

# Multilingual bookbinding terms: complex conceptual issues

Velios, Athanasios; Pickwood, Nicholas; Martin, Aurelie

Ligatus Research Centre,

University of the Arts London,

a.velios@arts.ac.uk

## 1 Introduction

The value of the study of bookbinding history has not been widely recognised for a number of reasons. One reason is the lack of research data. Bookbinding history has, so far, focussed on heavily decorated bindings (or treasure bindings). These are certainly interesting objects but they represent a tiny fraction of the books produced as they were only made for rich buyers. There are no significant resources published about ordinary bindings and their production. To produce such a representative resource collaboration among many libraries is required and we consider the Linked Open Data initiative as a good framework to work towards that target. Libraries which are willing to share their records can do so following a semantic framework with the CIDOC-CRM being an obvious candidate. A first requirement for a implementing a methodology for sharing bookbinding records on the semantic web is the development of a thesaurus of bookbinding terms. The Ligatus research centre at the University of the Arts London (UAL) secured a networking grant from the Arts and Humanities Research Council (AHRC) in the UK to develop a thesaurus of bookbinding terms in collaboration with the Foundation of Research and Technology Hellas (FORTH).

## 2 Language of bindings thesaurus

### 2.1 Cultural representation

The challenges of the development of multilingual thesauri have been investigated previously. Hudon (1997 and 2010) and De Santis et al. [2012] have described the problem of cultural misrepresentation in multilingual thesauri where

typically the bulk of the development is done in one language (often English) and other languages are forced to fit in a cultural structure which is not always ideal. We address this issue by involving experts from a wide range of cultural backgrounds within Europe. Every major book production centre is represented in the thesaurus development team which consists of 35 experts. The working language for the thesaurus is indeed English but the development team does not bring any cultural bias about the thesaurus structure.

## 2.2 Remote versus local development

Limited funding in the field of historic bookbinding would normally force the development team to work remotely and on a voluntary basis over a long period of time. Since the work for this thesaurus is funded externally, it was possible for the large development team to jointly work on the thesaurus in one location during the so-called *thesaurus development sprints*. *Development sprint* is a term used in the free software communities where a group of software developers meet in one place for a short period of time to produce a piece of functional software. Borrowing that idea we organised three thesaurus development sprints which led to a significant part of the thesaurus being produced. Experts were divided into teams and were given a subject area to develop. A facilitator was responsible for documenting the discussions about different terms. The development was done in short sessions after which representatives of each team would meet to check for duplications or to ask for advice. The resulting terms from these sprints are not necessarily finished thesaurus entries. There is still a significant amount of academic research to be done to check the validity of each term, but this exercise meant that a cross-cultural body of work for the thesaurus was done in a short period of time.

## 2.3 Semantic web and other initiatives

The use of the thesaurus within the Linked Open Data idea is a core objective of the project. Having produced a number of terms during the development sprints we chose SKOS as a popular system for organising our bookbinding concepts. The terms were inserted and processed in a thesaurus management system (WebTMS by FORTH) which follows the SKOS standard. The terms are being published online with the intention that they will be used as URIs within the context of the semantic web. The development team recognises the pivotal role of the Getty Arts and Architecture Thesaurus (AAT) in cultural heritage and considers the proposed bookbinding terms in relation to the AAT records. Where possible these records are matched. Alternatively the development team will engage with the AAT editors to recommend and justify alterations in the AAT records.

## 3 Binding structures for multilingual thesauri

### 3.1 CIDOC-CRM and top terms

As mentioned in the previous section, the thesaurus being produced is to be used alongside other semantic technologies. With the CIDOC-CRM being a strong candidate for mapping the bookbinding history domain the development team consulted with CIDOC-CRM experts to decide on the main thesaurus top terms. There are:

- objects (similar to CRM E19 Physical Object), which include the separate components of a binding,
- features (similar to CRM E26 Physical feature),
- materials (similar to CRM E57 Material),
- techniques (similar to CRM E29 Design or Procedure).

Other top terms are added as necessary (such as terms to cover E39 Actors or E73 Information Object) but they accommodate far fewer terms.

### 3.2 Equivalence

While the problem of having too many terms with the same meaning can be easily solved by choosing a preferred term, there are other relationships of terms from different languages which are worth investigating. A good document outlining specific challenges with terms in multilingual thesauri is the Guidelines for multilingual thesauri by the International Federation of Library Associations and Institutions [2009]. An important consideration when building multilingual thesauri is the problem of *equivalence* of terms from different languages. Equivalence identifies the type of relationship among translated terms or similar concepts in the different languages.

