

## Harmonisation of Environmental Terminologies

Gerhard Budin

Department for Philosophy of Science, University of Vienna, Sensengasse 8, A-1090 Wien, Austria  
gerhard.budin@univie.ac.at

### Abstract

A critical review is given of the current status of →terminology and →thesaurus projects in the field of environment on the European and international levels. For this purpose the following dimensions may be distinguished:

- **technology:** telematic-based co-operative applications are being developed, the creation of →terminology databases accessible via the World Wide Web has gained prime importance. New software has been programmed, or existing packages have been adapted to new needs. This also includes translation technology and corpus analysis programmes;
- **methodology:** ongoing projects in the framework of the European Environment Agency show that (traditional) methods of →terminology, nomenclature, classification and →thesaurus development should be further developed and integrated into a comprehensive knowledge organisation methodology. This approach also has to address the problem of thematic (vertical) fragmentation and specialisation, lacking concomitant cross-disciplinary (horizontal) re-integration of →terminological knowledge;
- **organisational strategy:** the thematic fragmentation mentioned above is mirrored on the organisational level – institutions specialising in different aspects of the wide field of environment hardly co-operate and communicate with each other. Co-operative →terminology development requires much more cross-disciplinary discussion, e.g. in order to harmonise conflicting →definitions of the same →concept used in different themes.

The overall process of →terminology development and management requires a balanced and dynamic, interactive relationship among these dimensions. Any technology is useless without the appropriate methodology, the best methodology does not work without the necessary organisational support, the best strategy will remain theoretical without choosing the right methodology and the necessary technology.

Keywords: terminology, terminology management, co-operative development, environment, harmonisation, European Environment Agency, technological development, methodology

## 1 Introduction

In this paper I would like to present some information on the multi-annual sub-project on “Consistency of Terminology and Coding” and on the development of a harmonised European multilingual environmental →thesaurus, as part of the work programme of the European Environment Agency (EEA). The EEA is carrying out a vast variety of projects for gathering and disseminating information on the state of the environment in Europe.

The tasks covered in this project can be classified on four different levels:

- the strategic level, where a general strategy of developing and harmonising environmental →terminologies and nomenclatures (coding systems) is designed;
- the methodological level of developing a →terminology management methodology tailored towards the identified needs and requirements of various user groups with different roles in the overall framework;

- the technical level concerning the technical feasibility of implementing the strategies and methodologies mentioned above and setting up an experimental World Wide Web (Web) page for this endeavour;
- the information level of promoting the use of the harmonised →thesaurus and of the classification and coding systems.

The motivation for designing and implementing the strategies and methodologies mentioned above is a pragmatic one: without consistency in the →terminologies that are used in all discourse within the European framework of environmental information, no efficient understanding will be possible. Environmental experts working in diverse areas with different aims increasingly feel the need for harmonisation and standardisation of the →terminologies that they use, already in monolingual, but all the more in multilingual discourse. Co-ordinated →terminology development in the highly heterogeneous and rapidly developing field of environment and with all the closely related fields such as chemistry, biology, forestry, agriculture, medicine and health, etc. has become a major strategic task. It will enable experts from different specialities to efficiently communicate with each other, since their highly specific →terminologies develop into diverse directions. This inevitably leads to terminological problems, e.g. a →term that denotes different →concepts within the same field (polysemes and homonyms) or a →concept being represented by different →terms (→synonymy), fuzzy and vague →definitions of the →concepts used, incompatible →terminologies in various languages which leads to severe translation problems, etc.

The setting up and operation of a multilingual →terminology database is an efficient tool to eliminate such inconsistencies in environmental discourse. The consistent use of a common indexing and retrieval language is the pre-requisite for precise and reliable storing of environmental information and its retrieval for various purposes.

## 2 Strategies of User-oriented Information

On the strategic level it is important to identify the users (institutions, experts) with their different expectations, aims, needs and requirements, and the general framework (networks of co-operation, bi- and multilateral agreements and contracts, etc.).

As mentioned in the overview, the population of such a database available in the WWW needs both the integration of existing databases or relevant parts thereof and the creation of new information by active members in the network and by including links to existing relevant databases also available in the Web. This heterogeneous cluster of available information and the creation of new information need a careful implementation strategy tailored to the different cases distinguished here.

User groups may be distinguished according to many different criteria. One is the subject field (domain) they are working in. This criterion immediately implies the problem of classification of these subject areas. It is of limited use, since an important percentage of →terminology is not limited to one specific subject field, but used in several ones (but not always in the same way). In addition, environment is a multi- and cross-disciplinary area par excellence: an expert in environmental law or a politician responsible for environmental policy in a certain country or province may need specific information from biochemistry, biology, forestry, etc., and thus a specific knowledge of the relevant →terminologies involved.

