The Corpus of the Croatian Church Slavonic Texts and the Current State of Affairs Concerning the Dictionary of the Croatian Redaction of Church Slavonic Compiling

Vida Vukoja
Old Church Slavonic Institute, Zagreb
vidal@stin.hr

Abstract

Croatian Church Slavonic is a literary, bookish idiom used in Croatia from the XI/XII until the XVII c. based on Old Church Slavonic, an idiom created by Sts. Cyril and Methodius, and shaped by the Croatian vernacular.

The Croatian Church Slavonic corpus consists of the texts excerpted from 11 breviaries, 4 missals, 3 psalters, 3 rituals, 15 miscellanies. It also incorporates all the 26 fragments dated from the period up to and including the XIII c. and several auxiliary sources. The corpus was created over a period of thirty years (from 1959 to the early 1990s) at the Old Church Slavonic Institute in Zagreb. It is a historical, referential, representative paper card-file of excerpts. It is also a parallel corpus, as it contains Latin and Greek parallel texts, those that were identified as closest to the actual source texts for the translational Croatian Church Slavonic texts.

The *Dictionary of the Croatian Redaction of Church Slavonic* has been compiled on the basis of the Croatian Church Slavonic corpus. The fascicles of the *Dictionary* have been published since 1991. So far, 1 (1991)–19 (2012), with the dictionary articles A–ŽRЬTVA (according to the Old Cyrillic alphabet) have been printed.

Keywords: corpus of the Croatian Church Slavonic; Dictionary of the Croatian Redaction of Church Slavonic; (Paleo)slavistic lexicography

1 Basic Information on the Croatian Church Slavonic Language (acr. CCS)

The Cyrillo-Methodian landmark mission, with its far-reaching influence on Medieval European culture and history, felt on a broad scale even today, commenced in 863, when the Thessalonian brothers arrived in Great Morava. Cyril and Methodius created a new language, now known as Old Church Slavonic (acr. OCS), for the purpose of translating Greek biblical and liturgical texts, literary texts as well as those concerning administration and law. That language did not match any particular Slavic vernacular, as it was constructed to potentially serve all the Slavic peoples ready to be evangelized in the

Slavic language (i.e. OCS), and willing to embrace literacy in a newly-invented script for its notification – Glagolitza or the Glagolitic script, created probably by Cyril.

After the end of the mission, a certain number of Cyril and Methodius' disciples arrived in the territory populated by the Croats. Under the influence of the Croatian vernacular, a new Church Slavonic language system came to be. That literary, bookish idiom used in Croatia from the XI/XII until the XVII c. is known as the Croatian Church Slavonic. It had a privileged status throughout the Croatian Middle Ages within the Croatian diasystem,[1] characterized by the Croatian/CCS diglossia[2], as it was a liturgical language, whose usage regularly marked a high literary style.

Two things should be mentioned in order to signal the importance of CCS. The first one considers European culture and history, and the even broader context of the Catholic Church history. CCS was the only close-vernacular idiom which gained and retained the explicit permission of the Pope to be used for liturgical purposes (besides Latin, Greek and Hebrew). Therefore, ahead of the decision of the Second Vatican Council allowing Catholic liturgy in vernacular, a CCS mass was served in the Roman St. Peter's basilica. The second thing to mention considers Croatian literacy. Namely, CCS is the first Croatian literary language, used from the end of XI c. until 1561. Its significance is reflected in the fact that it is the language of the *Baška tablet* (Cro. Bašćanska ploča, dated in 1100), one of the first and most important Croatian written monuments, but also the language of the first Croatian incunabula, *Missale Romanum Glagolitice,* which is the first missal in Europe not published in the Latin script and Latin language. Six out of nine Croatian incunabula were printed in CCS. The preserved Croatian texts in the Cyrillic script date from later periods, as is the case with the texts in Latin script.

2   The Corpus of the Croatian Church Slavonic texts (abbr. the CCS Corpus)

2.1   Basic Information on the Corpus of the Croatian Church Slavonic texts (abbr. the CCS Corpus)

Here, the term "corpus" is understood as a language material, a cluster of texts, purposefully collected to testify choices and combinations of choices made by users of a particular language (Sinclair 2003:167; cf. Svensén 2009: 43). The CCS corpus is primarily prepared for the compiling of the *Dictionary of the Croatian Redaction of Church Slavonic* (acr. DCRCS), but due to its features, it serves as a prime source for various linguistic investigations and other types of research.

---

1   For the meaning of the term diasystem v. Weinreich (1954), cf. Brozović (1970 [1967]): 14. For the CCS and Croatian vernacular and literary idioms as constituents of one common diasystem v. Katičić (1992).
2   For the character of the Croatian/CCS diglossia v. Mihaljević (2010).

The history of the CCS corpus started with the Fourth International Slavistic Congress, which took place in Moscow in 1958[3], where the suggestion to compile a thesaurus of the Church Slavonic language was embraced by the leading (paleo)slavists. The planned thesaurus was supposed to incorporate all the national Church Slavonic versions (or redactions): Bulgarian, Macedonian, Czech, Russian, Romanian, Bosnian, Serbian. It is important to mention that during that very time, the first fascicles of the landmark *Slovník jazyka staroslověnskeho* (abbr. *Slovník*; v. Štefanić 1962; Nazor 1991: I), based on the corpus of the canonical OCS texts, started to appear. The form and principles applied to the corpus prepared for the *Slovník* compiling, and the *Slovník* compiling itself, decisively influenced the CCS corpus and the DCRCS compiling respectively. The decision of the Croatian lexicographers to follow their Czech colleagues was not due to any lack of ingenuity, but was incited by the long-sighted priorities to shape the CCS corpus as similarly as possible to the corpus for *Slovník* compiling and to shape the DCRCS as similarly as possible to the *Slovník*. The aim of those decisions was to enhance the possibilities for comparative research ultimately aimed at attaining the structural knowledge of: (a) particular national versions of Church Slavonic, (b) the (Old) Church Slavonic language system, taken apart from any (national) vernacular's influence.

