

## Retro-Digitization of Croatian Pre-Standard Grammars

By Marijana Horvat\* & Martina Kramarić<sup>‡</sup>

*In this article, we will present the rich linguistic heritage of the Croatian language and our attempts to ensure its preservation and presentation to the general public by means of the "Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism – RETROGRAM" project. There is a long tradition of grammatical description in the history of the Croatian language. The first grammar book of the Croatian language was written at the beginning of the 17<sup>th</sup> century and the first grammar book written in Croatian was compiled in the middle of the 17<sup>th</sup> century. In later years, when literary and linguistic activity were transferred from the Dalmatian area to the northern and eastern part of Croatia, the Latin model for the description of the Croatian language was still present, even though German was also used. There were a large number of grammars written up to the second half of the 19<sup>th</sup> century, which are considered pre-standard Croatian grammars. They are the subject of research within the project "Pre-standard Croatian Grammars" at the Institute of Croatian Language and Linguistics. This research proposal "Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism" aims to create a model for the retro-digitization of the chosen eight Pre-standard Croatian Grammars (written from the 17<sup>th</sup> until the 19<sup>th</sup> century). The retro-digitization of Croatian grammar books implies the transfer of printed media to computer-readable and searchable text. It also includes a multilevel mark-up of transcribed or translated grammar text. The next step of the project is the creation of a Web Portal of Pre-standard Croatian Grammars, on which both the facsimiles and the digitized text of the grammars will be presented. Our aim is to present to the wider and international public the attainments of the Croatian language and linguistics as an important part of Croatian culture in general.*

**Keywords:** pre-standard Croatian grammars, history of the Croatian language, retro-digitization, Extensible mark-up language, Text encoding initiative, web portal of pre-standard Croatian grammars

### Introduction

Working with historical linguistic material always implies additional challenges for scholars. This is especially so if the works in question were not published or if they were written in manuscripts. Due to their fragility and the possibility of damage, it is only possible to work with them in specialized institutions. On the other hand, the old scripts in which they were written and the lack of standard orthography makes them incomprehensible to the wider public. The transcription of the manuscripts in modern script is time consuming and requires many specialized skills and knowledge (paleography, textology, codicology, comparative linguistics, history of culture, and philology)<sup>1</sup>. Critical editions of texts preserved in manuscripts and written in scripts that have not been in use for centuries are therefore of great help and importance to all the interested public. These are the main reasons why scholars and philologists are increasingly

---

\*Scientific Advisor with Tenure, Institute of Croatian Language and Linguistics, Croatia.

<sup>‡</sup>Senior Research Associate, Institute of Croatian Language and Linguistics, Croatia.

<sup>1</sup>The same is stressed in Renhart (2020, p. 311).

interested in digital technologies which in particular cases<sup>2</sup> can improve and accelerate their work. These technologies can also be very helpful in presenting the historical language heritage to the wider interested public.

Nowadays the digitization<sup>3</sup> of historical material in general is essential because of the need to preserve historical material for the purpose of its future use. One of the main reasons for the digitization of old texts is to protect the original document from possible damage caused by improper handling. At the same time, a digital copy is the most reliable backup of the original document.

Retro-digitization of the Croatian linguistic heritage is an important and comprehensive component of the Digital Humanities network in Croatia, and the main goals of this process are the same as in other areas of cultural and historical activities: the preservation of source material, original documents and making them more visible and more easily accessible to the general public (Bago 2014, p. 78). So far, none of the projects of retro-digitization have included Croatian grammar books from the pre-standard period of the Croatian language (before the establishment of the common standard language and orthography). Besides, these grammars have so far not been presented to the public on an easily accessible website or web portal, which has so far made scientific research of their content more difficult.

Within the current project under the title "Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism – RETROGRAM", funded by the Croatian Science Foundation, the most important grammar books from this period will be digitized<sup>4</sup> and their presentation and accessibility to experts and the general public will be improved. In Croatia, this innovative and comprehensive approach of presenting the cultural and language heritage is already used for old manuscripts, vocabularies, etc. but so far not for grammar texts. Even in the European context<sup>5</sup>, this project is unique because it will result in an open-access

---

<sup>2</sup>Such as the OCR (Optical Character Recognition) process to convert manuscript images into Microsoft Word .doc or .docx formats or developed programs like Transkribus for Handwritten Text Recognition (<https://transkribus.eu/Transkribus/>).

<sup>3</sup>By the digitization in this context we mean both the scanned images (digitized) and machine-readable (digital) texts. There is distinction between 'digital' and 'digitized' editions and according to scholars an edition cannot be called truly digital if by printing it we do not have a loss of functionality (Sahle 2008).

