Terminology Planning for the Croatian National Terminology Database STRUNA

Ivana Brač and Maja Lončar

Institute of Croatian Language and Linguistics, Republike Austrije 16, Zagreb, Croatia
ibrac@ihjj.hr, loncarm@ihjj.hr
http://struna.ihjj.hr

Abstract. The paper examines issues that have emerged during the development of the Croatian national terminology database Struna as the primary tool for the implementation of national terminology policy. Struna has been accessible over the Internet with a simple, searchable database (struna.ihjj.hr) since February 2012. The language and terminology situation in Croatia is discussed. The Croatian national terminology database Struna and a rather unique model of collaborative terminology work is described. This case study aims at helping other similar terminology projects benefit from our experience.

Keywords: Croatian terminology, national terminology database, terminology planning, terminography

1 The Language Situation in Croatia

Due to Croatia's specific political circumstances, the Croatian language was in a subordinate relationship to other languages not native to Croats during certain periods of history. Also, different parts of Croatia were under the influence of other languages due to either geographical proximity or political power. The Latin language had a strong influence on all parts of Croatia, the Italian language prevailed in Dalmatia (southern Croatia), German and Hungarian prevailed in northern Croatia, while in Slavonia (eastern Croatia) Hungarian and Turkish were the most influential. Some of these languages were official languages in administration and public life. Until 1848, Latin was the official language of administration, a situation supported by Croatian politicians in order to oppose the idea presented in the Parliament in 1790 that official records should be written in Hungarian. By maintaining Latin as the official language, they held that the Croatian language was in less danger from a dead language than a living one. The period between 1850 and 1859 was a period of absolutist rule known as "Bach's absolutism", and the official language in schools and offices was German. Ban Josip Šokčević introduced the Croatian language in schools as obligatory subject in 1861 and as the official language of administration in 1868.

1 The Hungarian language was introduced in schools as obligatory subject in 1827.
2 Governor
Since Croats were unable to use Croatian as the official language for a long time, Croatian terminology – particularly legal, political, and economic terminology – did not develop, and so the need for the creation of such Croatian terminology arose. The foremost figure in the creation of Croatian scientific terminology at the time was Boškoslav Šulek, who felt that it was sometimes better to coin a Croatian word despite its not being the best possible choice since it is difficult to remove foreign words from the language once they are adopted. The process in which terminology work was performed is substantially different from today’s. The Regency Council compiled a list of German and Latin scientific terms from all professions, which was then sent to all grammar school principals, teachers, writers and various experts, who supplemented the list with Croatian equivalents. After two years, the list was returned to the Regency Council, which was composed of the most important linguists and scholars of that time, who spent a month and a half checking the resulting material. They were led by the principle [1, p. III]: "the dictionary should avoid extremes, especially excessive purism and unnecessary classicism." Among the most important of Šulek’s works are Njemačko-hrvatski rječnik (1860), Hrvatsko-njemačko-talijanski rječnik znanstvenoga nazivlja (1874-1875) and Jugoslavenski imenik bilja (1890). When creating dictionaries, Šulek relied on old Croatian dictionaries and lexemes from the literary language, and when he could not find an appropriate word, he first turned to the dialects (the Croatian language has three dialects: Štokavian, Kajkavian and Čakavian), then to other Slavic languages, and eventually he created his own word. This was followed by a period in which attention was paid to the convergence of the Croatian and Serbian languages, and in this process the development of Croatian terminology was neglected.

Article 12 of the Constitution of the Republic of Croatia, which was adopted in 1990, stipulates that the Croatian language and Latin script are official, and Article 10 of the Constitutional Law on national minority rights states that persons belonging to national minorities have the right to freely use their language and script, privately and publicly. The Croatian language is in a specific position in relation to Bosnian and Serbian. For example, the Hague Tribunal refers to the three languages as Bosnian/Croatian/Serbian, even though they are distinct languages which have often been studied under similar names at European and other world universities. The Charter of Fundamental Rights of the European Union states that: "Every person may write to the institutions of the Union in one of the languages of the Treaties and must have an answer in the same language" (Title III Article 41). Or as Laurén, Myking and Picht [2, p. 1] conclude: "we may say that a fully developed language is a democratic right and therefore also a human right."

