ISO-Standards for Lexicography and Dictionary Publishing

Marie-Jeanne Derouin
Langenscheidt Fachverlag

André Le Meur
Université du Québec en Outaouais

Many things have changed in the field of dictionary production during these last ten years. With the introduction of digital support and networking, the lifespan of dictionaries has been considerably extended. The dictionary manuscript has become a unique data source that can be re-used and manipulated time and again by numerous in-house and external experts. The traditional relationship between author, publisher and user has now been expanded to include other partners, such as data-providers-either publishers or institutions or industry-partners, software developers, and language-tool providers. All these dictionary experts need a basic common language to optimize their work flow and to be able to co-operate in developing new products while avoiding time-consuming and expensive data manipulations. Dictionary users also need to receive more reliable information about new lexicographic products. In this paper we will first of all present the ISO standardization for Lexicography which takes these new market needs into account, and then go on to describe the two new standards: Presentation/Representation of entries in dictionaries which was published in March 2007 and Lexicographical production and marketing: Concepts and vocabulary which was launched in the summer of 2007. In conclusion, we will outline the benefits of standardization for the dictionary publishing industry.

1. General points on ISO-Standardization

Standardization is achieved on national, regional and international levels. Most countries in the world have their own national standardization body which is organized into technical committees according to the application fields and develop their own national standards.

Most countries also contribute to the development of international standards which are accepted worldwide. Stakeholders from industry, academia and societies from all over the world are involved in creating international standards within the ISO (International Standardization Organisation).

International standards aim to prevent national variations and give priority to describing performance requirement and interfaces. Products are not standardized.

Standardization is achieved according the following principles:

Market adequacy: a project will only be launched if a need has been identified through a market survey in all member countries.

Voluntary implementation: experts work on standards on a voluntary and anonymous basis.

Consensus based: at each of the main stages of development, experts from all member countries have to give their comments and to vote for or against the circulated draft.

One vote per country.

Worldwide acceptance.

2. The current situation in dictionary standardization

Only one ISO-Standard for dictionaries has been published: the ISO 1951: *Lexicographical symbols and typographical conventions for use in terminography*. This appeared in 1973

and was revised in 1997. It mainly concerned the layout of print dictionaries and did not address the actual needs of dictionary making.

Now standards have to take into account the broadest variety of dictionary products: monolingual, bilingual, multilingual, general language and specialist dictionaries, in print form or electronic, as isolated products or integrated into other language tools.

With the introduction of digital support and networking, the lifespan of dictionaries has been considerably extended and the manuscript has become a unique data-source that can be reused and manipulated many times by editors and content management experts. Dictionary publishers are importing and accommodating more and more dictionary data from other sources such as industry and other institutions or from other publishers when acquiring perhaps a new dictionary series or co-publishing a new dictionary. In many cases they have to deal with international partners. Dictionary publishers are also content-providers for other language tools such as translation software or translation memory systems or other internet products.

In order to optimize their work flow and to avoid time-consuming and expensive data manipulations, all these different players need to be implementing the same current best practices in terms of data representation.

The international growth in dictionary production also makes it necessary to create lexicographical products which can be universally understood and therefore to develop a transparent vocabulary for users, lexicographers and publishers alike.

Taking all these new challenges into account, the ISO technical committee 37 for terminology and lexicography has just published the entirely revised ISO-Standard 1951 and launched a new project which we will be describing later in this paper. An information campaign has also been launched in order to increase the awareness of standards amongst dictionary publishers.

3. The revised ISO-Standard 1951: Presentation/representation of entries in dictionaries

3.1. *This standard aims:*

to support the creation and management of various types of dictionary

to allow dictionary content to be reused in different printed and electronic formats

to facilitate necessary production, exchange and management procedures

to propose a specific model based on current best professional practices

It has been approved by experts from over 20 countries and it applies to every kind of dictionary.

3.2. *The normative section indicates requirements to be met as stated and includes:*

Definitions concerning basic and unambiguous terms relating to dictionary structure and presentation, which are common in most types of dictionaries, and which are within the scope of this International Standard.

A formal description of dictionary entries with an overview of data elements and compositional elements. It describes data elements using numerous examples from over 40 existing dictionaries and their grouping into compositional elements needed for representing the most common dictionary entries. For each data element a conventional name and description is provided, based as far as possible on [ISO 12620:1999]. Free data elements, definable by the user, allow extensions of the model for "negotiated interchange".

