APPROACHING THE EU: TERMINOLOGY AND TRANSLATION DIGITAL RESOURCES

Andrijana ĐORDAN
ALFA BK University, Faculty of Foreign Languages, Belgrade, Serbia

Abstract: The Republic of Serbia applied for the European Union membership in 2009, received full candidate status in 2012 and started the membership negotiations in 2014. The accession process constantly keeps showing a need for high quality translation of the extensive EU documentation. This paper deals with the existing digital resources created from the beginning of the accession process. It focuses on two main projects, developed for translating EU legislation (*acquis communautaire*) into Serbian: *Evronim* – multilingual termbase and *Evrotekta* - bilingual English-Serbian corpus of EU legislation texts. The paper will also bring some reflections on the perspectives of similar digital language resources, as results of language policy and planning activities.

Keywords*: terminology, translation, digitalization*, *Evronim*, *Evroteka*, language policy and planning, EU integration.

1. Introduction

The process of the accession to the European Union has been recognized by the Republic of Serbia as an opportunity for making and implementing series of reforms in all areas of society, in order to provide a better quality of life for its citizens and future economic development. For these key reasons (but also for many others), the Government of the Republic of Serbia has pointed out EU membership as its top foreign-policy priority. The way to a full EU membership is long and complex, and we could say that Serbia stepped into it at least 10 years ago, but it is still on the beginning stages. In order to provide a clearer picture of the complexity of the task, we will mention some key dates in the process:

(1) 2008 → the Stabilization and Association Agreement (SAA) between the EU and Serbia was signed;
(2) 2009 → Serbia submitted its application for EU membership;
(3) 2011 → the European Parliament ratified in Strasbourg the SAA between the EU and Serbia;
(4) 2012 → The European Council has granted Serbia membership candidate status;
(5) 2013 → the SAA entered into force;
(6) 2014 (January) → the First Intergovernmental Conference was held in Brussels, which marked the formal start of Serbia's negotiation to join the EU;
(7) 2014 (December) → the Second Intergovernmental Conference was held in Brussels, which marked the opening of Chapter 32 and Chapter 35;
(8) 2016 (July) → the Third Intergovernmental Conference was held in Brussels which marked the opening of Chapter 23 and Chapter 24;
(9) 2016 (December) → the Fourth Intergovernmental Conference was held in Brussels, which marked the opening of Chapter 5 and Chapter 25;

(10) 2017 (February) →the Fifth Intergovernmental Conference was held in Brussels, which marked the opening of Chapter 20 and Chapter 26.

Even from this brief Serbian EU integration history and only 8 opened chapters (8 out of 35) it is very clear that the accession process involves an extremely extensive volume of documentation and this is the point where the issue of language starts to play a very significant part. All the documentation in fact is the *acquis communautaire* - the accumulated legislation, legal acts, and court decisions which constitute the body of EU law. The EUR-Lex Glossary[i] defines it as:

"The EU's 'acquis' is the body of common rights and obligations that are binding on all EU countries, as EU Members. It is constantly evolving and comprises:

(1) the content, principles and political objectives of the Treaties;
(2) legislation adopted in application of the treaties and the case law of the Court of Justice of the EU;
(3) declarations and resolutions adopted by the EU;
(4) measures relating to the common foreign and security policy;
(5) measures relating to justice and home affairs;
(6) international agreements concluded by the EU and those concluded by the EU countries between themselves in the field of the EU's activities".[ii]

The acquis has to be transposed into the national legislation in order to become the national law. Translating this documentation, a country demonstrates that it is able to become a functioning EU member. This clearly implies that the process of the acquis translation into the national language is one of the essential. Since this extensive volume of the translated documentation is going to be used in the negotiations, it directly influences a country's future.

From the all facts exposed above there is no doubt that is absolutely necessary to provide not any translation, but the translation of the highest quality.

Apart from *high quality*, a very important characteristic of EU translation are *speed* and *accessibility*. We have to bear in mind that the EU has 24 official and working languages (when it is about translation/interpretation, we are dealing with more than 250 language pairs and more than 550 language pair directions) and that the main task is to ensure a quick access to relevant information in all these languages.

