

Language Identity Through Cyrillic Script

From Romanian to Moldovan by Automatic Transliteration in the Wikimoldia Project

Christian Koch

Abstract. Wikimoldia was a website active from 2018 to 2019 on which the Romanian Wikipedia was transliterated into Cyrillic by using a PHP script. This paper discusses the technical background of the automatic transliteration performed in Wikimoldia and links the project of a Cyrillic-language Romanian Wikipedia to the political and linguistic controversies surrounding the status of the Moldovan language. It discusses how Wikimoldia can benefit the Cyrillic-socialised minorities in the eastern periphery of Romanian-speaking areas. The use of machine transliteration can also be of interest in the context of other multialphabetic languages.

1. Introduction

Wikimoldia sounds like the name of another of Wikipedia's many offshoots (cf. Wikimedia, Wikisource, etc.), but alludes to a geographical area, namely the Republic of Moldova. If one tries to open the page <http://wikimoldia.org> today, there will not be anything of what was offered here between September 2018 and September 2019: a Romanian Wikipedia in Cyrillic script.

The aim of this article is to reconstruct the technical functioning of this automatically generated page and to discuss the potentials and difficulties of the project. Furthermore, the paper deals with the sociolinguistic background that may have motivated the Wikimoldia project. Understanding linguistic diversity in Wikipedia as a contribution to minority language vitalisation (cf. Born, 2007; Coulmas, 2018, 198f.), Wikimoldia can also raise the question of an attempt to vitalise this language, however it has to be defined at the same time what kind of language Moldovan actually is. From a linguistic point of view, the term 'Moldovan language' is highly problematic and requires a critical analysis, not least in view of the fact that, even within Romance linguistics,

Christian Koch  0000-0002-6697-3468

Romanistik / Angewandte Sprachwissenschaft und Didaktik, Universität Siegen,
Adolf-Reichwein-Str. 2, AR-IF 229, D-57076 Siegen, Germany
E-mail: koch@romanistik.uni-siegen.de

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2020. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 5.
Fluxus Editions, Brest, 2021, pp. 1067–1082. <https://doi.org/10.36824/2020-graf-koch>
ISBN: 978-2-9570549-7-8, e-ISBN: 978-2-9570549-9-2

the Romance-speaking varieties in the far east of Europe are among the rather unknown territories. The first part of the article is devoted to this aspect, before the following sections take a closer look at Wikimoldia from a technical and functional perspective.¹

2. On the Status of the Moldovan Language

2.1. Stages of Language Naming

The term ‘Moldovan language’ (Rom. *limbă moldovenească*) has a different meaning in different political contexts and time periods. Gabinskij (2002, p. 133) indicates ‘Moldovan’ as a non-scientific everyday term² for the language of the Republic of Moldova, but also as a subglottonym of the glottonym ‘Romanian’ (ibid., p. 139), which may seem acceptable from a scientific point of view, if one speaks of a (geopolitical) variety of Romanian rather than of a language of its own. However, this also seems problematic because in dialectological descriptions the Moldovan dialect is understood as a geolectic area north of the Daco-Romanian dialect (cf. Olariu, 2017, p. 108) and this geolectic area is largely situated within the political borders of Romania.

The territory of the modern Moldovan Republic had an eventful history along the 20th century: Bessarabia, previously part of the Russian Empire, became in its majority a part of Greater Romania in 1918, then in 1944 it converted into the Moldovan Soviet Socialist Republic, before the independent Republic of Moldova was founded in 1991. The affiliation to the Russian Empire and the Soviet Union is still visible today in the presence of the Russian language and the Cyrillic alphabet. Since its foundation, the present state has been divided politically and, to a certain extent, linguistically into two regions on both sides of the Dniester River. Transnistria as a breakaway republic with half a million inhabitants in eastern Moldova confronts the country with a conflict which is still insoluble and hinders the integration into European institutions and in the rapprochement with Romania. The ongoing linguistic separation is based on the use of the Latin alphabet, as—with the exception of Transnistria—the Latin alphabet has been reintroduced in the independent Republic of Moldova: “decretarea limbii române ca limbă de stat și reintroducerea alfabetului latin, din 3 noiembrie 1990”³ (cited in

1. This article a slightly extended and revised English translation of Koch (forthcoming).

2. German original: “(nichtwissenschaftliche) Alltagsbezeichnung”.

3. Translation: the decree of Romanian as the national language and the reintroduction of the Latin alphabet, from 3 of November 1990.

Cimpoeșu and Musteață, 2018, p. 50). While the Declaration of Independence that is quoted here still refers to the Romanian language, a new decree was issued in 1994: “limba de stat a Republicii Moldova este limba moldovenească”⁴ (cited in Olariu, 2017, p. 22). This can be understood as a return to Soviet identity construction which the Romanians Dorin Cimpoeșu and Sergiu Musteață accuse in harsh words:

au fost legiferate tezele staliniste false despre apartenența etnică și lingvistică a populației românești prin introducerea în legea fundamentală a sintagmelor ‘limbă moldovenească’ și ‘popor moldovenesc’ contrare adevărului științific și istoric⁵ (Cimpoeșu and Musteață, 2018, p. 61)

However, one can also more moderately assume a national identity building in which Moldova breaks away from its position as Romania’s satellite state and promotes linguistic independence with its own glottonym. A bit later, the national anthem entitled *Limba Noastră* has also been established, which, along with the national holiday *Limba Noastră (cea Română)*, emphasises the outstanding importance of the national language in Moldova. The text of the anthem, which goes back to a much older poem by Alexei Mateevici (1888–1917), uses the politically more neutral possessive determiner (‘our language’) instead of a glottonym.