<sup>4</sup>According to the distinction mentioned above although our project in its name contains digitization, our goal is to create a digital edition of the grammar text, which will be machine-readable and can be searched full-text.

<sup>5</sup>As an example of good practice in the European context, the Portal of Slovenian grammar and orthography handbooks (from 1584 until today) should be mentioned (<http://www.fran.si/slovnice-in-pravopisi/>). This portal contains facsimiles of each grammar and orthography handbook including metadata (e.g., century, author's name and short biography, original title and the translation of the title if appropriate, language data, information on the importance and purpose of the work, etc.), but texts are not searchable. A similar idea of a digital database of historical linguistic texts appears on the *Classiques Garnier Numérique* website (<https://www.classiques-garnier.com/numerique/>). It contains the *Great Corpus of French Language Grammars, Commentaries and Treatises on Language* (14<sup>th</sup>–17<sup>th</sup> centuries) with a searchable database but with no open access. An open access digital resource of Czech historical grammars (from the 16<sup>th</sup> to 19<sup>th</sup> century) is available on the website *Vokabulář webový/Mluvnice oddělení vývoje jazyka Ústavu pro jazyk*

web portal which will include facsimiles of selected grammar books, their transcription or translation, and an index of historical grammar and linguistic terminology. It will also be equipped with thematic search possibilities at the morphological level.

### History of the Croatian Grammatology

In the history of the Croatian language and long before the standardization of the Croatian language<sup>6</sup>, there was a lengthy tradition of writing and composing grammar and orthographical books and generally volumes in which historical linguistic material, not only in the Croatian language, was interpreted. Since these Croatian grammars were written before the second half of the 19<sup>th</sup> century and before the establishment of the common standard language and orthography based on the Štokavian dialect, they describe all three dialects of the former Croatian literary language: Čakavian, Kajkavian and Štokavian. The standardization of the Croatian literary language based on the Štokavian dialect variant began in the 17<sup>th</sup> century. This process was finalized in the 19<sup>th</sup> century during the time of the Croatian National Revival or the Illyrian movement. The grammars described in this article thus belong to the pre-standard period of the Croatian language. In the literature, the common title "Croatian pre-standard grammars" refers to a heterogeneous group of grammar books with common structural features and methods of description, and whose metalanguage or object language<sup>7</sup> is one of the Croatian literary languages that precede the modern Croatian standard language based on Štokavian dialect, that is, which were written before the formation of the standard Croatian language. The first grammar books were written within the religious orders, of the Jesuits and Franciscans, and were used to teach the Franciscan and Jesuit youth the Croatian or Latin language<sup>8</sup>.

The beginning of the standardization of the Croatian language is connected with the Reformation, or the Catholic Revival. The Catholic Church raised the question of a common Croatian language in which religious books would be written and through which it could renew religious life on Croatian territory among the South Slavs. The problem was the lack of a common language because of the dialect diversity of the Croatian language. The goal was to create a common literary language (*lingua communis*) which could readily be understood by everyone in the dialectically diverse Croatian area.

---

český AV ČR (<https://vokabular.ujc.cas.cz/>). Grammar books are represented by facsimiles and followed by a list of key words that can be traced and attested in the facsimiles.

<sup>6</sup>Before the choice of the common standard language and orthography.

<sup>7</sup>With the term object language, we refer to the language which is the object of grammatical description.

<sup>8</sup>Bartol Kašić, the author of the first Croatian grammar *Institutionum linguae Illyricae libri duo* (1604), was a Jesuit. In the established Academy of the Illyrian language (*Academia linguae Illyricae* 1599) at the Jesuit College in Rome, Bartol Kašić was the first teacher of the Illyrian language. Future missionaries, who would be sent to the Croatian territory during the Reformation, were to learn the language of the Croatian people in that Academy. So, Kašić's task was to write a grammar book of the Croatian language for students, future missionaries, to teach them the Croatian language.