In order to speed up and facilitate the translation, but also to ensure a constant high quality result and unified and functional terminology, there have been developed various digital tools and resources. In this paper we will present two relevant for the Serbian language – *Evronim*, a multilingual terminology database and *Evroteka*, a bilingual corpus.

2. *Evronim* – a multilingual termbase

Evronim represents a multilingual terminology database (termbase), made in the process of translation of the acquis into Serbian language. Obviously *Evronim* has been created as a product of translation, but it has also has become one of key tools in the same process. The aim of its creation is to collect EU terminology (which includes a wide range of areas) in Serbian and to make it correct, clear and unified. In this way *Evronim* becomes a strong support to translators, a reliable tool for all the parties included in EU integration process, but also for larger groups of different users in need for high quality terms.

The aquis includes regulation of a large number of different areas, so the terminology included in *Evronim* comes from all those fields. The creators of the termbase remind that in same cases a term can have different meanings in different areas and *Evronim* treats it as separate entry.

Apart from being world *lingua franca*, English is the language of the most EU documents, so *Evronim* will always ofer a term in English, its Serbian translation and, where possible, its equivalents in other languages. The terbmbase also contains entries in the following official EU languages: Bulgarian (BG), Czesch (CS), German (DE), Greek (EL), Spanish (ES), French (FR), Hungarian (HU), Italian (IT), Polish (PL), Romanian (RO), Croatian (HR), Slovak (SK), and Slovenian (SL). Some terms also appear in Latin (LA).

The table below shows what information on a term (apart from a term itself in various languages) *Evronim* offers[iii]:

Creation Date	date that the term was entered in the database
Change Date	date of the last update
Entry Number	ID number of term in the termbase
Project	project within which the terms were translated or checked
Source Language	source language of the term
Subject EN	subject field (in English)
Subject SR	subject field (in Serbian)
Term Reference	source document of the term
Target Term Reference	reference in the target language
Relevant Institution	competent authority that provided, recommended or reviewed the translation
Reliability	reliability of translation on the scale from 1 to 4: 1 – draft term, yet to be confirmed (as a rule for new concepts) 2 – term in use, but not yet confirmed by expert or legal reviewer 3 – terms reviewed and recommended by expert and/or legal reviewer 4 – verified term (from the primary legislation)
Definition	term definition in the specific field
Definition Reference	definition source
Context	term context from the document on the source or target language

Context reference	context source
Note	additional information, notes and comments
Note Reference	note source
Linguistic Notes	notes, comments and additional information concerning grammar, language usage etc.
Hyponym	narrower term
Hyperonym	wider term
Related concepts	concepts connected to given concepts through different types of associative and lexical relations

Table 1. *Evronim* term information

On May 7, 2017 the termbase contained 18919 entries and it was last updated on May 5, 2017 (what suggests that it is constantly developing).

When it is about *Evroterm* use, we can say that it is very user-friendly. Entering a term (that can contain one or more words) into the entry field you get a list of hits that contains your query. Clicking on one of hits you can get all the information related to the hit. It is possible to run the search in any of the languages from the database, but Serbian and English are mostly used. *Evronim* is linked to *Evroteka* (a bilingual corpus that will be presented in the next chapter) and if a term appears underlined, clicking it you get a list of segments containing it (*Evroteka's* result). As we have mentioned, it is very user-friendly – various supplementary symbols can be used for a more advanced search, which is well guided by the instructions for use.[iv]

Figure 1. *Evronim's* result for the query *chapter*

In Figure 1, there is an example of *Evronim* search result obtained for the query *chapter* (English language search).

3. *Evroteka* – a parallel bilingual corpus

Another important digital translation tool, developed also for the needs of EU integration process, is *Evroteka* – a bilingual (English-Serbian) parallel corpus. It presents a collection of different legislation texts and their segments, formed as a result of the acquis translation into Serbian. Trados software is used for this translation and corpus creation (its Translator's Workbench is used fot translation memories creation).