---

3 Translated by authors: "...valjalo se je čuvati dvijuh skrajnosti toga rječnika, naime preterana purizma i nepotrebna klasicizma."
4 German-Croatian Dictionary (1860), Croatian-German-Italian Dictionary of Scientific Terms (1874-1875), Yugoslav Plant Lexicon (1890)
5 http://narodne-novine.nn.hr/clanci/sluzbeni/232289.html
6 http://narodne-novine.nn.hr/clanci/sluzbeni/310287.html
7 "Pripadnici nacionalnih manjina imaju pravo slobodno služiti se svojim jezikom i pismom, privatno i javno."
Today, the Croatian language, like other (minor) languages, is in danger from the English language in certain areas, which can lead to domain loss [2, p. 5]: "loss of ability to communicate in the national language at all levels of an area of knowledge deficient because of the further development of the necessary means of professional communication." For example, the Ministry of Science, Education and Sports rates scientific papers published in English or any other foreign language higher than those published in Croatian. Also, an increasing number of classes at Croatian universities are being taught in English\(^8\), which does not allow Croatian scientific terminology to develop. The same problem exists in business as well, because companies that operate in Croatia whose managers do not speak Croatian use English as their official language, contrary to the constitutional provisions. Therefore, Croatian employees use English terms in daily communication, thus neglecting the need for the creation of Croatian terms. According to UNESCO [3, p. v] "people whose mother-tongue is not (or not sufficiently) developed from the point of view of terminology and special purpose languages or who are denied the use of their mother-tongue in education and training, for accessing information, or interacting in their work places, tend to be disadvantaged." The Croatian language is a language with rich inflection in which terms are easily formed by compounding, deriving, calquing and abbreviating.

Croatia became a sovereign and independent state on June 25\(^{th}\), 1991 through a decision of Croatian Parliament. In 2008, seventeen years after declaring independence (16 years since being internationally recognized), the Croatian language was internationally recognized when the National and University Library in Zagreb, together with the Croatian Standards Institute, the National Library of Serbia and the Institute for Standardization of Serbia, sent a request to the International Organization for Standardization to change the international language codes (alpha-3 code) [5] from scr (Serbo-Croatian-Roman) to hrv (Croatian) and scc (Serbo-Croatian-Cyrillic) to srp (Serbian).

2 The Terminology Situation in Croatia

"The standard language should be able to express all human knowledge, concrete and abstract, people must be able to speak in it confidently about the entire universe."[6] Therefore, it is important to develop terminology for areas currently under rapid development and for whose concepts English terms are currently being used.

During the 20\(^{th}\) and 21\(^{st}\) centuries, special dictionaries were created in Croatia, most of which are monolingual and in the fields of literature, philosophy, law, zoology, medicine, etc. However, this has not been followed by any systematic attempt at harmonizing terminology throughout the Croatian language and covering terminology from all fields. Standardization at the national level began in Croatia in 1992 at the State Office for Standardization and Metrology. In 2004, a Croatian Government act (Regulation on the Establishment of the Croatian Standards Institute) established the Croatian Standards Institute as an independent, non-profit public institution. Its aim is to serve as a national standardizing body and to increase the level of standardization in order to improve the safety of products and processes, protect health, human lives

\(^8\) See more about in [4].
The Croatian Standards Institute absorbed the infrastructure of the State Office for Standardization and Metrology, as well as its material resources and experts, who had been working on standardizing the following areas: basic standards, architecture, engineering, general electrical engineering, electrical engineering, electronics, telecommunications, computer science, metallic materials, non-metallic materials; health, environment, and medical equipment; transportation, materials handling and packaging; services, products for household and leisure; chemicals, chemical engineering, and food products.

It is worth mentioning that the State Office for Standardization and Metrology translated Heidi Suonutti's manual "Guide to Terminology" into Croatian in 1999 [7], and the Croatian Standards Institute, in cooperation with the Institute of Croatian Language and Linguistics, translated and published UNESCO's "Guidelines for Terminology Policies" into Croatian in 2009 [8].