With the first grammar book of the Croatian language, which was written at the beginning of the 17<sup>th</sup> century (Kašić 1604), Croatian was among the first European languages to have its grammatical description. Although the grammar was written in Latin, this was a grammar of the Croatian language that predominantly covered the southeast complex of Croatian literature. This was the reason why the examples in Čakavian/Štokavian literary language stylization with the Ikavian reflex of jat are presented<sup>9</sup>. This grammar thus contains some grammatical categories (for example, the ablative and the gerund) which do not belong to the structure of the Croatian language. However, some of the grammatical descriptions presented in the Kašić grammar are still present nowadays in the standard language Croatian grammars, such as the three nominal declension patterns determined by the ending in the genitive singular. Kašić names the Croatian language that he describes as Illyrian, which was the common name for the inhabitants on the wider southern Slavic area and their language, to whom Jesuit missionaries were sent to preach the Catholic faith and teach the religious youth the Croatian language<sup>10</sup>. The name Illyrian was taken over from the Antique tradition. The name Illyrians referred to the old tribes which inhabited the provinces of the Balkan Peninsula, and the name of those tribes led to the name of the Roman province Illyricum which was established in the 1st century BC. Under the name Illyrian, Kašić refers to a language of the southeast complex of Croatian literature which was based on Čakavian-Štokavian literature. Therefore, in Croatian philology, Kašić's name is connected to the beginning of the standardization of the Croatian language where he intentionally chose the Štokavian language, although he was born a Čakavian (Katičić 1981, Gabrić-Bagarić 2010, pp. 151–153).

The first grammar books written in Croatian were compiled in the middle of the 17<sup>th</sup> century. The first such grammar was written by the Jesuit friar Jakov Mikalja under the title *Gramatika talijanska ukratko ili kratak nauk za naučiti latinski jezik* (1649/1651) but that grammar describes the Italian language (the term Latin refers to the Italian language) while its meta-text is written predominantly in the Croatian language. Italian morphology paradigms and grammatical terms are presented with Croatian language equivalents. Although this is a grammar of a foreign language (Italian), the examples and the declension and conjugation patterns are written in Croatian, and Croatian is also the metalanguage of the grammar. It is a Štokavian grammar and is almost consistently Jekavian. Mikalja's grammar is part of his extensive trilingual dictionary *Blago jezika slovinskoga*

---

<sup>9</sup>In his grammar, the Ijekavian reflex of jat is mentioned only once, in the section where Kašić writes about letters (*Pars prima, De litteris*). Čakavian elements are mainly represented in phonology, especially in accentuation, and Štokavian elements are covered in the morphology part of the grammar book (Katičić 1981, p. 30).

<sup>10</sup>Gessner (1555) describes the Illyrian language in the 16<sup>th</sup> century. He writes that the Illyrian language is the most widespread language on the European territory and is spoken by many different inhabitants. His review of the Illyrian language is concluded with lists of inhabitants who spoke the Illyrian language, among whom he mentions the inhabitants of the Croatian territory (for example, *Dalmatae, Chroati, Istrii* etc.). Another name for this language is the Slavic language (*Slavonica lingua*). (For more about this topic, see Perić Gavrančić (2020, pp. 7–28)). In the Croatian tradition of writing dictionaries and grammar books, the names Illyrian and Slavic were both used to name the Croatian language until the 19<sup>th</sup> century.

(Croatian, Italian, Latin) and it is also the oldest source of Croatian grammatical terminology.

Traditional Latin education and the tradition of literacy in Latin in Croatia, and the incorporation of the coastal Croatian part into the wider Mediterranean space, were the reasons for the first grammars to be mainly oriented towards three languages: Croatian, Latin and Italian. Considering the western orientation of the Croatian language and literature and the long tradition of Latin literacy and translations from Latin, the predominant influence of Latin is apparent in the grammatical descriptions as well, on all language levels. The first Croatian grammars were written in Latin or they describe Latin, given that in Croatian schools education was conducted for a long time in Latin. Indeed, Latin was the language of cultural life, science and schooling.

In later years, when literary and linguistic activity was transferred from the Dalmatian area to the northern and eastern part of Croatia, the Latin model for the description of the Croatian language was still present, but the German language was also used.

The description of the language in the first Croatian grammars is based on the grammar of the Jesuit Álvares (1572)<sup>11</sup>. From the 19<sup>th</sup> century, Croatian grammar books also followed the model of the grammar written by Grigely (1809) who introduced the syntax of cases, since that grammar was prescribed by the Austrian government for use in schools.

Particular grammars, such as the grammar of the Franciscan friar Toma Babić and the grammar of another Franciscan friar Lovro Šitović are in fact adaptations of the Álvares grammars. Babić's grammar (1712) and Šitović's grammar (1745) follow the tradition of the so-called Álvaresian grammar model. Paradoxically, these Latin language-teaching textbooks provide more information on the formation of the Croatian literary language than the first description of the Croatian language, *Institutiones linguae Illyricae* by Bartol Kašić, where the Croatian language is presented only in the translation of Latin morphological paradigms (Perić Gavrančić 2015, p. 58). Their value stems from the fact that they are the source of the earliest standardization process of the Croatian language, as Latin grammatical paradigms and terms are followed by Croatian translations.