A formal generic dictionary model XmLex which offers a balance between strict formal structures which permit automation and user friendliness whilst still conforming to traditional lexicographic methods. This model also satisfies four requirements which enable data to be independent from tools and media (paper, CD-ROM, internet). They are:

Complete separation between logical structure and display.

Non ambiguity: all the relations between data elements can be computed and can be interfaced with any lexical database.

Flexibility: the generic model allows subsets to be specified according to specific needs.

Compatibility with currently available XML-tools

More information about the xml model (known as XmLex) defined in ISO 1951, examples of encoding and xsl libraries for displaying or transforming XmLex data are to be found at *http://www.xmlex.net/lexicography/xmlexintro.pdf.*

The most important application shown in these demonstrations is the automatic transformation of lemma-oriented entries into monosemic units that can be imported into most of the terminology management systems of translation memories. The principle of this transformation has been presented at EURALEX 2006 and the first industrial product based on this technology has been put on the market in Germany at TEKOM 2007 by across Systems GmbH integrating Langenscheidt's specialised dictionaries in the across terminology system crossTerm.
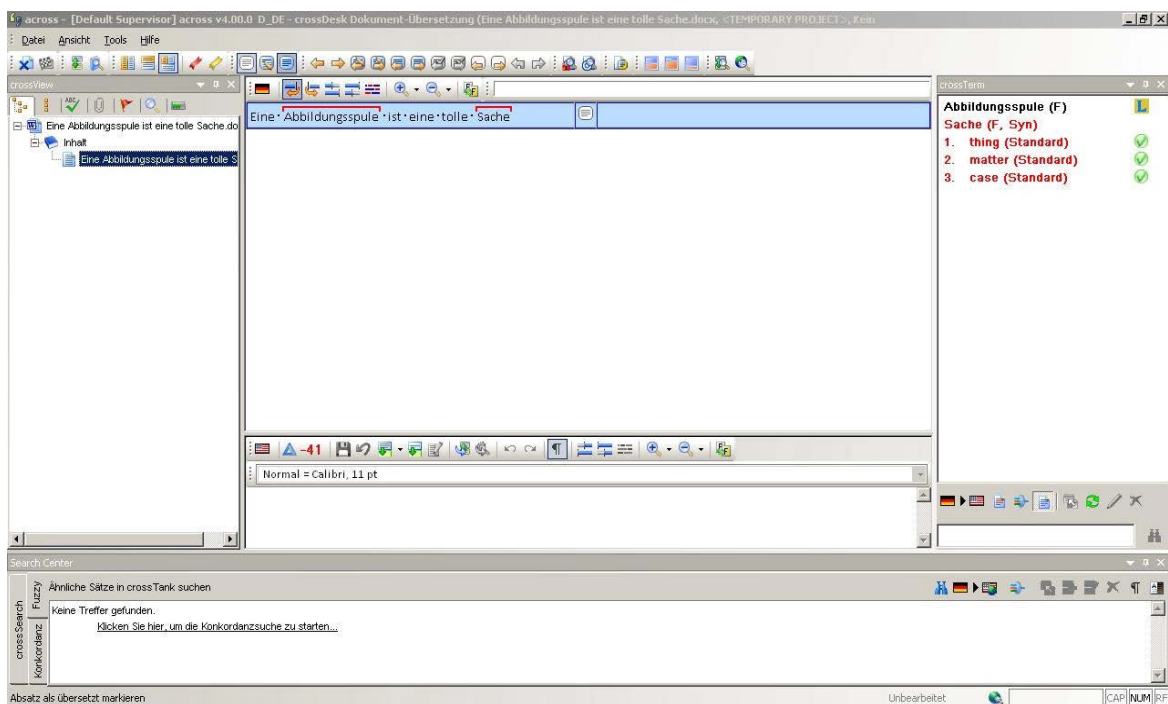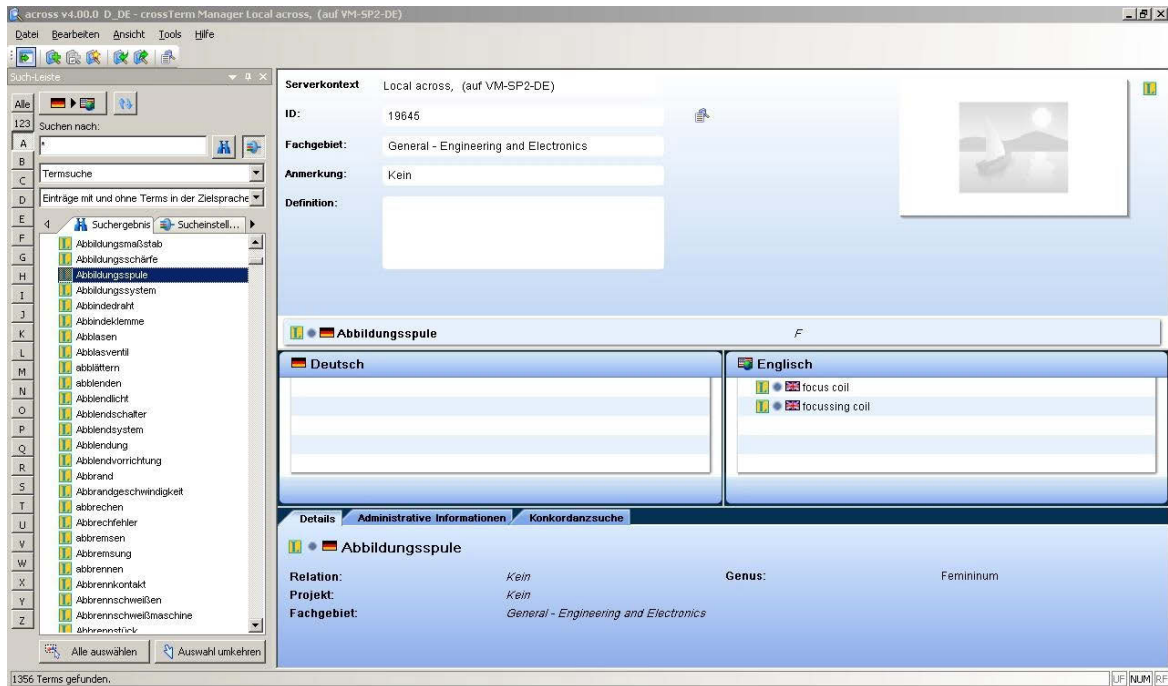
Table 1: Screen capture across 1