The Miroslav Krleža Institute of Lexicography has made approximately 60,000 entries from the General and National Encyclopaedia (20 volumes) available on its website9, as well as 53,355 articles from the Croatian Family Lexicon and a number of entries from its other publications: Encyclopaedia of Istria, Film Lexicon, Football Lexicon and the Croatian Biographical Lexicon10. Their approach is however lexicographic and encyclopaedic.

3 Terminology Policy

Aware of all the problems that may occur due to negligence of terminology in the Croatian language, the Ministry of Science, Education and Sports' Council for Standard Croatian Language Norms, at the initiative of Dr. Dunja Brozović Rončević, sent a request to the minister in 2007 to initiate a project entitled Izgradnja hrvatskog strukovnog nazivlja (Development of Croatian Special Field Terminology, referred to here by its Croatian acronym Struna), on which subject field experts and terminology and language experts would work jointly. National Foundation for Science, Higher Education and Technological Development of the Republic of Croatia (today called the Croatian Science Foundation) decided to fund the project within the frame of its program Socio-Cultural Transition from an Industrial into a Knowledge-Based Society because "a country's relative level of development can be measured by the average capability of its citizens to use information for the sake of knowledge transfer and capacity building" [3, p. 2].

The Foundation announced a call for proposals for a national coordinator, for which the Institute of Croatian Language and Linguistics (IHJJ) was chosen as a public institution to carry out the planning and implementation of national terminology policy. Dr. Milica Mihaljević was in charge of the first phase of the coordination project (2008). UNESCO [3, p. 15] recommends the active participation of those institu-

---

9 http://proleksis.lzmk.hr/
10 http://enciklopedija.lzmk.hr/
tions that are most dedicated in their commitment to language planning policies. Official approval for the final terminology policy was officially adopted, and a mandate was given to the implementation body to proceed with the implementation process [3, p. 29]. The project has been led by Prof. Dr. Maja Bratanić since 2009.

During the project coordination phase, an in-house terminology manual and terminology database (e-Struna) with a first version terminology management system were created\(^\text{11}\), and training workshops for the terminological education of subject field experts and future terminologists were organized. The initial principles used for terminology work are the following [9, p. 10]:

1. Terminology is part of the Croatian standard language.
2. Both subject field experts and linguists should be involved in the creation of terminology.
3. Linguists should participate in creating terminology from all fields, and subject field experts should participate in the creation of terminology for their respective fields.
4. Terminology from different fields should be mutually harmonized.
5. Synonymy is undesirable, and relationships should be established among synonyms (preferred term, admitted term, deprecated term, obsolete term, colloquial term).
6. In order to create a terminology system, it is necessary to define all terms.\(^\text{12}\)

UNESCO [3, p. 13] emphasises that "in collaborative terminology work, best results are achieved – especially for language development and terminology standardization – when linguists, terminologists, and domain experts work together in committees." The working methods for implementing this terminology policy differ to an extent from the working methods typical for national terminology centres. The selection of projects is based on public calls for projects. The Croatian Science Foundation announces calls twice a year for which "experts employed at universities, faculties, polytechnics, public institutes, professional associations, non-governmental organizations and state agencies, or other experts on legal basis involved in the work of the stated institutions, independently or in cooperation with other institutions and/or experts from the country and from abroad”\(^\text{13}\) may apply. The estimated budget for each project is 100,000 HRK (approx. €13,300). A project description, financial plan, work plan, CV and list of publications of the project leader and of all associates, and proof

---

\(^\text{11}\) UNESCO [3, p. 3] states that "in terminology planning and in particular in the framework of a national terminology policy, a national terminology database often is used as one of the primary tools for the implementation of that policy".

\(^\text{12}\) "1. nazivlje je dio hrvatskoga standardnog jezika, 2. u izgradnji nazivlja trebaju sudjelovati stručnjaci pojedinih struka i jezikoslovci, 3. jezikoslovci trebaju sudjelovati u izgradnji nazivlja svih struka, a stručnjaci pojedinih struka u izgradnji nazivlja svoje struke, 4. nazivlja različitih struka trebala bi biti međusobno usklađena, 5. sinonimija u nazivlju nije poželjna i trebalo bi uspostaviti odnose među sinonimima (preporučeni naziv, dopušteni naziv, nedopušteni naziv, zastarjeli naziv, žargonski naziv) 6. da bi se mogao izraditi terminološki sustav, potrebno je sve navedene nazive definirati."

of employment of the project leader must be included with candidates’ applications. The project lasts for one year, and project managers are required to submit periodical and final project reports (as well as financial reports) and to enter at least 1000 terms into the database.