The Jesuit Della Bella is the author of the Italian-Latin-Croatian dictionary *Dizionario italiano, latino, illirico* (1728) with instructions for reading, writing and accentuating Croatian words and a short grammar of the Croatian language written in Italian (*Istruzioni grammaticali della lingua illirica*)<sup>12</sup>. Its basic purpose was to help in the teaching of the Croatian language in missionary activities on

---

<sup>11</sup>The Álvares grammar book, soon after its publication, was ordered as the compulsory hand book for all Jesuits schools and it was published in more than 400 editions. Grammar books in many countries were written using it as a model as was the case in Croatia in the pre-standard period of the Croatian language. In the first Croatian grammars, especially in the grammar written by Bartol Kašić, also the influence of the grammars written by Roman grammarian Donat (*Ars minor*, IV century) and Aldii Pii Manutii (*Grammaticarum institutionum libri IV*, 1501) can be noticed (Katičić 1981, pp. 10, 13, 68).

<sup>12</sup>It was published also as a separate work (Della Bella 1837).

Croatian territory<sup>13</sup>. This grammar was published 124 years after the first Croatian grammar (Kašić's *Institutiones*) and thus played an important role in the continuing standardization of the Croatian language. It describes the Ragusan literary language. The special value of this grammar book lies in the fact that Della Bella described the accentual system of the Croatian language where he noticed the connection between the form of the word and its accent, and the accentual variants within one declension type. Another novelty compared with earlier grammar books was perhaps his analysis of the category of animacy, of the verb aspect and the better division of words relating to lexical-grammatical classes and subclasses. Della Bella's grammar book had a great influence on the formation of the Croatian standard language, especially with the regulation of the monosyllabic jekavian replacement of both the short and the long reflex of jat.

Tadijanović's grammar *Svašta po malo iliti kratko složenje imena i riči u ilirski i njemački jezik* (1761) is the first Slavonian-German grammar handbook and the first secular work in renewed Slavonian literature after liberation from Ottoman dominion. It contains a little Croatian and German grammar, a Croatian-German dictionary, conversation forms and letters and a list of proverbs. This handbook is aimed at learning the German language and at expanding knowledge of the author's own Illyrian (Croatian) language, not only for Slavonians but for Croats in general.

Relković is the author of the *Nova slavonska i nimačka gramatika. Neue Slavonische und Deutsche Grammatik* (1767). Relković's handbook is a German-Croatian contrastive grammar and it is very extensive (600 pages). It is written in Croatian but almost all paradigms and grammatical rules are also presented in German. Relković wrote his grammar in the Stokavian dialect but, in terms of lexis, he took all three dialects into account. Relković's grammar is a general language manual but also an extensive source of linguistic terminology. Relković wrote his handbook under the influence of the Enlightenment and his intention was not only to cultivate his mother tongue, but also to encourage young Croatian people to learn German and vice versa.

Marijan Lanosović's work belongs to the opus of Slavonic writers (in the second half of the 18<sup>th</sup> century) who played an important role in the process of the standardization of the Croatian language. His grammar *Uvod u latinsko riči slaganje s nikima nimačkog jezika biličkama za korist slovinskih mladića složen* (1776) is the first Latin grammar compiled with Croatian as the metalanguage. Its particular value lies in promoting the Croatian literary language, given that Lanosović was a high school teacher in a school in Osijek where German was the official teaching language. In such circumstances, in the period of Germanization, Latin language lessons offered many opportunities for cultivating the Croatian language.

From the 18th century, the Kajkavian literary language is described in grammars as well. The first grammar *Gründe der kroatischen Sprache zum Nutzen der deutschen Jugend verfasset* (1779) compiled by Ivan Vitković remains in

---

<sup>13</sup>Jesuit Della Bella (Foggia, Italy, 1655 – Split, 1737), among others, acted as a missionary and preacher in Dalmatian and the Republic of Dubrovnik, and from 1696 to 1702 was the rector of the Jesuit College in Dubrovnik.

manuscript form and thus the first Kajkavian grammar is *Einleitung zur kroatischen Sprachlehre für Deutsche* (1783) written by Ignacije Szentmártony. Although there is no information about the author of this grammar, there is consensus among scholars that this grammar was compiled by Ignacije Szentmártony. The grammar structure differs in a way from other Kajkavian grammars, as it does not contain a chapter on syntax or an added dictionary. Another important Kajkavian grammar is one written by Kornig (1795). In his description, Franjo Kornig largely relied on Szentmártony's grammar, especially in grammar structure and methodology. The grammar is extensive (419 pages) and its structure corresponds to other Kajkavian grammar books, which means that the first part contains a description of the phonology and orthography, while the central part covers morphology descriptions followed by a chapter on syntax and a dictionary as an annex.