For that reason, the arrival of the reputable and experienced Czech paleoslavistic lexicographer, František Václav Mareš, one of the principle collaborators at the compiling of *Slovník*, was very much welcomed, as was his cooperation with the then-young Croatian paleoslavists, in the task of his laying-out of the main principles for the CCS corpus and the compiling of the DCRCS drafted in Mareš (2007[1962]), as well as setting up the long-lasting work of the DCRCS compiling. It took about thirty years to complete the CCS corpus (from 1959 until the beginning of the 1990s). Despite its apparent excessive duration, it was actually the expected length for such a demanding task.

2.2 The Constituents, Card-files and Features of the CCS Corpus

The CCS corpus consists of selected CCS sources, manuscripts and incunabula dating from (XI/)XII to mid-XVI c., with the priority given to earlier and integral versions of particular texts. Its constituents are as follows: 11 breviaries, 4 missals, 3 psalters (1 with Psalter commentary), 3 rituals, selected texts from 15 Croatian Glagolitic miscellanies, all the fragments dating from the period up to and including the XIII c. (altogether 26 pieces), auxiliary sources are excerpted in the cases of lexicographically interesting lemma occurrences: another 2 missals and 2 breviaries.[4] A vast range of text genres found their place within the CCS corpus: liturgical texts (including biblical passages), biblical and apocryphal texts, sermons and homilies, moral and didactical texts, legal texts, legends and visions, hagiographies, disputations and other literary texts.

---

3   For the information on the prehistory of the CCS corpus v. Nazor (2008).
4   For the exhaustive list of the sources, constituents of the CCS corpus v. Nazor (1991).

The exact number of the CCS corpus tokens is not known, but is should be somewhere between 1 400 000 and 2 100 000. The excerpts are of various lengths and multiplied so that they can be organized within two card-files (an example of excerpt cards can be seen in the Figure 1).
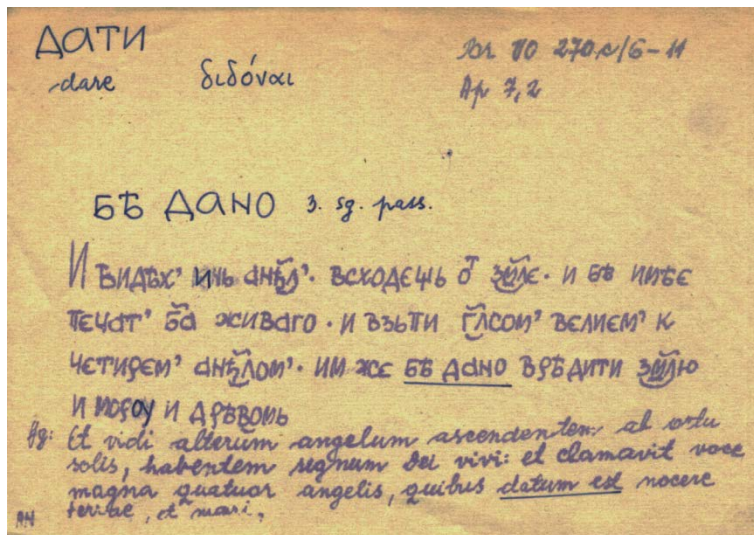

Figure 1: An excerpt card (used in the sources card-file and in the azbuka card-file).

Along the excerpt cards run the so-called parallel cards with the variations of the lexical constituents of the excerpt as found in other sources (an example of a parallel card is given in Figure 2.).
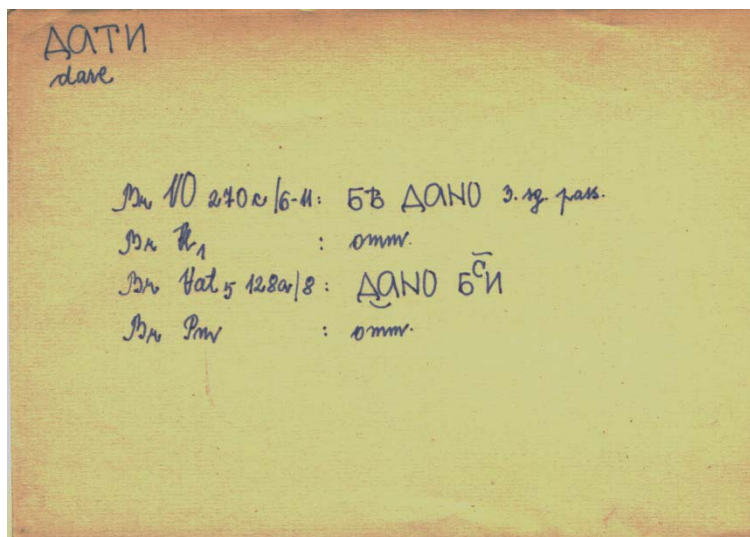

Figure 2: A parallel card (used in the sources card-file and in the azbuka card-file).

