples.

<i>Country</i>	<i>1. language</i>	<i>2. language</i>	<i>in all (50)</i>
Finland	Finnish – 10 samples	English – 10 samples	20
Russia	Russian – 10 samples	English – 10 samples	20
United states	English – 10 samples		10

Table 1: The research material (blogs) by country and language

The material collection bases on theoretical sampling (see Seale 1998, 329; Strauss and Corbin 1990, 176-179) and its criteria are diary-like content, provision of indexing terms¹ for the readers, similar search features of the used platform, and the background of the author, and the publication country and language.

This study focuses on examination of the linguistic methods, which are used to describe the content of the blogs, and the functions of the indexing. The method used is discourse analysis. Indexing is reflected to the functions and text-types found in the actual blog texts. When analyzing cultural differences the comparison pairs are Finland, Russia and United States, and in their linguistic aspects Finnish, Russian and English (as American English and *lingua franca*) are explored.

The **research questions** are:

- What kinds of functions the indexing and titles of blog texts reflect? What kinds of motives are found in the documentation of subjective blog writings?
- Are there differences in style and function of blog indexing and titling?
- Are there clear documentation discourses between specific countries versus languages versus certain background of the authors? – What kind of customs exists in the Finnish versus American versus Russian versus international material?

Indexing is reflected to the content of the blog texts, titles and search possibilities. The focus is on functions and styles. For example the geographical place of writing or reference is interesting also from the perspective of information seeking, and it can be expressed in many different ways (cf. Lindgren 2005).

3 Results

Despite of their universality and growing popularity blogs as a form of social media are still little studied (see Kjellberg 2010). The linguistic and cultural aspect of social media information seeking, described in the research plan, represents a unique approach. The topic of the research concerns directly the most challenging tasks within information science today, which is to achieve a better understanding of the fundamentals of human information behaviour, and the ways in which it changes over time (see Bawden & Robinson 2008, 9).

¹ Different blog-platform providers use different kinds of terms. For example, in Wordpress the term "tags" is used for the keywords which describe the content, whereas in Blogger they are referred to with the term "tunnisteet".

Research on indexing of social media is useful for all the aspects of the production, management, and seeking of digital information resources. The study clarifies the nature of subjective web communication by exploring its documentation, and thus helps to specify the traditional documentation sources. The cross-disciplinary approach brings new perspectives to the research field and helps to develop the research analysis method by borrowing expressive and useful concepts and tools especially from communication studies. The project has also a novelty value in describing a rich variety of culture and language bound problems in the development and study of multilingual and multicultural blogosphere. The results of the study are useful as well in private as in public sector for actors and developers involved in global Internet communication and social media.

References

- Aitchison, Jean 1991 (1984): *Language Change: Progress or Decay?* 2nd ed. Suffolk: Fontana Paperbacks. 258 p.
- Bawden & Robinson 2008: "The dark side of information: overload, anxiety and other paradoxes and pathologies." *Journal of Information Science*, Vol. 35, No. 2, 180-191 (2009)
- Blood, Rebecca 2003: "Weblogs and Journalism: Do They Connect." *Nieman Reports* 3/2003, s. 61-62. [Electronic resource, read 6.10.2009, URL: <<http://www.nieman.harvard.edu/assets/pdf/Nieman%20Reports/backissues/03fall.pdf>>]
- Bühler, Karl 1965 (1934): *Die Sprachtheorie*. 2nd ed. Stuttgart: Fischer. 434 p.
- Fairclough, Norman 1992: *Discourse and Social Change*. Cambridge, England: Polity Press. 259 p.
- Gillmor, Dan 2004: *We the Media. Grassroots Journalism by the People, for the People*. O'Reilly Media. Sebastopol. [Electronic version, read 5.10.2009, URL: <<http://oreilly.com/catalog/wemedia/book/index.csp>>]
- Hedemark, Å., Hedman, J. & Sundin, O. 2005: "Speaking of users: on user discourses in the field of public libraries". *Information Research*, 10(2) paper 218, 19 pp. Retrieved in October 31st 2005 from the World Wide Web URL: <<http://InformationR.net/ir/10-2/paper218.html>>
- ISO 5963-1985 = International Organization for Standardization 1985: *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms*. Geneva: International Organization for Standardization.
- Jakobson, Roman 1981 (1958): "Linguistics and Poetics". *Selected Writings, III: Poetry of Grammar and Grammar of Poetry*. Ed. Stephen Rudy. The Hague: Mouton. p. 18-51.
- Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero 1993: *Diskurssianalyysin aakkoset*. Tampere: Vastapaino. 238 p.
- Katan, David 2004 (1999): *Translating Cultures : An introduction for Translators, Interpreters and Mediators*. 2nd ed. Manchester, UK: St. Jemore Publishing. 380 p.
- Kjellberg, S. (2010). *Forskarbloggar: Vetenskaplig kommunikation och kunskapsproduktion i bloggösfären*. Lunds universitet. 226 p.
- Kovala, Urpo 2005: "Kommunikaation anatomia". IN: *Tarkkoja siirtoja*. Urpo Kovala, Katarina Eskola, Kimmo Jokinen, Vesa Niinikangas ja Esa Sironen (eds.). Jyväskylä, Jyväskylän yliopisto, Nykykulttuurin tutkimuskeskus. [Electronic resource, read 1.9.2009 URL: <<http://www.arthis.jyu.fi/julkaisut/tarkkojasiirtoja/kovala.html>>]
- Lehtonen, Mikko 2000 (1996): *Merkitysten maailma. Kulttuurisen tekstintutkimuksen lähtökohtia*. 3rd ed. Tampere: Vastapaino. 255 p.