Projects are evaluated by the Evaluation Committee and the Institute of Croatian Language and Linguistics as the national coordinator both semi-annually and at their conclusion. The evaluation analyses the results achieved in relation to planned objectives, financial justification of costs, investments in education and public presentation of the project. The evaluators are the foremost Croatian linguists. This is consistent with the ISO 15188 [10] recommendation that the external funding body to be reported to carry out an evaluation at the end of each project.

Project management guidelines for terminology standardisation set out in ISO 15188 [10] divide the management process into preparation, design, implementation and review phases. The preparation and design phases occur during the application procedure. The implementation phase consists of collecting and recording terminological data. The final phase is the preparation of the final report and financial review.

Both the call for projects and evaluation highlight the importance of cooperation between subject field experts and terminology and language experts. At the beginning of each project, workshops are organized in which subject field experts become familiar with the basics of the terminology work and with the most frequent errors in language and in definitions. They simultaneously receive an introduction to the principles of the terminology database and learn how to enter and edit term records into it.

Priorities have been set for those disciplines that are developing the fastest – computer science, economics, finance, EU acquis etc. – but the projects that are selected are those that most successfully meet the conditions prescribed by the call for projects. Calls are still being developed in order to more accurately define which fields require the creation of new terms and to establish for which specific areas dictionaries and other terminological resources are already available for implementation in the database.

The projects selected in the first four rounds have covered the fields of chemistry, aviation and air traffic, cartography and geoinformatics, dental medicine, polymers, corrosion, marine engineering, EU law, physics, civil engineering, mechanical engineering, anthropology and maritime studies. Fifth-round projects will commence in mid-June of 2012, and will deal with terminology from the fields of anatomy and physiology, hydraulics, mathematics, and archaeology. Projects vary in size considerably: from 600 (polymers) to 4,500 (dental medicine) terms. This is not fully consistent with ISO 15188 [10] recommendation that the number of concepts per workgroup should be less than 200. Our experience has shown that those projects which include over a thousand terms are difficult to handle and harmonize with other projects, or rather that they require the collaboration of a greater number of terminologists and subject field experts.

Besides those projects financed by the Croatian Science Foundation, the Institute of Croatian Language and Linguistics has been cooperating with the Croatian Standards Institute, which has made twenty-five ISO Croatian standards available to Struna. The database currently includes about 32,000 terms.
4 National Terminology Database

Since the Institute of Croatian Language and Linguistics is the only national institution in charge of terminological work in Croatia, it is important to develop effective coordination and workflow management. "With computer-assisted cooperative (and network-based distributed) terminology work, the preparation, processing and maintenance of terminological data can be carried out faster, more efficiently and in line with modern quality management." [3, p. 20]

Project leaders usually have some or even considerable experience with lexicographic work. However, in most cases, the subject field experts have not been trained for terminology work, so the database interface is designed to be as user friendly as possible. It is designed as a tool to record, process, store, and output terminological data. The database was made in a MySQL environment and is designed to be easily improved and updated when necessary (and improvements are indeed made continuously). As UNESCO [3, p. 33] stresses, "the procedures as well as the respective workflow management for net-based computer supported cooperative work should be designed very carefully in order to become efficient and effective; this may need some time for implementation." The first version of e-Struna, which was created in 2008, had 18 data categories and a simulated dictionary format. The fourth version, currently in use, has 46 categories and meets all requirements for processing terminology entries for a total of 16 different terminology collections. Struna’s relational database schema is compatible with TBX (defined by ISO 30042 standard [11]), ensuring easy data exchange with all other digital terminology collections in the world.

The database interface includes many categories with drop-down menus and category correspondence, in which terminologists and language experts can discuss various issues.

