The Croatian National Termbank Struna – the workflow of a terminological project

Ana Ostroški Anić
Institute of Croatian Language and Linguistics, Zagreb
E-mail: aostrosk@ihjj.hr, ana.ostroski@gmail.com

1 Introduction

The Development of the Croatian Special Field Terminology program (known by its Croatian acronym Struna) began in 2007, and was financed by the Croatian Science Foundation until 2013. It is being carried out at the Institute of Croatian Language and Linguistics, which was chosen to serve as the national coordinator. The main aim of Struna is to gradually make standardized Croatian terminology for all professional domains available to the public by means of coordinating the work of domain experts, on the one hand, and terminologists and language experts, on the other. The termbank e-Struna developed for that purpose is a national terminology database intended as the primary tool for the implementation of a more or less explicitly recognized approach to terminology planning.

The specific nature of the cooperation between the domain experts and terminologists required a methodological framework for practical terminographic work. A uniform approach to the terminological description guaranteed by a national terminology project environment has been crucial to making the various subject fields in the database as structured and uniform in description as they possibly can be (Bratanić and Ostroški Anić 2013). On the other hand, the variety of the domain knowledge included in Struna, and the various characteristics of each domain – its conceptual structure and dynamics, specific communicative settings and intended users – have called for the adjustment of initial terminological principles as well as for the modification of the existing and the introduction of new data categories.

Struna was released online in February 2012, and includes today the terminologies of 18 domains with about 107 000 terms (30 000 of which are preferred Croatian terms, 14 000 synonyms and 61 000 equivalents in English and other languages). Every domain terminology included in Struna is the result of terminographic work done in a separate project over a period ranging from 12 to 18 months. The workflow of every project differed according to the number of people included in the data acquisition and the data analysis, and based on the unique characteristics of the special field the terminology of which was being compiled. However, a general outline of a particular terminological project within Struna will be presented, along with suggestions for modifications and improvements in the existing terminographical workflow.

2 Lexicographical workflow

2.1 Preparation

In order to start working on the compilation and description of a particular domain's terminology within Struna, domain experts interested in such work needed to apply for funding from the Croatian Science Foundation. The phase of planning started with the project proposal in which the intended project manager – usually a university professor – was obliged to submit, among other information, an organizational plan with an elaborated schedule of the workflow and the description of their team members' responsibilities on the project. If chosen for funding, the project would then enter the second phase of planning, one that was led by the terminologists and language experts at the Institute of Croatian Language and Linguistics. This included meetings with the project manager in order to establish the course of cooperation, and to give guidelines on the organization of the work, e.g. how to allocate team members to individual subdomains, how to gather material for the corpus, analyze the sources, prepare a brief map of key concepts of the domain and their conceptual relations, advise on writing a sample of terminological entries to be analyzed by the terminologists before the process of data acquisition, etc. Since the majority of data acquisition and part of data analysis was to be done by domain experts who had little or no experience in terminological work at all, further instruction was provided through thematic workshops (e.g. on general terminological principles, on writing terminological definitions, common orthographic and other linguistic mistakes, etc.) and through individual meetings.

2.2 Data acquisition

The majority of the projects within Struna had no computerized corpus at their disposal for the extraction of relevant information. Those that did decide on compiling a corpus for this purpose had chosen to do so using Wordsmith tools, and applying basic methods in searching the corpus and analyzing term concordances. Their corpora consisted mostly of university textbooks and related course materials, relevant MA and PhD theses and papers from well established

scientific journals. In any case, specialized dictionaries, lexicons and encyclopaedias sometimes served as an additional source of information. Generally speaking, textbooks are the best source for knowledge-rich contexts (Meyer 2001) and for acquiring terminological definitions. The advantage of having higher education professionals in a team is that many of them are authors of such course materials, and could therefore provide primary sources in an electronically readable format.

2.3 Computerisation

Most textbooks, scientific papers and other material used for corpus compilation – but also as a source of manual term extraction – are usually available from the authors and received in readable formats. Images are manually entered into the database as well as formula, symbols, diagrams, and other similar information.
The database used for the entire terminographic work is a relational database designed in a MySQL environment. Its structure was designed in accordance with the TEI P5 guidelines for text markup and the TBX standard format for the representation and exchange of terminological data. It is particularly relevant for a terminological database to be exchangeable with similar terminological resources (Wright and Budin 2001).

2.4 Data processing

Term extraction is usually done by domain experts, except in cases when terminologists work with experts on compiling a list of term candidates. So far, no term candidates have been extracted by applying methods of automatic term extraction alone.

2.5 Data analysis

The process of data analysis can be divided into several phases, depending on the team members included in the analysis. The analysis encompasses: a) writing terminological entries (done by domain experts and edited by the project manager, who is an expert herself/himself), b) editing the entries (done by terminologists who amend the definitions, check term synonyms and equivalents, and proofread the entries), and c) the final revision (done by a senior terminologist and the project manager) during which more entries can be added in case the conceptual system defined requires further elaboration through concept definitions. It is during this stage that possible term duplicates are detected, missing term equivalents added, related concepts supplied or corrected, and the harmonization of conceptual information done both on the project level and between related projects.

2.6 Preparation for online release

The final stage of the data analysis overlaps with the preparation for online release. The main task of this phase is, however, to ensure another proofreading (usually of printed versions of the entries) and check that all multimedia information is functional and that the graphic display of the entries is displayed properly. Although during this phase entries can be corrected or deleted (Klosa 2013), this is usually avoided since these activities are mainly done in the final stage of the data analysis. In most cases, terminology of an entire project is released at the same time.

2.7 Afterlife

The phase of terminology update and maintenance hasn't yet been defined in Struna in much detail. So far it has depended on individual efforts of terminologists and project managers of completed projects. User response has been the most valuable source for possible updating of the entries. In most cases the updating referred to correcting definitions as far as their content is concerned. Such initiatives are still scarce, so a better approach would be to envisage a periodic checking for changes in the field described and included in Struna. The planning of the domain terminology regular updates and amendments should be agreed upon and described in detail before the end of the project. Tasks included in a plan of the termbank's afterlife management would have to include a description of periodic meetings with the project managers (possibly in the form of a Terminology Council or a similar body of distinguished domain experts that took part in the work of Struna), regular user analyses and monitoring the publication of newer editions of primary sources of information for the terminological entries of every special domain.

3 Time span of the different phases

Duration (months)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Preparation																
Data acquisition																
Computerisation																
Data processing																
Data analysis																
Preparation for Online release																
Afterlife																

Table 1 Process phases of dictionary project and their time span

4 References

Bratanić, M. and A. Ostroški Anić (2013). The Croatian national termbank Struna: A new platform for terminological work. *Collegium antropologicum*, 37(3), 677-683.

Klosa, Annette (2013) The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus H./Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herberst Ernst (Hgg.): Dictionaries. An international Encyclopedia of Lexicography. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin, Boston: de Gruyter, S. 517-524. (Handbücher zur Sprach- und Kommunikationswissenschaft; 5.4).

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography. A conceptual and methodological framework. In: D. Bourigault, C. Jacquemin and M. L'homme (Eds.). *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins.

Wright, S. E. and G. Budin (2001): *Handbook of terminology management. Volume 2: Application-oriented terminology management*. Amsterdam: John Benjamins.