The same as *Evronim*, *Evroteka* is also ongoing and it is being constantly develped. On May 7 it contained 140369 words (5997 translation units) in which there were 4003 terms from *Evronim*. It was last updated on March 23, 2017. The corpus describes every contained segment with a specific degree of reliability (defined on the scale of 1 to 5), depending on the phase when it was inserted in it. The phase refers to EU legislation translation process phases, which are: translation, expert, legal and language review and verification.

Evroteka search is based on the word level and it can be search on both, Serbian and English language, providing always a bilingual output. As sarch results, beside the query text segment, *Evroteka* offers its field, phase- degree of reliability, ID number of the document, its title or the abbreviation. Sometimes it also provides a link that takes you to the entire text of the legal act in question. If the search words is contained in *Evronim*, you will also get a link to it. If a user needs an advaced search, he/she can use supplementary symbols[v] (similar to *Evronim* search).

The following print screen picture shows the results for the same query we have previously used in the *Evronim* presentation (word *chapter* entered in English). Of course, since *Evroteka* search results are texts it is impossible to capture the whole output in one print screen picture, but to present it better, here is provided only the first part of the result.
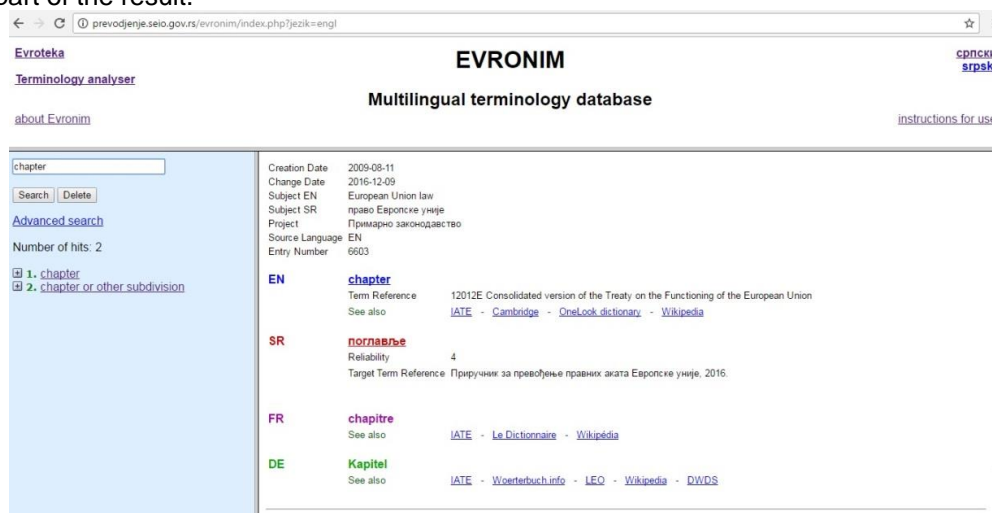

Figure 1. *Evroteka's* (partial) result for the query *chapter*

4. *Evronim's* and *Evroteka's* precursors

Neither *Evronim* nor *Evroteka* are authentically Serbian termbase/parallel bilingual corpus. Actually, all the republics of former Yougoslavia have their versions of these digital resources. Everything started in 2000 with *Evroterm*[vi], a termbase established in the preparation of the Slovene version of the acquis. Its parallel corpus is called *Evrokorpus*[vii] and it was started in 2002. Evroterm's technology was later used in the rest of Yougoslav countries and Albania:

(1) Serbia: Evronim – term base of the European Integration Office of the Republic of Serbia (since 2009);
(2) Bosnia and Herzegovina: BiHterm – termbase of the Directorate for European Integration (since 2011);
(3) Montenegro: Monterm – termbase of the Montenegrin Ministry of Foreign Affairs (since 2009);
(4) Macedonia: termbase developed by the Sector for Preparation of the National Version of the acquis (since 2007);
(5) Croatia: Euroterm – termbase developed by the Ministry of Foreign Affairs (publicly available from 2007 to 2009).

Basically, one digital pattern was used to create all the listed tools and resources which can be also considered as good examples of cooperation in the EU integration process.

