

Data categories for multilingual harmonisation-oriented terminology work

Elena Chiocchetti
European Academy of Bolzano (EURAC)
Institute for Specialised Communication and Multilingualism
Viale Druso 1
39100 Bolzano
echiocchetti@eurac.edu

The present abstract is based on the experience gathered within the LexALP¹ project, which aims at the detailed analysis and subsequent harmonisation of the specialised terminology used within and around the Alpine Convention, in particular as regards legal terms. The final objective of the project is to have a group of experts determine translation equivalents in the four official languages French, German, Italian and Slovene, in order to foster clear and efficient communication within the Convention. This can only happen after an in-depth comparison of the terms used by the Alpine Convention with other relevant national and supranational legal systems. Within the project, terminology work is carried out from different locations and stored directly online in a dedicated terminology data base (<http://www.eurac.edu/lexalp>).

The Alpine Convention is a framework agreement signed by all Alpine countries (Austria, France, Germany, Italy, Liechtenstein, Monaco, Switzerland and Slovenia) and the European Union for the protection and sustainable development of the Alps.² Given the involvement of several countries with different legal backgrounds and official languages, comparative and harmonisation-oriented terminology work proved a challenge from the very beginning. In particular, the initial definition of data categories to be used in the online term bank posed several problems, which were then underlined even more during practical terminology work.

Next to the types of data fields, the so-called data categories or data elements, commonly present in most term banks (term, definition, context, source etc.) a specific data category had to be foreseen in order to indicate the *harmonisation status* of the terms, as determined during the final decision process by the group of experts (closed data category: harmonised/rejected/unknown). In this phase it was also decided to add an open data category 'note', on the basis of the experience that full equivalence is rarely to be encountered when working on different languages and legal systems. Hence, comparative notes for easing the harmonisation process of the Alpine Convention terms would have been necessary.

In the ISO 12620 norm the data category 'note'³ is not underspecified. This data category is described as "supplemental information pertaining to any [...] element in the data collection, regardless whether it is a term, term-related, descriptive, or administrative". Two more specific data categories are suggested for use whenever possible, 'usage note' and 'transfer comment'. As regards the first, ("a note containing information on the usage of the associated term") its subcategory 'geographical usage' is used in the LexALP term bank to indicate the legal system to which the term belongs. The second category, 'transfer comment' ("note included in a term entry providing more explicit information on the degree of equivalence, directionality or other special features affecting equivalence between a term in one language and another term in the same or a second language") was originally not foreseen, because notes were meant to be written mainly for terminology entries concerning the Alpine Convention and with the goal of facilitating the harmonising process rather than the translation process. Furthermore; analysing the different degrees of equivalence between all legal systems involved singularly would have gone beyond the aim of the project.

¹ Legal Language Harmonisation System for Environment and Spatial Planning in the Multilingual Alps (<http://www.eurac.edu/lexalp>)

² http://www.alpenkonvention.org/page1_en.htm (05/12/2006)

³ Admitted names for this data category are also 'remark' and 'comment'.

During practical terminology work over four languages and nine legal systems⁴, it was realised that the use of this note field was very diverse, according to the specific situation/problem the terminologists encountered. At the moment there are at least five different types of information in the data category 'note':

- harmonisation notes (e.g. comments referred to two or more languages at Alpine Convention level, underlining the different implications of the terms used by the Alpine Convention when considered from the point of view of the other - mostly national - legal systems analysed; comments explaining differences in usage; comments proposing alternative or more correct terms, etc.) which are meant to ease the work of the harmonisers
- comparative notes not concerning harmonisation work (e.g. comments referred to two national legal systems)
- legal notes referred to one legal system only
- linguistic notes
- other 'personal' comments of the terminologist

Apart from the first category, these types of notes were not considered essential for the goals of the project. However, they undoubtedly constitute an added value and should be kept in the term bank. This leads to the problem of having a data category containing a rather wide range of information, which is not ideal, neither for data filtering nor for possible data exchange with other terminology data bases.

Even though the human user has no problem at all in understanding the different scope of the notes, the goal of this contribution is to show what kind of notes could and should be used for this type of harmonisation-oriented multilingual terminology work which focuses on legal terminology as well as to propose a clear categorisation on the basis of our experience, so as to be able to differentiate these types of notes in future.

Bibliography

- ISO 12620:1998. *Computer Applications in Terminology - Data Categories*. Geneva, International Organisation for Standardisation.
- Chiocchetti, E. & Lyding, V. (to be published): Multilingual Corpus for Terminology Work: The LexALP Corpus of Legal Texts. Poster presented at the Conference *Multilingualism across Europe: findings, needs, best practices*. Bolzano, 24th-26th August 2006.
- Lyding, V./Chiocchetti, E./Sérasset, G. & Brunet-Manquat, F. (2006): The LexALP Information System: Term Bank and Corpus for Multilingual Legal Terminology Consolidated. Paper presented at the *MLRI-Multilingual Language Resources and Interoperability* workshop. Sydney, 23rd July 2006. <http://acl.ldc.upenn.edu/W/W06/W06-1004.pdf>.
- Laurén, Ch./Myking, J & Picht, H. (1998): Terminologie Unter der Lupe. Vienna, TermNet, p. 274-301.
- Sérasset, G./Brunet-Manquat, F./Chiocchetti, E. (2006): Multilingual Legal Terminology on the Jibiki Platform: The LexALP Project. Paper presented at the *COLING/ACL2006*. Sydney, 17th - 21st July 2006. <http://acl.ldc.upenn.edu/P/P06/P06-1118.pdf>.
- Wright, S.E. (2001): Data Categories for Terminology Management in *Handbook of Terminology Management*. Amsterdam/Philadelphia, J. Benjamins, p. 552-571.

⁴ Namely: Austria, France, Germany, Italy, Switzerland and Slovenia at national level and European and International law at international level, the latter being the legal systems of reference of the Alpine Convention. Also Liechtenstein and Monaco are parties to the Alpine Convention; however, their legal systems are not terminologically processed within LexALP. The Alpine Convention level itself is treated as the ninth legal system in the term bank, even though, strictly speaking, it does not constitute a legal system per se but could be considered part of international law.