Lindgren, Tim 2005: "Blogging Places : Locating Pedagogy in the Whereness of Weblogs." KAIROS 10:1. [Electronic resource, read 6.10. 2009 URL: <<http://kairos.technorhetic.net/10.1/binder2.html?coverweb/lindgren/index.htm>>]

Niemi-Pynttari, Risto 2007: *Verkkoproosa: tutkimus dialogisesta kirjoittamisesta*. Ntamo, Helsinki. 497 p.

Nykyri, Susanna 2010. *Equivalence and Translation Strategies in Multilingual Thesaurus Construction*. Doctoral thesis. Åbo, Åbo Akademi University Press. 413 p.

Ojala, Marydee 2005: "Blogging : For knowledge sharing, management and dissemination". *Business Information Review* 22(4) 2005, p. 269-276.

Olsson, M.R. 2004: "Understanding Users: Context Communication and Construction". Challenging Ideas: ALIA 2004 Biennial Conference, Gold Coast, Australia, September 2004 Teoksessa: *Challenging Ideas*, ALIA, online, s. 1-9. [Electronic resource, read 8.12.2004, URL: <<http://conferences.alia.org.au/alia2004/pdfs/olsson.m.paper.pdf>>]

Potter, Jonathan & Wetherell, Margaret 1990: "Discourse: noun, verb or social practice?" *Philosophical Psychology*, Vol. 3/1990, p. 205-217.

Pälli, Pekka 2003: *Ihmisyhmä diskurssissa ja diskurssina*. [Human group in discourse and as discourse] Acta Electronica Universitatis Tamperensis; 231. Tampere University Press, Tampere. 247 p.

Reiss, Katarina 1977: "Die literarische Übersetzung als Kommunikationsleistung." IN: Karl-Heinz Bender, Klaus Berger & Mario Wandruszka (Hrsg.) *Imago linguae. Beiträge zu Sprache, Deutung und Übersetzen*. München: Wilhelm Fink Verlag, p. 487-501.

Seale Clive (ed.) 1998: *Researching Society and Culture*. London UK: SAGE Publications. 349 p.

Slater Dan 1998: "Analysing cultural objects: content analysis and semiotics." IN: Seale Clive (toim.) 1998: *Researching Society and Culture*, s. 233-260. SAGE Publications, London UK. 349 p.

Strauss, Anselm & Corbin, Juliet 1990: *Basics of Qualitative Research. Grounded Theory Procedures and Techniques*. SAGE Publications, USA.

Suojanen, Päivikki 1993: "Identiteetti, perinne, vallan viestit." Teoksessa: Suojanen, Päivikki & Suojanen, Matti K. (toim.): *KULTTUURIN KALEIDOSKOPISTA : Kirjoituksia kielestä ja kulttuurista*. Tampere: Antrokirjat. 174 p.

Suoninen, Eero 1992: *Perheen Kuvakulmat: Diskurssianalyysi perheenäidin puheesta*. Tampereen yliopisto, Sosiologian ja sosiaalipsykologian laitos. Tutkimuksia, sarja A, nro 24. 137 p.

Talja, Sanna 1997: "Constituting 'information' and 'user' as research objects: a theory of knowledge formations as an alternative to the information-man theory", IN: Vakkari, P.,

Savolainen, R., Dervin, B. (ed.), *Information Seeking in Context*, p. 81-96. LONDON: Taylor Graham, London.

----- 1998: *Musiikki, kulttuuri, kirjasto : diskurssien analyysi*. Tampere: TAJU. 289 p.

----- 1999: "Analyzing qualitative interview data: The discourse analytic method." *Library & Information Science Research* 21 (4): 459-477.

Varantola, Krista 1990: *Tekniikan suomi yhdentyvässä Euroopassa : Sanastotyön merkitystä koskeva selvitys*. Helsinki: Tekniikan Sanastokeskus ry (TSK).

Vehmas-Lehto, Inkeri 1999: *Kopiointia vai kommunikointia : Johdatus käännösteoriaan*. Helsinki: Oy Finn Lectura Ab. 143 p.

Widén-Wulff, G., Huvila, I. & Holmberg, K. 2008: "Library 2.0 as a new participatory context." IN: M. Pagani (ed.), *Encyclopedia of Multimedia Technology and Networking*, vol. II. 2. ed. Hershey: Information Science Reference. p. 842-848.

Wierzbicka, Anna 1997: *Understanding Cultures Through Their Key Words. English, Russian, Polish, German, and Japanese*. New York: Oxford University Press. 317 p.