Table 2: Screen capture across 2

3.3. *The informative section provides useful recommendations and presents:*

Means of presentation of entries in print and electronic dictionaries with layout aids and compacting mechanisms. Depending on the medium, one or more layout aids are used. Text formatting and typographical conventions can be used for all media to differentiate entry elements. The tabular form is especially recommended for dictionaries with more than two languages. For electronic media, data bank templates can be used as a specific form of presentation.

Numbering systems

Table of function of lexicographical symbols

Examples of XML-Encoding

> *Dictionary entries*
>
> **dam1** *[dæm]* **1** n **a** *(wall)* [river] *barrage* m *(de retenue), digue* f *;*[lake] *barrage (de retenue).* **b** (water) *réservoir* m*, lac* m *de retenue.***2** vt **a** *(also ~ up)* river *endiguer ;* lake *construire un barrage sur.* **to ~the waters of yhe Nile** *faire or construire un barrage pour contenir les eaux du Nil.* **b** flow of words, oaths *endiguer.* **3 comp** *dambuster (*bomb*) bombe* f *à ricochets ;* (person) *(aviateur* m*) briseur* m *de barrages* (se réfère à un épisode de la seconde guerre mondiale).
> **dam2** *[dæm]* n *(animal) mère* f.
> **dam3** *[dæm]* adj, adv **a** *= **damn 4,5**.* **b** *(US) ~ **Yankee** sale* Yankee or nordiste.*

Xml encoding of the first entry

```
<DictionaryEntry identifier="damFR-0-1" sourceLanguage="en" targetLanguage="fr">
  <HeadwordCtn><Headword>dam</Headword>
  <Pronunciation>daem</Pronunciation> </HeadwordCtn>
  <HomographGrp>
    <PartOfSpeech value="noun"/>
    <SenseGrp>
      <SenseIndicator>wall</SenseIndicator>
      <TranslationBlock>
        <TypicalComplement>river</TypicalComplement>
        <Translation>barrage <GrammaticalGender value="masculine"/>
          <Optional>de retenue</Optional></Translation>
        <Translation>digue</Translation> <GrammaticalGender value="feminine"/>
```