The first card-file is established according to the sources (informally, the sources card-file) and it contains approximately 420 000 cards. The second card-file is established according to the azbuka sequence of the lemmas (informally, the azbuka card-file), with more than 400 000 cards. The azbuka

card-file contains fewer cards than the sources card-file because not every token is taken into consideration for the compiling of the entry of the DCRCS, and consequently, its card doesn't appear in the latter card-file.

There are three auxiliary card-files. The first one to be mentioned is the card-file of the CCS lemmas (systematized according to the Old Cyrillic azbuka) with approximately 18 100 cards (of which approximately 8500 nouns, 5400 verbs, 2500 adjectives; an example of a card is shown in the Figure 3).
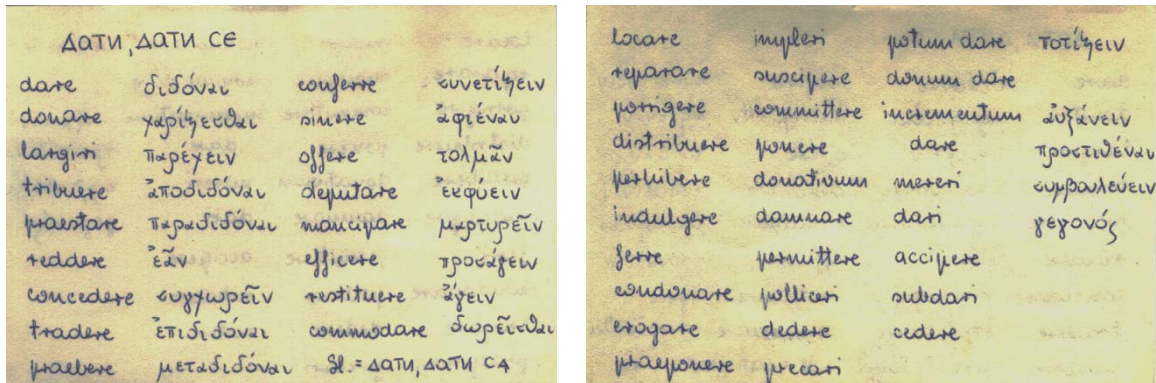

Figure 3: A card from the card-file of CCS lemmas (*recto* and *verso*).

The second one is the card-file of the Greek parallels of the CCS lemmas (systematized according to the Greek alphabet) with approximately 60 000 cards (two examples are given in the Figure 4).
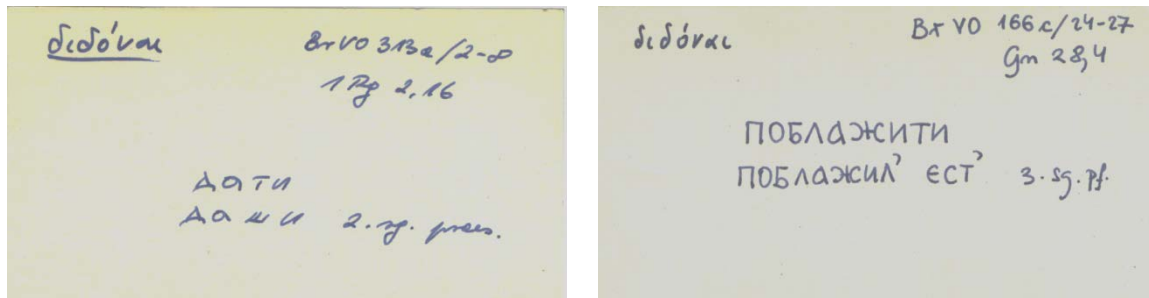

Figure 4: Two cards from the card-file of Greek parallels.

Also, there is the card-file of the Latin parallels of the CCS lemmas (systematized according to the Latin alphabet) with approximately 200 000 cards (two examples can be seen in the Figure 5).
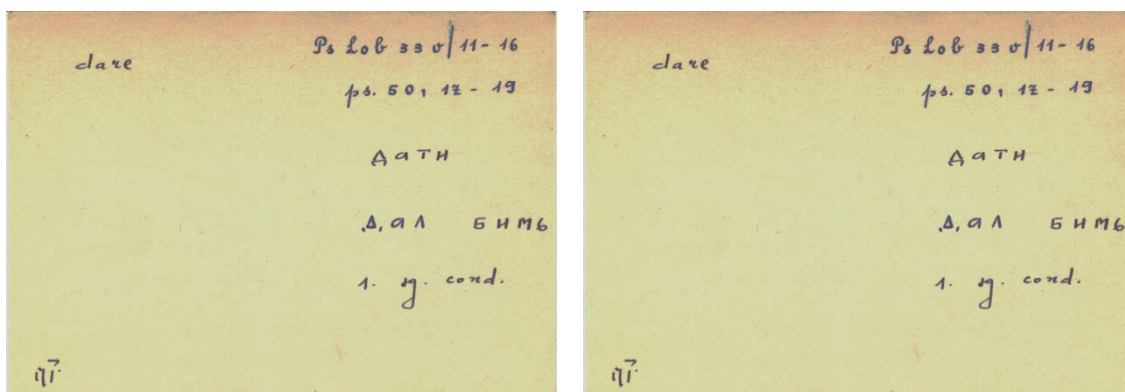
Figure 5: Two cards from the card-file of Latin parallels.

The main features of the CCS corpus are shaped by the principles articulated by Mareš (2007[1962]), and they can be listed as follows:

(1) The CCS corpus is *historical*: it contains texts originating from (XI/)XII to XVI c.

(2) The corpus is made with the aspiration to enable the production of a *thesaurus*, i.e. a dictionary containing the entire vocabulary range confirmed in the CCS texts.