#### 3.2.1 Exact equivalence

Ideally, every concept of the thesaurus should have a term in every language. The source terms of one language should convey the same meaning when translated to the target terms of another. In reality achieving this result is difficult because of the cultural and linguistic particularities of each group in a given domain. However, in the case of bookbinding it is important to emphasise that the binding structure of the codex is by and large the same across many cultures and periods. There is great variation and rich typologies of the different structural components but the overall structure remains remarkably similar. Therefore, given that a bookbinding thesaurus primarily deals with the description of binding components, perhaps it is valid to say that in this case achieving exact equivalence of at least the broader terms is more likely than in other domains. Indeed in our bookbinding thesaurus there are examples of terms with exactly equivalent terms in other languages.

### 3.2.2 Partial equivalence

It is often the case that the meaning of a term in one language is slightly broader than that in another. For example, a *brochure* binding in French is generally understood as one with the so-called *drawn-on covers*. No assumption in terms of the thickness of the book is made. A frequent translation of this term in English is *pamphlet*. *Pamphlets* could have *drawn-on covers* but this is not a requirement and also it is a term used for small ephemeral publications only, so the size of the book is important. When a case of partial equivalence is observed then the respective terms are arranged hierarchically and treated as either compound or preferably as orphan terms.

### 3.2.3 Compound terms

In partial equivalence, an interesting exercise is to try to describe the source term with a compound term in the target language and although in some cases this is feasible by using valid compound terms, the exercise may result in the production of compound terms which are not really used by domain experts. For example the term *αγοπρόβατα* in Greek could be translated as *sheep and/or goats* in English. In Greek it can be used to indicate the provenance of skins as either *sheepskin* or *goatskin*. However there is no term *sheep and/or goats* in English and we therefore prefer to consider these source terms as orphan terms as described next.

## 3.3 Orphan terms

The circulation and sharing of terms across different languages is a frequent phenomenon in the bookbinding world. It is not unusual for foreign terms to be used to describe specific types of bindings. Therefore, we encourage the development of a single structure of broader and narrower terms which is then populated with terms from different languages. We employ exact equivalents when available or consider orphan terms in the rest of the languages. A good example is the German term *kapitalbund* which is a sewing component only evident in the German tradition without an equivalent in any other language.

## 4 Thesauri in surveys

In this paper we have focussed on a thesaurus of bookbinding terms but it is important to emphasise that a bookbinding thesaurus will be typically used during a survey of bindings in a library. The survey itself would require more information on top of the binding record, such as the name of the surveyor or the date of the survey. It will also require data confirming the completeness or accuracy of each record. It is often the case that the absence of a reference to a bookbinding component in a record is ambiguous. Is the reference missing because the component is not there (therefore we are talking about a type of binding)? Or is it missing because the surveyor forgot to observe it? Or more

likely, is it missing because the surveyor could not observe it (which is often the case with internal components of books in good condition)? And what happens when the surveyor has reasonable evidence to assume a component exists, but with an element of doubt? Should the thesaurus reflect these levels of uncertainty? Or should this be a concern of the survey methodology?

So far, in the bookbinding thesaurus we have tried to focus on descriptive terms for observable components and techniques. However, it is inevitable that our familiarity with surveys as a data collection exercise and the aforementioned questions affect the way our terms are chosen. For example the term *stitched without a cover* indicates the significance of the absence of a cover from a book. In this case the cover is singled out as an absent component among a number of other components and it is confirmed that its absence is part of the typology of the binding and not the result of a poorly executed survey. In our opinion further discussion for best practices on recording the accuracy and validity of records is required in relation to the production of thesauri and we intend to investigate this further within the context of the bookbinding thesaurus.

## Acknowledgements

The authors thank Martin Doerr and Maria Theodoridou for their continuous support and guidance in this project as well as all the thesaurus contributors for their invaluable work in this project.

## References

Barbara De Santis, Melissa Tiberi, and Fulvio Mazzocchi. Language and culture representativeness in multilingual thesauri development. *Magyar Terminológia*, 5(2):211–224, December 2012. ISSN 1789-9486, 2060-2774. doi: 10.1556/MaTerm.5.2012.2.1. URL <http://www.akademiai.com/content/m678930680m23314/>.

Michele Hudon. Multilingual thesaurus construction-integrating the views of different cultures in one gateway to knowledge and concepts. *Information Services & Use*, 17(2/3):111, July 1997. ISSN 01675265. URL <http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9712125918&site=ehost-live>.

Michèle Hudon. The world through words. *The Indexer*, 28(4):146–150, December 2010.

International Federation of Library Associations and Institutions and Working Group on Guidelines for Multilingual Thesauri. *Guidelines for multilingual thesauri*. International Federation of Library Associations and Institutions, The Hague, 2009. ISBN 9789077897355 9077897356.