5. Creators and users of *Evronim* and *Evroteka* – an insight from the perspective of language policy and planning

As it has just been mentioned, *Evronim* and *Evroteka* were established following the pattern and technology of *Evroterm* and *Evrokorpus*. It is obvious that they are based on the same software (they even name the same person as a contact for software issues), but of course, they have different teams of terminologist in charge of base development.

The European Integration Office of the Government of Serbia is a body responsible for these tools. Since the tools are closely related to the EU integration process and translation of the aquis, as the negotiations progress, they change, adapt and improve.

From the point of view of language policy and planning (more precisely, terminology policy and planning), it would be interesting to reflect on the fact that a governamental body is responsible for these terminological (in other words, mostly linguistic) tasks.

Language policy and planning (LPP) became a subject of scientific research in the second half of the 20th century (but its activities, of course, were present before) and today it draws a lot of researchers' intention. Summarizing the definitions given by modern researchers (Filipović, 2009; Hornberg, 2006; Ricento, 2006, etc.), we could say that language policy and planning is an official (governmental, belongs to public institutions), organized, long term activity on various aspects of language status, linguistic forms and domains of language use. It includes status LPP, corpus LPP and acquisition planning.

Terminology policy and planning represents a segment of language policy and planning. It is of great importance especially for languages for specific purposes (LSP)

or special languages/specialized languages, how they are also called. As Cabré (1999: 59) notes, language for general purposes (LGP) is used in *unmarked* situations.

"In contrast, we speak of special or specialized languages to refer to a set of subcodes (that partially overlap with the subcodes of the general language), each of which can be "specifically" characterized by certain particulars such as subject field, type of interlocutors situation, speakers' intentions, the context in which a communicative exchange occurs, the type of exchange, etc. Situations in which special languages are used can be considered as "marked"", (Cabré, 1999: 59).

Terminology, as a key part of specialized language, provides specialized knowledge at the level of concepts (basic units of subject-field), and terms are recognized as main *content carriers* of expert knowledge.

It is clear that on one hand, terminology plays an important part in specialized subject fields and on the other, it is a concern of those responsible for terminology policy and planning. In our particular case, we deal with EU legal terminology in modern Serbian language.

From recent research done in the field of terminology in Serbian (Jakić, 2014; Đordan, 2016) we see that usually we lack of coordinated, organized and institutionalized work. In most cases, experts in specific fields are creators and users of terminologies of their respective subject-fields. Since the official terminology policy (driven by language planners) offers no timely, adequate and useful solutions, professionals in need of terms as linguistic pictures of their concepts, need to produce those linguistic content carriers. From theoretical point of view, it is actually about two approaches to LPP: *top-down* and *bottom-up* approaches.

Top-down approach is the one shaped by government and official institutional framework which generally includes language professionals, but usually forgets to consult those form the specific subject fields. On the other hand, *bottom-up* approach to LPP represents the activities carried out by the users, or in the case of terminology, by special field experts. This implies that creators of terminology in use are actually subject field experts themselves. This put us into a situation where we have two different approaches to LPP that do not recognize each other. Instead of cooperation, we get two different professional groups dealing with the same issue, without recognizing each other and their results.

In our case *Evronim* and *Evroteka* could be represented not as a perfect, but as a good example of meeting point of the two different LPP approaches. They are created and developed by professionals who need to transpose EU legislation into Serbian. Obviously, the task is linguistic/translational, legal but at the same time it belongs to the area of a specific field that is a subject of a particular legal document. Taking all the mentioned fact into account, we could conclude that this type of a translation process should include:

(1) linguists with an excellent knowledge of both Serbian and English languages, experienced translators and terminologist;
(2) experts from a particular field which is a subject of legislation;
(3) legal specialists;
(4) IT experts (if it is about digital resources, such as *Evronim* and *Evroteka*).

Only an organized, coordinated work realized by a team formed of the mentioned experts can provide high quality terminology. Building such a system of terminology in Serbian language is of great importance for providing:

(1) quality documentation (in a particular case high quality national legislation, which directly influences EU integrations and country's future);

(2) development of different language domains, language vitality and domain and language maintenance.