The last in a series of pre-standard linguistic handbooks are the grammars written by Josip Voltić and Francesco (Franjo) Maria Appendini. Voltić wrote his grammar *Grammatica Illirica* (1803) as part of a trilingual dictionary of the Illyric, Italian and German languages. If we look back and exclude Mikalja's dictionary, to which the grammar of the Italian language was added, Voltić's *Ričoslovník* is comparable only to Della Bella's *Dizionario Italiano, Latino, Illirico* with a grammar of the Illyrian language as an annex (1728). Voltić was also the first to describe the Croatian language comparatively to Italian and German, bringing his handbook closer to the wider circle of European readers.

Francesco (Franjo) Maria Appendini was born in Italy (Poirino), but from 1791 he lived in Dubrovnik. Although he learned Croatian, he wrote very little in the Croatian language. His most prominent work is *Grammatica della lingua illirica* (1808). It is written in Italian with Croatian examples in morphology paradigms. It is based on Ragusan speech and was compiled for practical purposes – for Italian inhabitants in Dubrovnik to learn Croatian. Appendini used Della Bella's grammar as a template but also relied on Kašić's grammar in some segments.

The first Croatian grammars which were written according to the tradition of the first Croatian grammar, Bartol Kašić's grammar, were not geared towards other Slavic languages at all. The difference was made by Juraj Križanić in his grammar *Gramatično izkazânje ob rúskom jeziku* written in Siberian Tobolsk in 1665. That grammar is, in fact, a grammar of the common Slavic language (a pan-Slavic grammar) and includes a description of Russian, Czech, Polish, Bulgarian, Serbian and Croatian. The grammar is unique in the Croatian grammatical tradition through its content and description of the languages covered. Juraj Križanić created his type of pan-Slavic grammar based on the language facts from the six Slavic languages. Križanić wanted to unite the Roman Catholic and Russian Orthodox churches and he shaped the pan-Slavic language on the basis of Russian. Since the goal of his grammar was to distinguish a particular Slavic language from its general type, Križanić also marked the discrepancies of the Croatian literary language from the general language. The Croatian language that Križanić described is a trilingual hybrid, which includes elements of all three Croatian dialects – (Čakavian, Štokavian and Kajkavian), which was the literary language

used in the 17<sup>th</sup> century in the northwestern part of Croatian territory.

There are a large number of grammars written up to the second half of the 19<sup>th</sup> century which are considered pre-standard Croatian grammars. So far, we have mentioned only a small part of that rich Croatian historical linguistic heritage. The grammatical and orthographical handbooks of the pre-standard period of the Croatian language are the subject of research within the project "Pre-standard Croatian Grammars" in the Institute of the Croatian Language and Linguistics. Within that project, there is a long tradition of editing reference books in the field of historical linguistics and it is very important to continue and foster this tradition in the future. Since 2001, fourteen pre-standard grammar handbooks have been published in the Institute of Croatian Language and Linguistics. This valuable linguistic material which is worth researching has not so far been part of the digitization processes. With the research proposal presented here, we will change this situation.

### **About Project "Retro-Digitization and Interpretation of Croatian Grammar Books before Illyrism"**

This research proposal "Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism"<sup>14</sup> aims to create a model for the retro-digitization of the chosen eight Pre-standard Croatian Grammars (written from the 17<sup>th</sup> to the 19<sup>th</sup> century). The retro-digitization of Croatian grammar books implies the transfer of printed media to computer-readable and searchable text. This project also includes a multilevel mark-up of transcribed or translated grammar text. The digitization process will include the mark-up of grammar book text segments and morphological paradigms by TEI (Text Encoding Initiative) tags and will be conducted on transcriptions or translations of the selected grammar books. The sources include eight printed grammars written in Croatian, Latin, German or Italian in the period from the beginning of the 17<sup>th</sup> century to the beginning of the 19<sup>th</sup> century<sup>15</sup>. The Croatian metatext and examples are written in different pre-standard Croatian orthography manners<sup>16</sup>. The preparatory work, which includes

---

<sup>14</sup>The project which began on 25 November 2019 is financed by the Croatian Science Foundation.

<sup>15</sup>Kašić (1604), Mikalja (1649), Della Bella (1728), Tadijanović (1761), Szentmártony (1783), Voltić (1803). These grammars have already been published as reference books by philologists from the Institute of Croatian Language and Linguistics and associates of this project. In addition, two more grammars which have not yet been published as reference books are added as source grammar texts which will be digitized: Appendini (1808), Lanosović (1776). These two grammars will be published as reference books in 2021.