666

```
    </TranslationBlock>
    <TranslationCtn>
      <TypicalComplement>lake</TypicalComplement>
      <Translation>barrage <GrammaticalGender value="masculine"/>
        <Optional>de retenue</Optional></Translation>
    </TranslationCtn>
  </SenseGrp>
  <SenseGrp>
    <TranslationBlock>
      <SenseIndicator>water</SenseIndicator>
      <Translation>réservoir<GrammaticalGender value="masculine"/></Translation>
      <Translation>lac<GrammaticalGender value="masculine"/> de retenue</Translation>
    </TranslationBlock>
  </SenseGrp>
  </HomographGrp>
</DictionaryEntry>
</Dictionary>
```
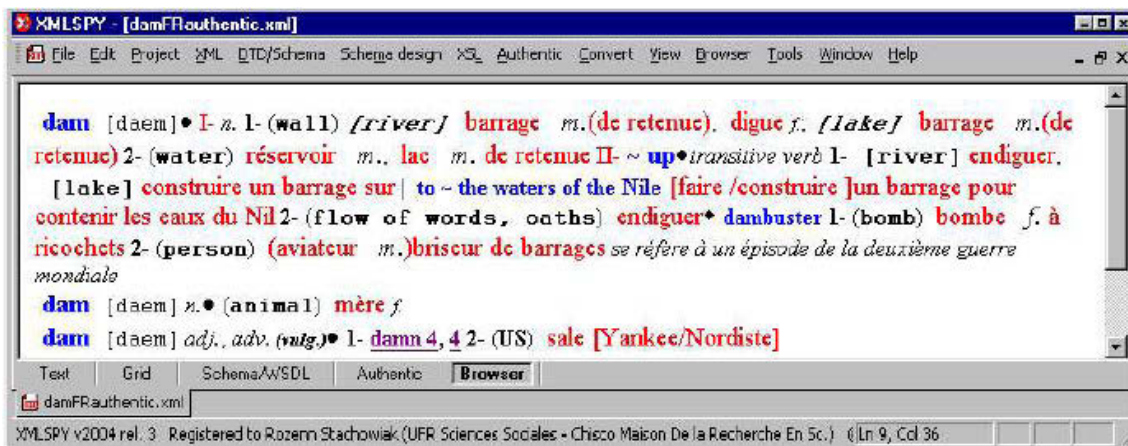
HTML display

Table 3: Example of XmLex (excerpts from XmLexIntro.pdf)

3.4. *ISO 1951 is a useful guide:*

for the development of new dictionaries in all publishing media

for the production of new dictionary products based on existing data collections (extraction and merging of dictionaries)

for the integration of dictionaries from other series in an existing dictionary programme

for providing dictionary data in other language tools (Translation Software, Translation Memory tools)

for optimizing the editorial work flow

It is available from every national standardisation body such as DIN in Germany or AFNOR in France.

4. The new standardization project: Lexicographical production, marketing concepts and vocabulary

4.1. *General information*

The new ISO project was launched at the last meeting of the ISO technical committee for terminology (ISO TC37) in August 2007. It addresses every type of dictionary and aims:

to ensure greater transparency of lexicographical products in the production and marketing of dictionaries for the benefit of both users and dictionary publishers

to allow genuine comparisons between products

to facilitate transactions between publishers

to improve the dictionary publisher's image among dictionary users

4.2. *Current proposals for the content:*

Description of the products (all types of dictionary, different components of print and electronic dictionaries etc.)

Process of dictionary making (setting up a product project, compilation of data, data and corpora management, editorial work, product naming, publishing etc)

Marketing: product distribution, merchandising, copyright trading, royalties etc. (the first draft is now in progress. It will be circulated by the end of 2008).

5. Conclusion

The standardization of dictionary products is beneficial for dictionary publishers. New market demands require new dictionary products and dictionary users need global language solutions. The interoperability of data saves time and cuts costs. The importance of standardization is growing as dictionary prices have to be kept low because of the increasing development of free internet products and consequently of price consciousness among dictionary users. The dictionary market is becoming more global and it is becoming increasingly necessary for there to be a basic common language used throughout dictionary publishing. While standardization may not be as indispensable as in other branches of industry, conformity to an ISO-Standard will improve quality assurance and contribute to a better appreciation of dictionary products.

Standardization is not intended to regulate dictionary production, however. Dictionary publishers should not be worried about the regulation of choice, content or layout of dictionary products. All the players in the dictionary industry are invited to participate in the development of lexicographical standards by providing expert advice to their national standardization bodies. This will ensure that these standards continue to reflect market requirements and allow everyone to stay informed about new developments in this field.

References

Derouin, M. J.; Le Meur, A. (2006). "Integrated Bilingual Specialist Dictionaries as Added Value for Translation Memories: The LexTerm Initiative." In *Euralex Proceedings. Translating and the Computer 28 Conference*.