Apart from the administrative data categories recorded automatically, e-Struna's current data categories are:

- preferred term (and short form of the preferred term)
- source of the term
- subject field and subfields
- interdisciplinary term label
- grammar/part of speech information (noun (grammatical gender: masculine, feminine, neuter; grammatical number: singular, plural), verb, adjective, adverb, multi-word terms, name)
- new term label
- working definition
- definition
- source of the definition
- context
- source of the context
- synonyms (admitted, proposed, deprecated, obsolete, colloquial)
- antonym
- subordinate term
Terminology Management

The project has a rather unique model of collaborative terminology work between subject field experts and terminology and language experts, which can be divided into six phases (Fig. 1).

1. A subject field expert enters a term with appropriate data categories – to be more specific, out of 46 available categories, the obligatory fields are: term, definition and English equivalent. The term is then forwarded to the editor.
2. The editor, usually the project leader, edits the entry and either returns it to the term author for improvement or transfers it to the terminologist.
3. The first terminologist edits all the entered data categories in accordance with terminological principles.
4. A language expert corrects orthographic and grammatical errors and in certain situations suggests alternative proposed terms which s/he considers to be a better fit with either the Croatian word formation principles or more eligible than the existing preferred term based
on some other criteria. (5) The second terminology expert checks whether the language editing has changed the meaning of the definition. (6) The editor finally reviews the proposed changes (the database interface having been designed to accept changes very easily) and locks the term. If the editor does not want to accept all changes, the main definition can be adapted. If the editor does not accept the changes at all, the term in question will appear in a different colour in the work interface, making it easy for the terminologist to see that the working definition has not been accepted. The project manager has the final say in locking the term as the terminology and language experts cannot independently decide on a definition’s validity.

At the end of this process, the editor-in-chief of Struna decides that the new terminology collection can be considered finalized and included into Struna, and it then becomes accessible through the search engine.

6 STRUNA as an Online Information Service

Struna is accessible over the Internet through a simple search of its database since February 2012. The website was designed to be visually attractive and user-friendly. The search engine and interface were created using HTML5, CSS3 and the library jQuery.

It is possible to search for terms in definitions, context and notes if advanced search is selected. Finding terms is facilitated by introducing suggestions into the search engine. Foreign terms (or their constituent parts) written in their original forms are highlighted in italics. Along with this, the corresponding language label indicating the term’s origin is displayed using the ISO language codes.

An enquiry form is available on the website, and all users are encouraged to address their comments or questions to the Struna team.

Terminological data, from a formal point of view, represents specialized knowledge at the level of concepts, and can be offered through online information services [3, p. 19]. None of the subject field terminology collections have yet been published as hard-copy dictionaries or glossaries (in line with the worldwide shift from conventionally published forms to on-line resources).

7 Challenges and Misconceptions

UNESCO [3, p. 11] distinguishes two approaches when dealing with terminology: descriptive and prescriptive terminology work. "While descriptive terminology work only observes and analyses the emergence of terms, prescriptive terminology work constitutes an agreement by users to adopt a term for common and repeated use in given circumstances. The latter comprises terminology unification, standardization and harmonization. Terminology standardization almost always involves a choice among competing terms.”

The Institute was chosen as the national coordinator to standardize Croatian terminology. This role allows the Institute’s language experts to suggest replacements for existing terms by proposing new terms they consider to be better-motivated, more
transparent, more precise, or more productive in terms of the formation of related
terms. All terms for the same concept are listed, but they are ordered by their norma-
tive status. The preferred term, considered as the best choice by subject field experts,
differs from its synonyms in the following aspects:

- admitted term – a term used by subject field experts, but due to linguistic reasons
  or overlap with similar terms in other fields, not selected as preferred term. This
  term should not be used in definitions of other terms.
- deprecated term – a term that contains a linguistic error or could be confused with
  another term in the same field or other fields; or a term that inaccurately or incom-
pletely expresses the concept to which it refers.
- obsolete – a term which is no longer used, but is important to record because of
  older literature.
- colloquial term – a term used by experts in everyday informal communication.
  Most are loan words.
- proposed term – a new category introduced for cases where subject field experts do
  not accept the language experts' solution for a preferred term. If language experts at
  the Institute of Croatian Language and Linguistics think that a preferred term is in-
adequate due to linguistic or semantic reasons, they suggest a proposed term. If
  subject field experts accept a suggested term, it becomes a preferred term, and if
  they do not accept it, it remains in the proposed term category.