(3) It is *referential*, general; i.e. it contains virtually all of the confirmed CCS lexemes in a number of instances and types of context which represent the situation in the entire (preserved) corpus of the CCS texts. Also, it contains various versions of the same lemma in terms of its graphic features (such as manners of noting /ь/, /ъ/, /m/, Latin or Cyrillic script initials in otherwise Glagolitic texts etc.) as well as its phonological, morphological, lexical, textological features. Therefore, the corpus meets the requirements for various types of research, be it linguistic or non-linguistic, such as textological, literary, liturgical, historical etc.

(4) The corpus is *representative*. It contains texts of different genres, mirroring in quantity the real share of particular genres in the entirety of CCS texts.

(5) In form, the corpus is *paper card-file* of excerpts, where excerpts are used for practical reasons, as expected considering the period it was set up. Still, it cannot be perceived as a mere collection of citations, as the sequences of the card-files contain integral texts, and even entire hundreds-of-pages long manuscripts, just as it is expected in the case of a proper corpus (and unlike expected for a collection of citation).

(6) The corpus *constituents* are *not strictly divided*. Despite being fundamentally systematized on the basis of the source type (missals, breviaries, psalters, rituals, fragments, texts selected from miscellanies), the corpus is not subdivided into compartments. But, if important versions of a particular text appear in different groups of sources, the differences are noted on a string of parallel cards.

(7) It has characteristics of *static* and *dynamic* corpora. A static corpus (also, *sample corpus*) is a structured collection of a limited number of examples of the lemmatized lexemes, which, once introduced in the collection, cannot be excluded from the corpus. And the CCS corpus is such a collec-

tion in its azbuka card-file. A dynamic corpus (also, *monitor corpus*) consists of entire texts and it is open to receiving a newly-found text or confirmation of lemmas shown to be desirable concerning the referential and representative features of the corpus, and the CCS corpus is open to the incorporation of such newly-found texts.

(8) The corpus contains *only written texts*, with a great majority of texts being manuscripts and a limited number of incunabula. As it concerns medieval language material, no spoken texts are expected.

(9) Besides the above reason, the CCS corpus can contain written texts only, as CCS itself is a *bookish* idiom (not spoken), *non-organic*, and *not* a *native language* of anyone.

(10) Texts for the corpus are *excerpted directly from the original sources or photocopies of the original sources* (not from any secondary editions of the selected texts).

(11) The corpus is *parallel*, not monolingual (on parallel corpora v. Svensén 2009:55-57; also Borin 2002). Whenever an excerpted CCS text is actually a translation from Greek or Latin (and most often this indeed is the case), if the parallel Greek or Latin text has been determined, the cards contain that parallel text below the CCS texts, as strictly aligned as possible.

(12) The corpus is formed in such a manner that the *basic version of the text is determined and differentiated* from the secondary versions of the same text found in various sources.

(13) The *critical, phototypical and facsimile editions* of the texts included in the corpus are *published*, the most ambitious undertakings being the capital editions of Hrvoje's Missal (Grabar et al. 1973) and the Second Novi Breviary (Pantelić & Nazor 1977).

(14) Approximately *18,100 lemmas* are *expected* in the dictionary once it is finished.

(15) Without exception, the texts incorporated in the CCS corpus are written in the *Glagolitic script*, but on the card-files of the CCS corpus they appear transliterated in the Old Cyrillic script, according to the decision made at the Fourth International Slavistic Congress. The principles of the transliteration from Glagolitic into Old Cyrillic script are appropriated from Jagić (the table of Glagolitic and Old Cyrillic correspondents can be found in Jagić 1879: XXXVII; cf. Bratulić 1981: 145-146). The lemmas of the DCRCS are also written in the Old Cyrillic script. Of course, Greek parallel texts are written in the Greek alphabet and Latin parallel texts are written in the Latin script.

(16) The *tokens* of the CCS corpus, appearing in the sources card-file, are *grammatically tagged* (parsed) and inserted into the azbuka card-file in all the cases except those of extremely frequent types, such as the forms of the verb *biti* ('to be') or the noun *bogъ* ('god').

(17) All the types are *lemmatized* in a *normalized form*. The principles of normalization are given in Grabar et al. 1991: VIII-XIII, XIX-XXV.

2.3 The European and Croatian Context of the CCS Corpus and the Compiling *of the DCRCS*

2.3.1 The Situation in the (European) Paleoslavistic Lexicography

Currently, after the completion of the landmark four-volumes OCS dictionary (i.e. *Slovník*), the Czech paleoslavistic lexicography has commenced five major lexicographic works based on several paleo-slavistic (meta-)corpora: *Etymologický, Slovník* V., revision of *Старославянский*[5], *Srovnavaci index k slovnikům zpracovavanym v ramci Komisie pro cirkevněslovanske slovniky*[6], *Řeckostaroslověnsky.* Among these, special attention should be drawn to *Srovnavaci,* as the CCS corpus is a component of its meta-corpus, and the DCRCS' lemmas are taken into consideration during the process of determining similarities and differences of various Church Slavonic idioms.

After publishing major Old and Middle Bulgarian[7] dictionaries (*Старославянский; Бончев* 2002-2012), Bulgarian paleoslavists took up the compiling of a voluminous digitized corpus *Компютърни и интерактивни средства за исторически езиковедски изследвания*[8], which is to become a permanent basis for the comprehensive diachronic description of Bulgarian (including its Church Slavonic constituent) from X to XVIII c. Also, Bulgarian paleoslavistic production abounds with dictionaries and indices of particular texts and sources (e.g. *Тасева* 2010, esp. pp. 533-818; *Димитров* 2010, 2013; *Илиева 2013a, 2013b).* Macedonian paleoslavists are working on the dictionary of the Macedonian Church Slavonic, based on the corpus comprising documents from the XII to XVI c. *(Речник).* Once the corpus of the Serbian Church Slavonic texts have been created (and it is currently in preparation), the Serbian Church Slavonic dictionary compiling can be expected (*Српскословенски* serves as its introductory fascicle).