<sup>16</sup>The first centuries of Croatian literacy in the Latin script are characterized by great inconsistency in writing and orthography because the letters were combined differently to mark the same voice. For example, the digraph *ch* could be read nowadays as *c*, *č*, *ć*, *h* or *k*. We can see the reasons for this obstacle in the attempt to write the Croatian phonological system with Latin letters (there were no Latin letters to mark the special Croatian voices, mainly palatals) and in the fact that the Latin script was taken over as the Croatian script from two different directions: from the Latin and Italian source in southern Croatian territory and from the Latin and Hungarian source in the other parts of Croatia. They are three main alphabet manners: Dalmatian, Kajkavian and Slavonian. The Dalmatian alphabet was modelled on the Italian way of writing, and was used by south Croatian and



the transcription of the Old Latin script or translation into Croatian, on the corpus of historical Croatian grammar books has already been done by philologists from the research group<sup>17</sup>. The manner of grammar editing depends on the particularities of each grammar, as not all grammars are written in Croatian, nor do all the grammars describe the Croatian language. The grammar books which are not written in Croatian are already translated and their text is edited in the reference books. The text of the grammars written in Croatian, but in the old script, is transcribed and edited as the reference book. For each particular grammar the transcribing manners are described in the study of the reference books according to their alphabet, and the main differences in the transcribing procedure are written in the TEI Header as the part of the full bibliographical description of the each digitized grammar. XML (Extensible mark-up language)<sup>18</sup> enables the preservation and digital edition of the text. TEI<sup>19</sup> is one of many XML languages, and can be easily modified for the editing of historical documents of any kind<sup>20</sup>. Further, it will guarantee the standardization of the digital editions and will offer a seamless exchange with other platforms, or their transformation into other formats. With TEI, the encoding process can be adjusted to the specific needs of each editor and to the special demands and purposes of the digital edition. It also guarantees the interoperability and possibility of exchange of digital resources with other platforms and their mutual communication (Bago 2014, p. 77, Renhart 2020, p. 310). Furthermore, it will guarantee the standardization of digital editions or their transformation into other formats.

The main challenge of the project is the fact that the TEI module for grammars has not yet been established, so our work in establishing this standard

---

Bosnian Franciscan writers. The Kajkavian alphabet was modelled on the Hungarian way of writing in the north-west (mainly Kajkavian) part of the Croatian language area. The Slavonian alphabet was formed in the 18th century combining the graphical solutions from the Kajkavian and Dalmatian alphabet. It became the most stable and most acceptable way of writing and was used over the whole territory, that is, outside its regional framework. Finally, it was chosen as the writing manner for schools. The alphabet issue was present in Croatian orthography ever since the Latin script entered into use as one of the former Croatian scripts (along with Glagolitic and the West Cyrillic script). This problem was resolved in the 19<sup>th</sup> century when Ljudevit Gaj composed the Croatian alphabet (*Kratka osnova horvatsko-slavenskoga pravopisa*na (Budim 1830), and later in an article *Pravopis* in the magazine *Danica*, 1835). This alphabet is still in use in Croatia with minor adaptations. Ljudevit Gaj gave his name to the Croatian alphabet which is still called *gajica*.

<sup>17</sup>For this type of text, it was not possible to use any digital tool developed to accelerate transcription because Transcribus is not trained yet for the Old Latin Croatian script. All the transcription work was done manually and in cases where the metalanguage of the grammar was foreign, that text was translated into Croatian as part of the tasks in the process of editing reference books.

<sup>18</sup>Retrieved from: <https://en.wikipedia.org/wiki/XML>.

<sup>19</sup>"TEI is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its main deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. The TEI Guidelines have become an accepted standard for digital text especially where there are concerns about long-term preservation, interchange, or interoperability" (TEI 2007).

<sup>20</sup>TEI is one of many XML schemas, and, from its development in 1987, it has become the most used standard for the digital encoding of literary and language texts (Pierazzo 2014, p. 14). It is widely used by librarians, publishers, and individual scholars to represent all kinds of text material for online research and teaching, and can be easily modified for editing any historical documents.

may have a significant impact on similar undertakings in the future. TEI tags are used to encode features crucial to the linguistic, philological, and textological interpretation of texts to enable further scientific research and interpretation of these grammar books. In this particular case, TEI tags will be used to encode the morphological paradigms of the pre-standard grammar books.

