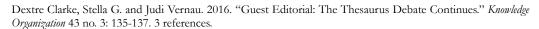# Guest Editorial:
The Thesaurus Debate Continues

Stella G. Dextre Clarke* and Judi Vernau**

*Luke House, West Hendred, Wantage, UK, OX12 8RR, <stella@lukehouse.org>
**Metataxis Ltd, 8 Thurleigh Avenue, London, UK, SW12 8AW, <judi.vernau@metataxis.com>

Stella Dextre Clarke has until recently been an independent consultant specializing in the design of thesauri and other types of knowledge organization systems. She is probably best known for her work on the national and international standards BS 8723 (Structured Vocabularies for Information Retrieval) and ISO 25964 (Thesauri and Interoperability with other Vocabularies). Her work on standards and on taxonomy development was recognized in 2006 when she won the Tony Kent Strix Award for outstanding achievement in information retrieval. Nowadays she is active as Vice-chair of ISKO-UK and Vice-President of ISKO.

Judi Vernau is a Director of Metataxis Limited, a consultancy specialising in information architecture and information management. She is a qualified librarian who made a move into reference publishing and the electronic structuring and categorisation of content in the 1980s. More recently she has worked on the development of information architectures to support enterprise information management, including the use of ontologies to support knowledge sharing and automatic categorisation. She has taught information architecture at London City University and Victoria University Wellington, and is currently Chair of ISKO-UK.

The urge to categorize is so intrinsic to the human brain that classification schemes, thesauri and other knowledge organization systems (KOSs) exert a fascinating magnetism for the community of members of the International Society for Knowledge Organization (ISKO). Despite, or indeed because of, sharing this fascination, in February 2015 ISKO-United Kingdom (ISKO-UK) felt the time had come to debate the proposition "This house believes that the traditional thesaurus has no place in modern information retrieval." As reported on the ISKO-UK web page for the event at http://www.iskouk.org/content/great-debate, the motion was resoundingly defeated. The debate was light-hearted, but the proposition remains a serious one, with much to be said on either side of the argument. Hence this special issue of *Knowledge Organization*, aiming to open the floor to the whole knowledge organization (KO) community, encourage continuing debate, and stimulate productive research and development (R&D).

In truth the proposition is not new. The debate began long before ISKO was founded, and is almost as old as the information retrieval thesaurus itself. So that all ISKO members could join our debate, a call for papers went out last June, inviting contributions on any relevant aspect: the result is this special issue.

The first short paper, from Stella Dextre Clarke, briefly traces the history of the debate as background to our continuing discussion. Another background paper, from Martin White, sketches a key area in the landscape of "modern information retrieval (IR)" in which the thesaurus might or might not find a place. While one region of this landscape—the World Wide Web with Google as kingpin—presents a scene uniformly familiar to most of us, the same is not true of retrieval practices inside public and private sector organizations. What is it like searching for information in corporate intranets, document and records management systems, knowledge management systems, customer relationship management systems, image libraries, museum catalogues, sound archives, news databases, multilingual digital asset management systems, etc.? Finding information is critical to the survival of many public and private sector enterprises, and to the national economies that rely on them. These are the target markets for vendors of enterprise search software, but do their products work effectively? White's paper points to widespread dissatisfaction with the technologies available, and to the correspondingly large opportunity for people with knowledge of KO techniques to put KOSs to work alongside them.

136

Knowl. Org. 43(2016)No.3
S. G. Dextre Clarke and J. Vernau. Guest Editorial: The Thesaurus Debate Continues

Only one ISKO member was brave and bold enough to speak out in favour of the debate proposition—Birger Hjørland, who suggests that while thesauri no longer have a place, there will always be a need for "more flexible semantic tools based on proper studies of domains." This seems a promising area for KO research with opportunities to yield practically useful results. We very much hope there will be follow-up in both the academic and the practitioner communities.