Most disagreements between subject field experts and language experts are caused
by proposed terms. Sometimes the subject field experts have difficulties accepting
proposed terms that are terminologically and linguistically more acceptable and insist
on terms already established in a subject field. Although "the replacement of well-
known and frequently used terminology would cause confusion to users already fa-
miliar with a specific term rather than improve communication goals" [12, p. 130],
IHJJ acts as the national coordinator entrusted with authority over terminology stan-
dardization. Therefore, it is sometimes necessary to accept a proposed term, which is
in line with UNESCO's statement [3, p. 17] that "it is important to have effective co-
ordination and workflow management for the documentation of terms in order to
avoid duplication and eventually to promote standardization". Struna is, therefore, a
prescriptive terminology database with a clearly assigned normative function, while
also being descriptive in its approach. As we move from theoretical categorisation
towards practice, i.e. towards the ontological level, simple boundaries are blurred and
both dimensions need to be taken into account [13, p. 90]. "Although the process of
harmonizing concepts and standardizing terms implies prescriptive terminology work,
harmonization must always follow only after descriptive work had taken place." [14]
Or in the words of H. Picht [cited by 13, p. 90]: "normative and descriptive terminol-
ogy work are not opposites – they are possible complementary stages."

The Struna terminology database was initially envisioned as a primarily monolin-
gual terminology resource aimed at Croatian LSP terminology standardization. Bear-
ing in mind its enormous global communicative power, potential for data sharing and
term base exchange, as well as the ever-growing needs of translators, it was eventu-
ally decided that the database should be multilingual, with definitions only in Croatian
and term equivalents in at least one foreign language (English), or preferably several other languages. Furthermore, foreign language equivalents serve as a control mechanism for harmonization of terms across domains. This is in line with UNESCO’s [3, p. 21] recommendation that the more a terminology policy is conceived as multilingual and non-exclusive from the outset, the larger potential markets can become.

Although processing terminology in a national terminology database (as opposed to a professional or field terminology database) has some obvious advantages (uniform terminographic description, easier harmonization of concepts, especially in interdisciplinary fields, faster term validation, etc.), a certain level of discrepancy in the approach has yet to be overcome. For example, concept overlap is difficult to avoid. Insufficient coordination in the delimitation of various domains results in cases of multiple records of equivalent concepts and terms in the database. In order to reduce such cases, the data category "interdisciplinary term" was also introduced. With the large number of people working on the database (currently more than 150), coordination among projects is becoming increasingly difficult. There is a certain discourse variation between domains (e.g. Latin terms in medical terminology, more appellations in EU law than are usual in terminology etc.), but so far no major discrepancies between projects have been detected.

A specific profile for Struna users is not implied, since the profile of its intended user is difficult to envision. It is left to field experts to predict the main category of users and their needs, which has repercussions on the terminological procedure (especially on the metalanguage of definitions). The experience of other terminology centres has shown that different user groups for different purposes require differing levels of complexity of terminological data. UNESCO [3, p. 4] therefore recommends that it is "highly economical to prepare multi-purpose terminological data from the outset for different users and applications."

8 Conclusion

By establishing standardized subject field terminology, Struna aims at improving knowledge and information circulation while providing unambiguous communication within and across professional domains. This is important for the status of Croatian as a standard language, soon to become an EU language.

The aim of this paper has been to point out some specifics that occurred during the development of the Croatian national terminology database. Budin [15, p. 89] mentions that "comparative and evaluative case studies on the reasons for the success and/or failure of language planning programs or parts thereof are, however, badly needed to optimize these efforts". Apart from Bhreathnach’s doctoral thesis [16], not many similar efforts seem to have been made over the course of the past twenty years.

UNESCO [3, p. 4] emphasises that "the study of existing examples and experiences elsewhere in the world are highly valuable in this process, and may be extremely useful, not the least in order to avoid traps and pitfalls others have already encountered, bearing in mind local sensitivities."
References:

1. Šulek, B.: Hrvatsko-njemačko-talijanski rječnik znanstvenoga nazivlja: osobito za srednja učilišta. Tiskom narodne tiskare Dra Ljudevita Gaja, Zagreb (1874)
6. Petrović, B.: Normativnost hrvatskog jezika. Lecture. Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb (2009)