Important state-financed projects were recently run or are being run at present in non-Slavic countries, among which two will be mentioned: *SlaVaComp. COMPutergestützte Untersuchung von VAriabilität im KirchenSLAvischen* in Germany (Freiburg, 2013-)[9] and *Die kirchenslavische Übersetzung der Werke von Gregorios Palamas und Barlaam von Kalabrien* in Switzerland (Bern, 2010-2013)[10].

2.3.2 The Situation in the Croatian Historical Lexicography

Currently, in Croatia, besides the DCRCS, there are three historical lexicographical projects in different phases of progress: *Dictionary of Croatian Kajkavian literary language* (Cro*. Rječnik hrvatskoga kajkavs-*

5   Emilie Bláhová is the redactor of the revisited and supplemented edition of *C*тарославянский.
6   The first volume should be published in forseeable future.
7   Bulgarian slavists use the term "Middle Bulgarian" for the idiom other slavists usually name "Bulgarian Church Slavonic", and "Old Bulgarian" for the idiom the other slavists usually name "Old Church Slavon-ic".
8   The corpus has been created at the Софийски университет „Св. Климент Охридски" under redaction of Dora Ivanova-Mirčeva. For more information on that corpus v. Totomanova (2012).
9   The project is led by J. Besters-Dilger and G. Schneider (University of Freiburg). More information can be found at the web-site: http://www.slavacomp.uni-freiburg.de/ [15/10/2013].
10   The project was led by Y. Kakridis (University of Bern) and financed by *Schweizerischer Nationalfond*.

koga književnog jezika), *Dictionary of the Croatian literary language from the National Revival to I. G. Kovačić* (Cro. *Rječnik hrvatskoga književnoga jezika od preporoda do I. G. Kovačića*; compiling of which was resumed in 2008, after 18 years of hiatus), *Old Croatian dictionary* (Cro. *Rječnik starohrvatskoga jezika*, in the middle of the elaborate preparation of its corpus).

Here, the last of the three lexicographical projects is the most important one, because the corpus which is being prepared for the *Old Croatian dictionary* compiling shares two very important features with the CCS corpus: (a) both are referential which also makes the dictionaries based on them referential; (b) both corpora consist of integral texts found in manuscripts and incunabula originated from the same period, namely (XI/)XII—XV/XVI c. But, the fact that the two corpora and the two dictionaries are compatible is even more important than their aforementioned overlapping features. Indeed, the two corpora and the two dictionaries taken together cover two major components of the medieval Croatian diasystem: CCS (in the case of the DCRCS) and literary medieval Croatian idioms (in the case of the *Old Croatian dictionary*).[11] Thus, to once have both corpora and both dictionaries at disposal is critical for the trustworthy diachronic investigation of the Croatian language. It. cannot be emphasized enough that it is virtually impossible to conclude the investigation of the medieval Croatian diasystem without both of the corpora and respective dictionaries published.

The research of the CCS component of the Croatian diasystem appears to be more demanding for a contemporary linguist then the research of the literary medieval Croatian idioms due to the fact that the CCS corpus of texts is much less familiar, even obscure, to the majority of linguists dealing with the diachrony of the Croatian language, which is not much of a surprise if the following is kept in mind:

(1) it is written in an unusual script (i.e. Glagolitic);

(2) it is written in an idiom (i.e. CCS) not entirely comprehensible if relying solely on the good command of the Croatian vernacular diasystem;

(3) in numerous cases, the linguistic features of a given CCS text cannot be interpreted correctly if its Greek or Latin parallel text is unfamiliar to the investigator;

(4) CCS documents are scattered not only across Croatia, but also across the world (including countries such as Russia, Austria, Slovenia, United Kingdom, USA and so forth);

(5) in many cases, the CCS sources are at least partially damaged due to age and mal-manipulation, sometimes even to the point of being barely readable.

It can be added that the above-mentioned features of the CCS corpus give a clue to a range of competences CCS lexicographers have to have at their disposal, as well as to the indispensability of their contribution to the research of the overall diachrony of the Croatian language.

---

11   For the comparison of the two corpora v. Vukoja (2012).

3 The Contribution of the DCRCS Compiling to Croatian Studies and to the Historical Lexicography in General

3.1 The Contribution of the DCRCS Compiling to Croatian Studies

Being referential and representative, besides its other features, the CCS corpus provides the most effective basis available for the relevant research of the CCS idiom. If conducted in a methodologically correct manner, the conclusions drawn from the analyses of the corpus can be taken as trustworthy for the whole of the CCS idiom.

In general, every research of the CCS has to take into consideration three layers of factors: (i) the inherited Old Slavonic state, with its characteristic Greek and Latin influences; (ii) the Latin influences that CCS acquired through the adaptation to the Western (Church) tradition, which should be taken separately from the Romance influences acquired through spoken language; (iii) the influences of Croatian vernacular idioms. At present, systematic research of CCS is conducted on the basis of the CCS corpus in several fields: grammar (phonology, morphology, syntax, semantics), lexicology, textology, translation theory and practice, research on the medieval conceptualization of feelings.