As an official language, Moldovan had a coding in the ISO 639 standard as “mo”/“mol,” which was however already abolished in 2008 (cf. <https://iso639-3.sil.org/code/rum>). Finally, in 2013 it was decided to revert to the designation *limbă română* in official language use, probably also in order to strengthen the ties with Romania and thus the bridge to the European Union. This step was justified by the designation in the declaration of independence:

prevederea conținută în Declarația de Independență referitoare la limba română ca limbă de stat a Republicii Moldova prevalează asupra prevederii referitoare la limba moldovenească conținute în articolul 13 al Constituției⁶ (cited in *ibid.*, p. 24).

In Transnistria, the renaming of the language was not applied, but Moldovan in Cyrillic script remained the official language (along with

4. Translation: the official language of the Republic of Moldova is the Moldovan language.

5. Translation: false Stalinist theses on the ethnic and linguistic affiliation of the Romanian people were established by law, introducing in the basic law the expressions “Moldovan language” and “Moldovan people,” which contradict scientific and historical facts.

6. Translation: the provision on the Romanian language as the official language of the Republic of Moldova contained in the Declaration of Independence takes precedence over the provision on the Moldovan language in Article 13 of the Constitution.

Russian, Ukrainian and, regionally, Gagauzian). It should also be mentioned here that the Romanian language area extends beyond the eastern border of Moldova into Ukraine where the Romanian-speaking minorities in northern Bucovina and the Ukrainian part of Transnistria (or Bessarabia), but less frequently in Transcarpathia, feel close to Moldovan identity (cf. Dahmen, 2018, 345ff). Due to the “official” language names used today, the name *limbă moldovenească* (or *ли́мбэ молдовеняскэ*) is increasingly narrowed down to the language of Transnistria and the mentioned regions of Ukraine. The justification for the use of this glottonym is linked to politically sensitive Transnistrian normative concepts. From a scientific point of view, the term ‘Moldovan’ can be used in this context as a functional term for ‘Cyrillic Romanian’ without any political classification. The fact that it is primarily the writing system and less the diatopic variety that is decisive has to do with the way in which transliteration is carried out. This will be the subject of the next section.

2.2. Principles of the (New) Cyrillic Script of Romanian

The adjective *new* indicates that the principles of Cyrillic script in the 20th century are not related to the use of Cyrillic in early Romanian writing since the 16th century, because Latin and Cyrillic letters had already been in competition with each other for more than 300 years before the Latin alphabet became established (cf. Onu, 1989). With the beginning of Joseph Stalin’s rule over the Soviet Union, the Cyrillic alphabet became the identity-giving symbol for the majority of the regional languages in USSR. In the part of Transnistria not belonging to Greater Romania, the Cyrillic script was introduced in 1928 for Romanian writing and, with a brief interruption between 1933 and 1937, became firmly established (cf. Kramer, 1989, p. 15). As mentioned above, Moldova then became Soviet in 1944 and the Cyrillic alphabet was introduced throughout the Soviet Republic, where it remained until independence in 1991.

The most important principle of the Cyrillic transliteration⁷ from 1928 onwards was—as with all newly written languages of the Soviet Union—the greatest possible harmony with the phoneme-grapheme correspondences of Russian. However, the languages were also granted a

7. According to Zikmund (1996, p. 1592), ‘transliteration’ is understood as the language-indifferent transliteration opposed to target language-specific transcription. In this distinction, the term ‘transcription’ would also be conceivable, since numerous specific features of Russian are present in Cyrillic writing. However, unlike the transcription of proper names, for example, it is not a form of integrating written forms into the target language Russian, but rather the transcription of an entire language system, which is why we prefer the term of transliteration.

certain degree of autonomy, so that their graphemes did not need to reflect all the special features of Russian. Individual letters of the Russian alphabet were converted and new letters were introduced—exactly one in the case of Romanian. This will be illustrated by a few examples (Tab. 1).