This aspect of multi- and crossdisciplinarity is even more valid when it comes to a problem-oriented definition of tasks, including the institutionalisation of such tasks on the European level. This is in fact another criterion for distinguishing user groups: each institution with its staff and its specific duties and activities is a user group of its own. This is incidentally the main

reason why so many companies and institutions maintain their own →terminology databases. Even if comparable databases exist, they are of limited use to them, for many reasons, e.g. the languages covered in other databases and the level of detail of the coverage of specific subject fields are not or only partially relevant. Sometimes there are legal or financial barriers: relevant databases might be restricted to internal use, or external use would imply prohibitive costs (licence fees, costs for search operations). Another criterion is the function of an institution or person concerned, e.g. policy maker, scientist, administrator, etc. But there we face the same problem as with the criterion of institutionalisation.

Although we might be able to identify user groups according to such criteria, there will always be an area of *ad hoc* needs or rapid changes in previously identified requirements of individual institutions and experts. The system design will then still require an inherent flexibility in adapting to personal user profiles, a feature that even commercially available →terminology management systems include.

In addition to this multidimensional set of criteria for distinguishing user groups of experts, we are increasingly confronted with users that are not subject field experts, i.e. translators, technical writers, journalists, and other communication experts, and indexers, librarians, information brokers and other information specialists. A case study for such user groups and their needs concerning environmental →terminology was carried out in 1995 in the context of a European Language Engineering Project, called POINTER (Proposals for an Operational Infrastructure for Terminology in Europe; POINTER 1996). One of the conclusions of this case study was to propose networking procedures in the area of managing environmental →terminologies so that different user groups may benefit from each other's work more than in the past.

### 3 General Framework

This chapter focuses on the organisational context that governs the strategies and methodologies of developing and managing environmental →terminologies.

There are complex interactions among various institutions concerned, i.e. national information centres and focal points, European institutions, research centres, etc. The project mentioned above is part of the work programme of the EEA.

#### 3.1 Terminology and Thesaurus

The relationship between the project on the consistency of →terminology and coding and that on the European Environmental Thesaurus is that of mutual support. There is world-wide consensus in information science and →terminology science (as reflected in ISO standards ISO 2788, 5964, 704, 1087) that the →concepts of →*thesaurus* and →*terminology* are separate from each other, but that there is a bi-directional relationship between them. While →*terminology* can be as "a structured set of →concepts and their representations in a particular subject field", a →*thesaurus* is a documentation language that is used for indexing and retrieving information from databases or catalogues. In order to establish a →thesaurus, a specific →terminology must be used and transformed into the usual structure of a →thesaurus. Traditionally a →thesaurus is used by indexers and those who search for information. A →terminology, on the other hand, is analysed by →terminologists, documented and presented in the form of →dictionaries and/or →terminology databases by →terminographers and →lexicographers. It has a totally different structure and presentation form, and different purposes and applications, such as translation, technical writing, but also subject-oriented analysis and description of scientific problems, etc.

The bi-directional relationship is as follows: while →terminology is needed to populate a →thesaurus, a →thesaurus may be used to index and retrieve entries in a →terminology database. While a →thesaurus is by definition created for and used together with a certain information system, a →terminology collection (→dictionary, database) is by definition an open and neutral repository of conceptual and linguistic information with no special application in mind (and is thus multifunctional on the pragmatic level).

### 3.2 Enviroterm

As a bridge between theoretical and strategic considerations and the implementation level, a data model was specified for Enviroterm, the →terminology database. ISO FDIS 12620 “Terminology – Computer Applications – Data categories” (ISO 1997) is applied by using the data categories with their specifications contained in this standard document. This standard is being fully implemented in the scope of this project, as well as ISO FDIS 12200 (ISO 1997). As the database is implemented as a relational database, a meta-coding schema by MELBY and HARDMAN (1997) has been used, in order to stay in line with the ISO standards mentioned.

### 3.3 The Environmental Thesaurus

Currently the →thesaurus contains about 5,000 descriptors. Most of them include →definitions in English, other language versions include German, Italian, French, Spanish, etc. For more information on the →thesaurus refer to Bruno Felluga (felluga@itbm.rm.cnr.it) in Rome and Wolf-Dieter Batschi (batschi@uba.de) in Berlin.

## 4 Working towards a Web-based Collaborative Environmental Hyperglossary

Terminological information is increasingly available in the Web, usually in the form of mono- and multilingual electronic →dictionaries, sometimes with on-line translation features.

With interactive features increasingly available in the Web, several initiatives have started to create collaborative →dictionaries in the Web. To my knowledge the first initiative of this kind was started by Birkbeck College in the U.K.: under the leadership of Peter Murray-Rust and Lesley West, the “Collaborative Virtual Hyperglossary” for biomolecular →terminology has been designed as an integral part of an accredited computer-supported and distance-teaching tool in biochemistry courses of several universities (MURRAY-RUST 1995<sup>1</sup>).