With the DCRCS, Croatian lexicography would vastly improve the starting point for all the research on the Croatian diachronic (medieval) diasystem, as it implies diglossia consisting of the Croatian vernacular and CCS. The compiling of the DCRCS adds up critically to another enterprise of the Croatian lexicography, namely the creation of the corpus for the *Old Croatian dictionary* and its compiling. Without the DCRCS, the work of the lexicographers engaged in the project of *Old Croatian dictionary* would be considerably more difficult, even hardly accomplishable in numerous cases, as the CCS features in prevailingly vernacular texts would easily remain unnoticed, which, in turn, would result with a poor diachronical description of the Croatian language.[12] The DCRCS and its corpus offer help to all the scholars who for various reasons need to understand CCS texts. Due to the vast range of the CCS text genres, such scholars may be hagiographers, liturgists, general historians, historians of specific fields (feelings, medicine etc.) ethnologists, and so forth.

3.2 The Contribution of the DCRCS Compiling to the Historical Lexicography in General

In the context of the (international) historical lexicography, the corpus for the DCRCS and the DCRCS itself are one of the pivots of the international (paleo)slavistic lexicography. At the moment, the results achieved within the management of the CCS corpus and the compiling of the DCRCS are the constitutive elements of the work on *Srovnavaci*, but the CCS corpus and the DCRCS are also at the disposal of all the lexicographers working on or with various Church Slavonic corpora. They are both also

---

12 Cf. e.g. the difficulties in determining the origin of certain features of language forms used by Bartol Kašić, v. Vrtič (2009:117-118.218-219.291-293).

valuable help for all the linguists engaged with the national Church Slavonic idioms (Bulgarian, Macedonian, Czech, Russian, Romanian, Bosnian, Serbian). As national Church Slavonic idioms have been rightly recognized to be integral components of their respective national standard languages, the CCS corpus and the DCRCS support, within their capacities, the research of various Slavic languages. The presentation of the CCS material in both formats, corpus and dictionary, is most-welcome because CCS in its original text, written in Glagolitic and Old Cyrillic scripts usually proves to be a tough nut for slavists, especially those not versed in dealing with the Slavonic idioms.

4   The DCRCS Dictionary Article

The DCRCS is compiled in accordance with the general lexicographical theoretical knowledge (Zgusta 1971, Sinclair 2003; Svensén 2009), fashioned in accordance with the practical knowledge of the most experienced paleoslavistic lexicographers, those from the Czech paleoslavistic lexicographical tradition (v. Mareš 2007[1962]), but finally formed so that it recognizes and appropriately displays a specific range of the CCS features in semantics, grammatical forms, textological and translational traits (Nazor 1991; Grabar et al. 1991; Vukoja 2012).

The methodology of the DCRCS compiling is based on the compiling methodology of the earlier Old Slavonic dictionary, *Slovník*, in order to achieve the formal lexicographical concordance needed to enable, help and enhance various paleoslavic research.[13] However, the particularities of the CCS material has asked for extensive adaptations, e.g. in the areas of normalisation, differentiation of semantic variants etc. By all means, the DCRCS is made according to the highest standards of the relevant historical dictionaries.

The head of the dictionary article is written in the Glagolitic and Old Cyrillic alphabets, but its body is in the Latin script, except for the Greek parallels of the lemma and pertaining citations. The body of the dictionary article contains Croatian and English translations of the CCS lemma, Greek and Latin parallel lexemes, as well as an encyclopaedic identification (in the Latin language), if the given lemma is an anthroponym, a toponym or a technical term (most often liturgical). Every recognised meaning is accompanied by CCS examples followed by Greek and Latin parallel phrases as well as CCS variants of the lemma, if existing. In choosing the examples, not only semantic variations are sought to be presented, but also a range of different contexts (and text genres) as well as the range of the lemma forms. Special attention is given to the phrasemes, the differences in spelling (words abbreviated under a tilde, shortened by suspension or by omission of the first syllable with which the preceding word ends) and other particularities of the lemma. At the end of the dictionary article, relevant synonyms are enlisted. Also, possible presence of the given lemma in *Slovník*, Miklosich (1862-1865) and ARj is indicated.

---

13   For the basic methodological principles of the DCRCS compiling v. Mareš (2007[1962]); also Grabar et al. (1991: VIII-XXX).

5 The Current State of Affairs Regarding the CCS Corpus and the DCRCS Compiling

5.1 The Current State of the CCS Corpus Conversion into Digitized Form

At present, two main paper card-files (the sources and the azbuka card-files) are scanned and preserved in the JPEG format (digitization editor: Marica Čunčić, software: Antonio Magdić, scanned by the Croatian State Archives, ArchivePRO). This way, the safety of the data is largely improved, but very little is done in terms of the digital manageability of the corpus (digital readability and searchability, practicality of the DCRCS' compiling process). For internal use, one of the DCRCS compilers edited the JPEG-formatted sources card-file in a more user-friendly manner, and the DCRCS' compilers are using that version in their daily work. Still, paper card-files are indispensable in the majority of research cases.

At the moment, a combination of factors (among which the lack of sufficient financial support is the most notable one) contributed to the decision of putting on hold the project of digitization, which should end only when the full digital readibility and searchability of the CCS corpus is achieved. Another notable obstacle is that the cards in the card-files are written by more than two dozen different hands, which practically excludes all known Optical Character Recognition (acr. OCR) options, and which requires an exploration of the Intelligent Character Recognition (acr. ICR) possibilities. Still, the collaborators on the project of the compiling of the DCRCS, as well as the authorities of the Old Church Slavonic Institute are constantly looking for a solution that would make digitization viable.