### Web Portal of Pre-Standard Croatian Grammars

The next step of the project is the creation of a Web Portal of Pre-standard Croatian Grammars, on which both the facsimiles and the digitized text of the grammars will be presented. Since there is no common historical linguistic portal in Croatian philology on which we can present our digitized material, the web portal that we are planning will be unique in its content and in the material presented<sup>21</sup>. The planned portal will include metadata of the grammars and will present the results of comparative and philological research and analysis at all levels, from phonology to syntax. Since the pre-standard grammars are a rich source of pre-standard Croatian linguistic and grammatical terminology, we will compile an index of Croatian pre-standard linguistic terminology as part of the portal<sup>22</sup>. The portal will also be equipped with thematic searching possibilities at the morphology level. Connection to the facsimiles will enable comprehensive research on the orthography and traductological aspects of the selected texts.

### Conclusion

There is a long tradition of grammatical description in the history of the Croatian language. But, the creation of retro-digitized resources in Croatia is still in its early stages and the existing resources do not include historical grammars. With the project "Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism – RETROGRAM" this situation will change. The importance of this project is therefore significant not only for linguists and philologists, but also for historians and those interested in cultural heritage in general. Our aim is to present to the wider and international public the attainment of the Croatian language and linguistics as an important part of Croatian culture in general. Above all, the project will make it possible to link and exchange knowledge with other institutions and experts and similar ongoing processes in the digitization of historical heritage (*Europeana*, *Manuscriptorium* – Czech Digital Library etc.). The web portal of the pre-standard Croatian grammars will be open to further upgrading and will offer a model for similar research and a model for the preservation and presentation of the

---

<sup>21</sup>For more on the lack of a central historical web portal and on existing platforms, see in Kramarić (2020, pp. 186–206).

<sup>22</sup>Croatian pre-standard linguistic terminology will also be extracted from the following grammars, which are not part of the grammars included into the annotation process: Matijević (1771), Mikloušić (1796), Rožić (1820), Rožić (1821).

historical linguistic heritage. It would also lay the foundations for creating a reliable historical grammar of the Croatian language. The project in general, in its comprehensiveness, would give full insight into the Croatian pre-standard grammatical tradition, enabling further research, especially comparative research, opening new perspectives on pre-standard grammatology and systematizing pre-standard linguistic terminology.

### Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism IP-2018-01-3585.

### References

- Álvares E (1572) *De institutione grammatica libri tres* (Grammatical instructions of the Latin language in three books). Lisbon.
- Appendini FM (1808) *Grammatica della lingua illirica* (Illyrian grammar). Dubrovnik.
- Babić T (1712) *Prima grammaticae institutio pro tyronibus illyricis accomodata* (The first grammatical instructions adapted for the Illyrian youth). Venice: Per Bartolo Occhi.
- Bago P (2014) Označavanje natuknice u della Bellinu rječniku i povezivanje autora izvora citata s vanjskim resursima (Marking dictionary entries in Della Bella's dictionary and linking authors of the quotation sources to external resources). *Studia Lexicographica* 2(15): 77–102.
- Della Bella A (1728) Istruzioni grammaticali della lingua illirica (Grammatical instructions of the Illyrian language). In *Dizionario Italiano, Latino, Illirico*. Venice.
- Della Bella A (1837) *Principi elementari della grammatica illirica*. (Basic principles of the Illyrian grammar). Dubrovnik.
- Gabrić-Bagarić D (2010) Četiri ishodišta hrvatskoga standardnoga jezika (Four starting points in the history of the Croatian standard language). *Fluminensia: Časopis za Filološka Istraživanja* 22(1): 149–162.
- Gessner C (1555) *Mithridates (sive) de differentiis linguarum* (Mithridates (or) of different languages). Retrieved from: <https://bildsuche.digitale-sammlungen.de/index.html?c=viewer&bandnummer=bsb00039584&pimage=169&v=100&nav=&l=en>. [Accessed 19 February 2021]
- Grigely I (1809) *Institutiones grammaticae in usum scholarum grammaticarum regni Hungariae et adnexarum provinciarum* (Grammatical instructions for the schools in the Hungarian monarchy and the belonging provinces). Budae: Typis Regiae Universitatis Hungaricae.
- Institut za hrvatski jezik i jezikoslovlje. *Retro-digitization and interpretation of Croatian grammar books before Illyrism – RETROGRAM*. Retrieved from: <http://ihjj.hr/projekt/retrodigitalizacija-i-interpretacija-hrvatskih-gramatika-do-ilirizma-ndash-retrogram/89/>. [Accessed 19 February 2021]
- Kašić B (1604) *Institutionum linguae Illyricae libri duo* (A short grammar of the Illyrian language). Rome.
- Katičić R (1981) Gramatika Bartola Kašića. (Grammar by Bartol Kašić). *RAD JAZU* 388: 5–129.