Andreas Kempf and Joachim Neubert counter Hjørland's arguments with their case study showing how at least one multilingual thesaurus plays a key role in supporting retrieval in the field of economics. By their account, the benefits of this thesaurus, the STW Thesaurus for Economics, are not limited to direct users of the four libraries and databases for which it was originally developed. The growth of the World Wide Web and other networks has allowed the thesaurus to reach an unknowably large global audience, supporting access to economics literature that may originate from a wealth of other repositories. To achieve this success, the developers have been exploiting advances in technology and widespread adoption of the publishing method known as linked open data. Standards such as the W3C recommendation Simple Knowledge Organization Sytems (SKOS) (Miles and Bechhofer 2009), as well as ISO 25964 Parts 1 and 2 (International Organization for Standardization 2011 and 2013), have helped them by providing guidelines that encourage a community of developers to build tools, applications and resources that function interoperably across diverse platforms and domains.

The remainder of the papers all favour a continuing role for the thesaurus, sometimes complemented by other types of KOS. Given the context of a debate, some are fairly short and not all follow the usual style of scholarly papers. For example, some of them select references to support only one side of the argument. If this partisan stance provokes some readers to respond with research that demonstrates they are wrong, or even with a letter of protest to the editor, so much the better!

The paper from Douglas Tudhope and Ceri Binding reinforces the STW case study by describing a wave of similar developments. They claim that efforts to build and apply an ontology have sometimes been abandoned or modified in favour of the greater flexibility afforded by a thesaurus. It seems the rigour and expense of the ontological approach is easier to sustain on a small scale, or in a small domain, than in circumstances that need a large number of highly qualified developers and users. Their conclusion points to areas where more R&D would support progress.

Moving forward from the established developments described by Tudhope and Binding, Javier García pro-

vides a more speculative paper, a leap of faith in which the thesaurus publishing community exploits Wikipedia and its derived products to build an interconnected web of thesauri and other KOSs. Predicting the future is always a dangerous game, and wishful thinking can lead to big disappointments. But a pilot project to test this proposal could well provide insights and inspiration for new opportunities in which the social media take advantage of established thesaural structures.

In the last four decades of the 20th century, much of the thesaurus debate dwelt on comparisons of full text with thesaurus-driven index terms. It neglected the contexts where there is no "full text"—notably collections of photographs, artworks, videos or audio recordings. These types of material are nowadays in great demand, and indexing is indisputably still essential. Andrew MacFarlane points to the failure of artificial intelligence and other computer-based IR methods in this area—a clear example of the so-called "semantic gap." Is the thesaurus the best option for subject indexing in this important niche, or should it be replaced by some other type of KOS?

Returning to White's paper, he points to the value of "taxonomies and metadata" without specifying which of these KO tools is or are most useful, and how it or they should be built and implemented. Vendors as well as typical practising information managers and users employ the term "taxonomy" to mean anything from an ontology to a simple list of folder headings. Between these extremes lie all manner of thesauri, subject heading lists, synonym rings, filing plans, etc., with no very clear principles for which type of KOS functions best in what circumstances. Advances in technology and changes in societal expectations often tip the balance of advantage from one type of tool to another. And as Hjørland points out, the nature of the domain and the quality of construction have a big influence on the performance of any thesaurus. A small and little-recognized band of consultants is available to help managers design, choose or build and then implement the appropriate KOS (if any), usually on the basis of gut feel, artistry and experience as much as established principles. If only the KO research community could provide clear guidelines for this work!

We in the KO community are the best placed to study the theory and the practice of KOS applications. Conclusions from the papers in this issue of the journal point to only some of the many opportunities for R&D that could make a difference to people all over the world, as they struggle to find the information needed in today's complex society. Can we deploy our fascination with KO techniques to effect, leading to thesauri that earn their keep, or to alternative types of KOS that perform better? The editor of this journal awaits news of where we and you go next.

References

International Organization for Standardization. 2011. ISO 25964-1: *Information and Documentation—Thesauri and Interoperability with other Vocabularies - Part 1: Thesauri for Information Retrieval*. Geneva: International Organization for Standardization.

International Organization for Standardization. 2013. ISO 25964-2: *Information and Documentation—Thesauri and Interoperability with other Vocabularies - Part 2: Interoperability with other Vocabularies*. Geneva: International Organization for Standardization.

Miles, Alistair and Sean Bechhofer, eds. 2009. *SKOS Simple Knowledge Organization System Reference. W3C Recommendation*. http://www.w3.org/TR/skos-reference.