5.2 The Current State of Affairs Concerning the DCRCS Compiling

The fascicles of the DCRCS, one per year, each containing 64 pages, have been published since 1991. So far, 1 (1991)–19 (2012), with the dictionary articles A–ŽRЬTVA (according to the Old Cyrillic alphabet) have been compiled. The first 10 fascicles are bound in Vol. 1. (DCRCS 2000), which has been peer-reviewed as an extraordinary lexicographical accomplishment on several occasions.[14] The fascicles 1–19 exist also in the PDF format, with a limited range of text-searching options. They are available on the Institute's intranet, and placed at the disposal of any interested researcher.

The fascicle 20 is only days away from publishing. Once it is printed, Vol. 2 of the DCRCS is to be bound. Despite their dedication, five lexicographers who are engaged in the compiling of the DCRCS are not able to produce the fascicles at a faster pace, but hopefully with additional collaborators the pace will be intensified in the foreseeable future.

---

14 E.g. at the presentation of the DCRCS Vol. I. at the International Slavistic Congress in Ljubljana 2003, also Гркович-Мејџор (2007: 187).

6   Conclusion

The CCS corpus is an indispensable tool for the research of the CCS idiom as well as the prime-quality source for all the scholars who for various reasons need to consult the CCS texts. Its present two formats (paper card-file and JPEG) seek for a thorough digital conversion of the corpus, which is on hold at the moment due primarily to the financial reasons.

The DCRCS with its forthcoming 20th fascicle is in progress, although at a moderate pace, due to the limited number of available lexicographers. Hopefully, the compiling will be intensified in forseeable future.

Despite the difficulties just mentioned, the work on the CCS corpus and the DCRCS compiling should be continued due to its major importance in the context of Croatian as well as (paleo)slavistic studies.

7   References

ARj = *Rječnik hrvatskoga ili srpskoga jezika*, Vol. I–XXIII. *Zagreb: JAZU.* (1880-1976).

Бончев 2002-2012 = архимандрит Анастасий Бончев. *Речник на църковнославянския език*. Т. 1. (А-О; 2002), Т. 2. (П-Я; 2012). София: Народна библиотека «Св. Св. Кирил и Методий».

Borin, L. (ed.). (2002). *Parallel corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April, 1999.* Amsterdam: Rodopi.

Bratulić, J. (1981). Ediciona praksa hrvatskih istraživača i izdavača srednjovjekovnih tekstova u XIX i XX stoljeću (Historijski prikaz). In Д. Богдановић (ур.) *Међународни научни скуп: Текстологија средњовековних јужнословенских књижевности, 14-16. новембра 1977.* Београд: САНУ, pp.137-147.

Brozović, D. (1970[1967]). *Standardni jezik.* Zagreb: Matica hrvatska.

DCRCS 2000 = *Rječnik crkvenoslavenskoga jezika hrvatske redakcije.* Vol. I. (*a-vrêdь*.). 2000. Zagreb: Staroslavenski zavod Hrvatskoga filološkog instituta.

Димитров, К. (2010). Речник - индекс на Словата на авва Доротей (По ръкопис 1054 от сбирката на М. П. Погодин). Велико Търново: Университетско издателство "Св. св. Кирил и Методий".

Димитров, К. (2013). Авва Доротей. Слова. Среднобългарски превод. Гръцко-български словоуказател. Велико Търново: Университетско издателство "Св. св. Кирил и Методий".

Etymologický = *Etymologický slovník jazyka staroslověnského.* E. Havlová, A. Erhart & I. Janyšková (red.). Praha, Brno: Akademie věd České Republiky, Academia, Tribun EU. (1989-).

Grabar, B., Mareš, F.V. & Mulc, I. (1991). Oblikovanje i sastav natuknice. In *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. (sveščić 1., Uvod).* Zagreb: Staroslavenski zavod Hrvatskoga filološkog instituta. pp. VIII-XVIII. (transl. in Eng. pp. XIX-XXX.)

Grabar, B., Nazor, A. & Pantelić, M. (1973). Missale Hervoiae ducis Spalatensis croatico-glagoliticum = Hrvatskoglagoljskimisal Hrvoja Vukčića Hrvatinića = Croato-glagolitic Missal of Hrvoje, Duke of Split: transcriptio et commentarium. V. Štefanić (red.). Zagreb, Ljubljana, Graz: Staroslavenski institut, Mladinska knjiga, Akademische Druck- u. Verlagsanstalt. + Faximil.

Грковић-Мејџор, J. (2007). Rječnik crkvenoslavenskoga jezika hrvatske redakcije, I. svezak (a-vrêdь), Staroslavenski institut, Zagreb, 2000. *Прилози за књижевност, језик, историју и фолклор,* LXXII/1-4, pp.185-187.

Jagić, V. (1879). Quattuor evangeliorum codex glagoliticus olim Zographensis nunc Petropolitanus. Characteribus cyrillicis transcriptum notis criticis prologomenis appendicibus auctum adiuvante summi ministerii Borussici liberalitate edidit V. Jagi. Berolini: Apud Weidmannos.

Илиева, Т. (2013a). Старобългарският превод на Стария завет, том 3: Старобългарско-гръцки словоуказател към Книгата на пророк Иезекиил. София: Кирило-Методиевският научен център, Българска академия на науките.

Илиева, Т. (2013b). *Терминологичната лексика в Йоан-Екзарховия превод на* "De Fide orthodoxa". София: Самиздател.