This notion can be taken over and adapted for the environmental field, leading us to propose to implement an “Environmental Hyperglossary,” the Web-based “collaboratory” form of Enviroterm. The interactive mode of providing and retrieving data is managed by HTML-RTF filter as an interface to text processor formats. Data entry forms are provided by Common Gateway Interface (CGI) scripts and by Java applets. Retrieval of terminological data from the local database via the Web will have to be provided.

Two types of user groups are distinguished: active providers and passive consumers, latter group also has the possibility of giving feedback to the central management team. The

---

<sup>1</sup> See also <<http://www.vhg.org.uk/>>.

collaborative aspect is organised and managed by entry forms and validation and consolidation phases for receiving input from different sources, as part of an overall work flow management scheme. The central glossary manager collects, validates and processes all incoming data, e.g. basic entries with monolingual information, some →definitions and some cross-references. The transformation of the data model into input and output formats, interfaces and the management format of the Web site, guarantees a uniform data structure and a formal validation of data according to this specification. The validation of content, however, is a collaborative duty: the Web site offers, but only to the closed group of active providers, entries with the unchecked and uncompleted information, duly marked as “working records”, so that each member of this group can check and complete entries, according to his/her competence (concerning languages and subject field knowledge) and according to specific needs and interests. Consolidated records are then offered for general use on the open Web site to the passive consumers.

The organisation of an Environmental Hyperglossary in the Web could be described as follows:

In a first test phase a certain number of institutions would co-operate on the Web in the following way:

- there is a centre that manages the hyperglossary and the process of its growth, revision and use;
- every institution contributes data according to rules that are agreed upon (work manual as a help database);
- these rules concern technical issues (HTML coding), methodological issues (→terminology data capture and presentation) and administrative issues. The rules shall be simple, straightforward, intuitive and easy to follow and to remember;
- entries provided with provisional status will be checked by the centre on conformity with the rules and revised accordingly. Doubts concerning the actual content (e.g. a →definition of a →concept) will be put on a notice board for general discussion and subsequent joint decision. The entry concerned will be sent to providers as an unconsolidated entry, revised accordingly, and then it is added as a consolidated entry to the main part of the glossary;
- entries will usually be incomplete concerning languages and/or information on a →term/ →concept unit (like →definition, →synonym, restriction etc.). Each member of the group might want to add missing information. While in the first phase it might be desirable to record who contributed which information, it could be turned into a global acknowledgement and authors’ statement at the beginning of the consolidated part of the hyperglossary (this will have to be decided by the group).

In order to facilitate the start-up phase, a certain number of entries are included in the hyperglossary right from the start of the pilot phase. References (hyperlinks) are present to other terminological resources available in the Web, as well as to terminological meta-information, i.e. bibliographic references.

## 5 Conclusion

There is an ongoing and increasing need in harmonizing existing terminologies in environmental sciences and their related fields, and in developing coordinated and collaborative working procedures for the continuous innovation, extension and improvement of terminological resources.

## 6 References

### 6.1 Cited Web Sites

- European Environment Agency  
<<http://www.eea.eu.int/>>
- Ministry of Environment of Lower Saxony, Germany, project leader of European Topic Centre for Catalogue of Data Sources  
<<http://www.mu.niedersachsen.de/system/cds/>>
- International Society for Environmental Protection (ISEP) with its main project Central European Data Request Facility (CEDAR)  
<<http://www.cedar.univie.ac.at/>>
- POINTER home page  
<<http://www.mcs.surrey.ac.uk/Research/CS/AI/pointer/>>
- Virtual HyperGlossary  
<<http://www.vhg.org.uk>>

### 6.2 International standards

- ISO 704: 1987: Principles and Methods of Terminology
- ISO 1087: 1990: Terminology – Vocabulary
- ISO 2788: Documentation – Guidelines for the establishment and development of monolingual thesauri
- ISO 5964: 1985: Documentation – Guidelines for the establishment and development of multilingual thesauri
- ISO 8879: 1986: Information processing – Text and office systems – Standard Generalized Markup Language (SGML)
- ISO 10646: 1993: Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane
- ISO FDIS 12200: 1997: Terminology – Computer Applications – Machine-Readable Terminology Interchange Format (MARTIF)
- ISO FDIS 12620: 1997: Terminology – Computer Applications – Data Categories

### 6.3 Other References

- European Environment Agency (EEA), 1995: Multi-annual Work Programme 1994–1998. Copenhagen.
- MELBY, A.; HARDMAN, D., 1997: The Reltef™ Specification (unpublished manuscript).
- MURRAY-RUST, P., 1995: The Collaborative Virtual Hyperglossary. Paper presented at Tampere University, December 1995 (unpublished manuscript).
- POINTER, 1996: Final Report. Proposals for an Operational Infrastructure for Terminology in Europe. Brussels.