If we add digitalization as an important characteristic of almost every profession in the 21st century, we are making terminology products and services modern, easy accessible, available to a larger community, user-friendly and suitable for further development.

Of course, both *Evronim* and *Evroteka* have a lot of space for further improvement of its different aspects.

6. Conclusion

In this paper, we have presented two important translation and terminology digital tools for Serbian language – *Evronim*, a termbase and *Evroteka*, a parallel bilingual corpus. Both have been created in the translation process of EU legislation (*acquis communautaire*) into Serbian language. They are products of translation activities, but they have also become a strong support to translators and reliable resource to a larger group of users.

Apart from being modern and useful digital tools, *Evronim* and *Evroteka* are a proof that a way to a stable, reliable, linguistically acceptable and professionally useful terminology leads through reconciliation and meeting of the *bottom-up* and *top-down* language (more precisely terminology) policy and planning. Creating a terminology system of any specific subject field necessarily requires both, linguistic and professional (subject field) knowledge. As we have already pointed out, if the products of this cooperation are digital tools and resources, we are on a good way to develop a useful and accessible terminology which can be constantly and easily improved.

Unfortunately, our (Serbian) terminological practice usually does not show the existence of cooperation between *bottom-up* and *top-down* language policy and planning approaches. It is more often that these two approaches act completely separated (each through its institutions and representatives), ignoring the results of the other side.

If we consider a wider picture of terminological work, we find out that terminology management should be entrusted to an institution dedicated especially to terminological work and which would include experts from all relevant fields (as Jakić, 2014 suggests and Đordan, 2016 confirms).

Anyhow, while we are waiting for some bigger actions to be realized in the framework of Serbian terminology, we can say that *Evronim* and *Evroteka* can be considered as a good example of cooperation between language and legal professionals (representatives of a specific subject field) and as a step towards a brighter terminology future.

References

1. Cabré, M. T. 1999. *Terminology: Theory, Methods and Applications*. Amsterdam: John Benjamins Publishing Co.

2. Đordan. A. 2016. *Jezička politika i terminologija u profesionalnim domenima jezičke upotrebe u standardnom španskom i srpskom jeziku*. Belgrade: Faculty of Philology (doctoral dissertation, unpublished manuscript form).
3. Filipović, J. 2009. *Moć reči: Ogledi iz kritičke sociolingvistike*. Beograd: Zadužbina Andrejević.
4. Hornberger, N. H. 2006. "Frameworks and Models in Language Policy and Planning". In Ricento, T. (ed.), *An Introduction to Language Policy: Theory and Method*. Malden, MA, USA: Blackwell Publishing, pp. 24-42.
5. Jakić, G. 2014. *Terminologija organizacionih nauka*. Beograd: Fakultet organizacionih nauka.
6. Ricento, T. (ed.). 2006. *An Introduction to Language Policy: Theory and Method*. Malden, MA, USA: Blackwell Publishing.

Webography

1. http://eur-lex.europa.eu/summary/glossary/glossary.html?locale=en [accessed March 2017].
2. http://eur-lex.europa.eu/summary/glossary/acquis.html [accessed March 2017].
3. http://prevodjenje.seio.gov.rs/evronim/index.php?jezik=engl [accessed May 2017].
4. http://prevodjenje.seio.gov.rs/evroteka/index.php?jezik=engl [accessed May 2017].
5. http://www.evroterm.gov.si/index.php?jezik=angl [accessed April 2017].
6. http://www.evroterm.gov.si/evrokorpus/index.php?jezik=angl [accessed April 2017].

---

[i] http://eur-lex.europa.eu/summary/glossary/glossary.html?locale=en

[ii] http://eur-lex.europa.eu/summary/glossary/acquis.html

[iii] http://prevodjenje.seio.gov.rs/evronim/index.php?jezik=engl

[iv] More details available at: http://prevodjenje.seio.gov.rs/evronim/index.php?jezik=engl

[v] More details available at: http://prevodjenje.seio.gov.rs/evroteka/index.php?jezik=engl

[vi] http://www.evroterm.gov.si/index.php?jezik=angl

[vii] http://www.evroterm.gov.si/evrokorpus/index.php?jezik=angl