Katičić, R. (1992). 'Slovênski' i 'hrvatski' kao zamjenjivi nazivi jezika hrvatske književnosti. In *Novi jezikoslovni pogledi*. Zagreb: Školska knjiga, pp. 312-328.

Mareš, F.V. (2007[1962]). Návrh přípravných prací pro slovník jazyka církevněslovanského. *Církevněslovanská lexikografie 2006*. In Václav Čermák (sestavil); E. Bláhová, E. Šlaufová & V. Čermák (eds.) Praha: Slovanský ustav AV ČR, Euroslavica, 64-84. (Russian translation: Мареш, Ф.В. (1966). Проект подготовки словаря церковнославянского языка. *Вопросы языкознания XV*. Москва: Академия наук СССР, Институт языкознания, пп. 86-99.)

Mihaljević, M. (2010). Položaj crkvenoslavenskoga jezika u hrvatskoj srednjovjekovnoj kulturi. Свети Наум Охридски и словенската духовна, културна и писмена традиција (организиран по повод 1100-годишнината од смртта на св. Наум Охридски). Зборник на трудови од Меѓународниот научен собир. Охрид, 4-7 ноември. Скопје: Универзитет „Св. Кирил и Методиј", pp. 229-238.

Miklosich F. (1862-1865). *Lexicon palaeoslovenico-graeco-latinum*. Vienna: Guilelmus Braumueller.

Nazor, A. (1991). Uvod; Popis izvora; Navedena literatura. In *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. (sveščić 1., Uvod).* Zagreb: Staroslavenski zavod Hrvatskoga filološkog instituta. pp. I-III (transl. in Eng.: pp. IV-VII.). XXXI-XXXVI. XXXVII-XXXIX.

Nazor, Anica. (2008). Rječnik crkvenoslavenskoga jezika iniciran na IV. međunarodnom slavističkom kongresu u Moskvi 1958. godine (u povodu 50. obljetnice inicijative). In M. Samardžija (ed.) *Vidjeti Ohrid. Referati hrvatskih sudionika i sudionika za XIV. međunarodni slavistički kongres (Ohrid, 10.-16. rujna 2008.).* Zagreb: Hrvatsko filološko društvo, Hrvatska sveučilišna naklada, pp. 65-82.

Pantelić, M., Nazor, A. (1977). Uvod; Bibliografija. In *Drugi novljanski brevijar: hrvatskoglagoljski rukopis iz 1495*. Phototypical edition. Zagreb: Staroslavenski institut "Svetozar Ritig", Turistkomerc, pp. 7-37.

Řeckostaroslověnsky = Řeckostaroslověnsky index. Index verborum graeco-palaeoslovenicus. Praha: Slovansky ustav AV ČR, Euroslavica. (2008-).

Речник = Речник на црковнословенскиот јазик од македонска редакција. Том I. Вовед. А–Б. Скопје: Институт за македонски јазик. (2006).

Rosenwein, B.H. (2006). *Emotional Communities in the Early Middle Ages*. Ithaca: Cornell University Press.

Sinclair, J. (2003). Corpora for lexicography. In P. van Sterkenburg, (ed.) *A Practical Guide to Lexicography*. Amsterdam, Philadelphia: Benjamins Publishing Company, pp. 167-178.

Словарь = Словарь древнерусского языка (XI–XIV вв.). В 10 т. Москва: Институт русского языка Российской академии наук. (1988-).

Slovník = Slovník jazyka staroslověnskeho. Lexicon Linguae Palaeoslovenicae. I.-IV. Praha: ČSAV Slovansky ustav. (1958–1997).

Slovník V. = *Slovník jazyka staroslověnského.* Sv. V. (Addenda et Corrigenda). Praha: Euroslavica. (2010-).

Српскословенски = Српскословенски речник јеванђеља. Огледна свеска. Саставио: Виктор Савић. Уредник: Гордана Јовановић. Београд: Институт за српски језик САНУ. (2007).

Старославянский = Старославянский словарь (по рукописям X–XI веков). Под редакцией Р.М.Цейтлин, Р.Вечерки и Э.Благовой. Москва: Славянский институт академии наук Чешской республики, Институт славяноведения и балканистики Российской академии наук, "Русский язык". (1994).

Svensén, B. (2009). *A Handbook of Lexicography.* Cambridge: Cambridge University Press.

Štefanić, V. (1962). Problem rječnika južnoslavenskih redakcija staroslavenskog jezika. In *Slovo*, 11-12, pp. 181-187.

Тасева, Л. (2010). Триодните синаксари в средновековната славянска книжнина. Текстологично изследване. Издание на Закхеевия превод. Словоуказатели (Monumenta linguae slavicae dialecti veteris LIV). Freiburg im Briesgau: Weiher Verlag.

Totomanova, A-M. (2012). Digital Presentation of Bulgarian Lexical Heritage. Towards an Electronic Historical Dictionary. In *Studia Ceranea*, 2, pp. 219-229.

Vrtič, I. (2009). *Sintaksa Kašićeva prijevoda* Svetoga pisma. PhD. thesis. Filozofski fakultet Sveučilišta u Zagrebu, Zagreb, Croatia.

Vukoja, V. (2012). O korpusu Rječnika crkvenoslavenskoga jezika hrvatske redakcije i njegovu odnosu prema korpusima hrvatskoga jezika. In *Filologija*, 59, pp. 207-229.

Weinreich, U. (1954). Is a structural dialectology possible?. In *Word*, 10, pp. 388-400.

Zgusta, L. (1971). *Manual of Lexicography.* The Hague, Paris – Prague: Mouton – Academia.