- Kornig F (1795) *Kroatische Sprachlehre, oder Anweisung für Deutsche, die kroatische Sprache in kurzer Zeit gründlich zu erlernen, nebst beigefügten Gesprächen und verschiedenen Übungen* (Croatian language classes or instructions for Germans to learn the Croatian language thoroughly in a short time, with accompanying discussions and various exercises). Zagreb.
- Kramarić M (2020) Towards new approaches to editing of old manuscripts and documents. In M Tomić, M Willer, N Tomašević (eds.), *Empowering the Visibility of Croatian Cultural Heritage through the Digital Humanities*, 186–206. Cambridge Scholars Publishing.
- Križanić J (1665) *Gramatično iskazanje ob rúskom jeziku* (Grammar of the Russian language). Tobolsk.
- Lanosović M (1776) *Uvod u latinsko riči slaganje s nikima nimačkog jezika biličkama za korist slovinskih mladića složen* (An introduction to the Latin grammar with some notes in the German language compiled for the Slavonic youth). Osijek.
- Matijević J (1771) *Pomum granatum oder der wahre Kern der deutschen Sprache, kroatisch expliziert* (Pomum Granatum or the real core of the German language, Croatian explicit). Zagreb.
- Mikalja J (1649) Gramatika talijanska ukratko ili kratak nauk za naučiti latinski jezik. (Short Italian grammar). In *Blago jezika slovinskoga*. Loreto.
- Mikloušić T (1796) *Syllabus vocabulorum grammaticae Emmanuelis Alvares, in Croaticam linguam conversorum* (The wordlist of Alvares' grammar translated in Croatian). Zagreb.
- Perić Gavrančić S (2015) Lanosovićev prinos izgradnji hrvatskoga jezikoslovnoga nazivlja (The Lanosović's contribution to the development of the Croatian linguistic terminology). In M Samardžija (ed.), *Hrvatski jezik i pisana riječ XVII. i XVIII. stoljeća: Zbornik radova*, 56–60. Vinkovci: Privlačica.
- Perić Gavrančić S (2020) *Sic etiam Croati*: bilješke o hrvatskome jeziku u humanističko-renesansnim jezičnim repertorijima (*Sic etiam Croati*: notes on the Croatian language in the linguistic repertoires of renaissance humanists). *Povijesni prilozi* 39(58): 7–27.
- Pierazzo E (2014) *Digital scholarly editing: theories, models and methods*. hal-01182162. Retrieved from: <http://hal.univ-grenoble-alpes.fr/hal-01182162/document>. [Accessed 19 February 2021]
- Relković MA (1767) *Nova slavonska, i nimačka gramatika* (The new Slavonian and German grammar). Zagreb.
- Renhart E (2020) Glagolitic text editions and a perspective towards digital humanities. In T Katić, N Tomašević (eds.), *Zbornik Mirna Willer: Festschrift*, 303–314. Zadar: University of Zadar.
- Rožić A (1820) *Kratko naputjenje vu kruto hasnoviteh, i zevsema potrebnih temeljih dijačkoga jezika* (The short grammar of the Latin language). Budim.
- Rožić A (1821) *Prvi temelji dijačkoga jezika za početnike vu domorodnom jeziku* (The first elements of the Latin language written in Croatian). Varaždin.
- Sahle P (2008) *A catalogue of digital scholarly editions*. Retrieved from: <http://www.digitale-edition.de/>. [Accessed 19 February 2021]
- Šitović Ljubušak L (1745) *Grammatica Latino-Illyrica ex Emanuelis* (The Latin-Illyrian grammar modeled after Emanuel Álvares' grammar). Venice.
- Szentmártony I (1783) *Einleitung zur kroatischen Sprachlehre für Deutsche* (Introduction to the grammar of the Kajkavian language). Varaždin.
- Tadijanović B (1761) *Svašta po malo iliti kratko složenje imena i riči u ilirski, i njemački jezik* (A miscellany or a short overview of the grammar of nouns and verbs in Illyrian and German). Magdeburg.
- TEI (2007) *Guidelines*. Retrieved from: <https://tei-c.org/guidelines/>. [Accessed 19 February 2021]

2021]

Vitković I (1779) *Gründe der kroatischen Sprache zum Nutzen der deutschen Jugend verfasst* (The basis of the Croatian language written for the German students)

Voltić J (1803) *Grammatica Illirica* (Grammar of the Illyrian language). In *Ričoslovník ilričkoga, italianskoga i nimačkoga jezika s' jednom pridpostavlienom gramatikom illi pismenstvom: sve ovo sabrano i složeno od Jose Voltiggi Istrijanina*. Vienna.