TABLE 1. Examples of Romanian letters in Cyrillic script

| Example | Sound | Russian | „Moldovan“ | Romanian |
|---------|-------|-----------|------------|----------|
| (1) | [i] | ы | ы | â, î |
| (2) | [j] | ь | ь | î |
| (3) | [ə] | (ə ≐ [ɛ]) | э | ă |
| (4) | [ɕ] | - | ж̣ | g (+e/i) |
| (5) | [jʲa] | ия | ия | ia |
| (6) | [l] | ль | л | l |

1. The unrounded closed central vowel [i] is characteristic of both Russian and Romanian, and the Cyrillic writing can solve the problem of the two graphemes <â> and <î> of Romanian.
2. The letter <i> is used in Romanian to indicate, inter alia, the palatalisation of consonants which are syllable-final or usually word-final. In some Slavic languages, the so-called soft sign <ь> follows on the palatalised consonants. The difficulty of transliteration due to the different functions of <i> is explained below.
3. For the Romanian language, the shwa sound written as <ă> is characteristic and not uncommon even in stressed syllables. Russian knows the sound only in unstressed syllables as a reduction level of /a/ and /o/. At this point, an imitation of Russian grapheme-phoneme correspondence would hardly be possible, and instead the third-last letter <э> of the Russian alphabet is used, whose Russian phonological value /ɛ/ is not necessary for the representation of a specific sound in Romanian.
4. Romanian and Gagauzian have got their own letter for the affricate /ɕ/: <ж̣>, which is distinguished by a diacritic breve from <ж>, because <ж> as the transliteration of Rom. <j> is also needed. This shows the preference in Cyrillic scripts for diacritics over digraphs, although a spelling like <дж> would be intuitively easier to read.
5. In various Slavic languages, iotation plays an important role, i.e., some vowels have an approximate initial [j]. Russian has its own letters for this purpose (<ю>—[ju], <я>—[ja], possibly also <ë>—[jo]). The letter <e>, which is identical in both alphabets, is regularly used with iotation in both Russian and Romanian. Another special feature of Russian is the graphic marking of the intervocalic iotation, which

is also conceivable as an epenthesis in Romanian, but is not graphically marked. As the last letter in the name of *Викимолдия* shows, the transliteration is based on the graphic representation of the epenthesis according to the Russian model.

6. However, the name *Викимолдия* shows that <л> can work without a soft sign for the articulation of the lateral [l], since the articulation of velarised (or hard) [ɫ] is irrelevant for Romanian.⁸

Remarkable about the transliteration rules developed during the Soviet era is that regional peculiarities of the articulation of the Moldovan variety were not taken into account. Instead, the pronunciation standard of Daco-Romanian (cf. Gabinskij, 2002, p. 135) is fully applicable and no attempt has been made—as it has happened with other Soviet regional languages—to represent a variety in Cyrillic that did not have any established writing system before. The demarcation from Romania rather occurred in the field of lexis as a rejection of the so-called *limbă păsărească* ('bird language'), which denotes the more sophisticated language oriented towards the Romanian norm (cf. *ibid.*, p. 135).

3. Key Data on Wikimoldia

The site <http://wikimoldia.org> was launched in September 2018 and was then accessible for one year. We can presume that the termination of the online presence happened due to the fact that the contract of use for the domain was not renewed. Therefore, direct access to the URL is no longer possible. Via the large web archive *Wayback Machine* (https://web.archive.org/web/*/http://wikimoldia.org/), 520 pages are still available, but this is only a fraction of the several hundred thousand pages once available.

Wikimoldia can be seen as a phantom page to the Romanian Wikipedia, because the contents of Wikipedia are transliterated into Cyrillic with the help of a PHP script, that we are going to discuss in more detail in the following section. In addition, the name *Wikipedia* is replaced by *Викимолдия*, so that the artificiality of the pages is not obvious at first glance (Figs. 1 and 2).

In the top left-hand corner, it is noticeable that the signature of the logo "Wikipedia / Enciclopedia liberă" has not been transliterated or replaced, as this is a graphic element. The exact layout and hypertext structure are transferred to Wikimoldia; all hyperlinks to articles work and lead to corresponding pages in Wikimoldia. However, the search field at the top right is not functional (Fig. 3).

8. This does not mean that in Russian there would be a soft sign at this point. Rather, unlike in Romanian, *Молдавия* is articulated with velarised [ɫ].

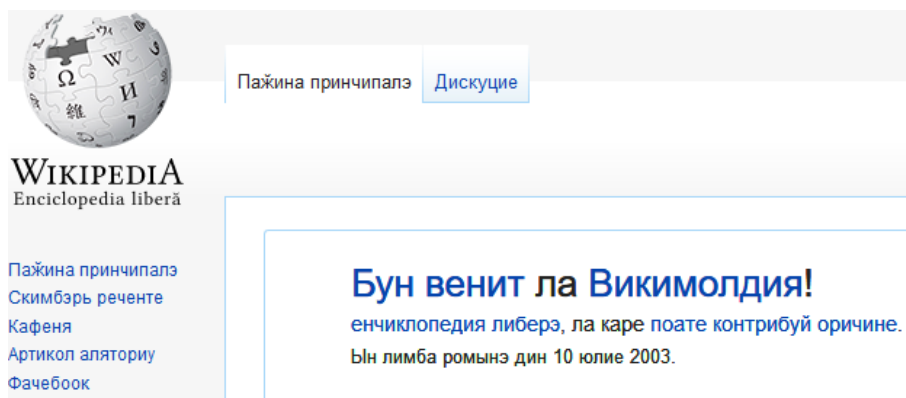


FIGURE 1. Partial screenshot of the Wikimoldia homepage

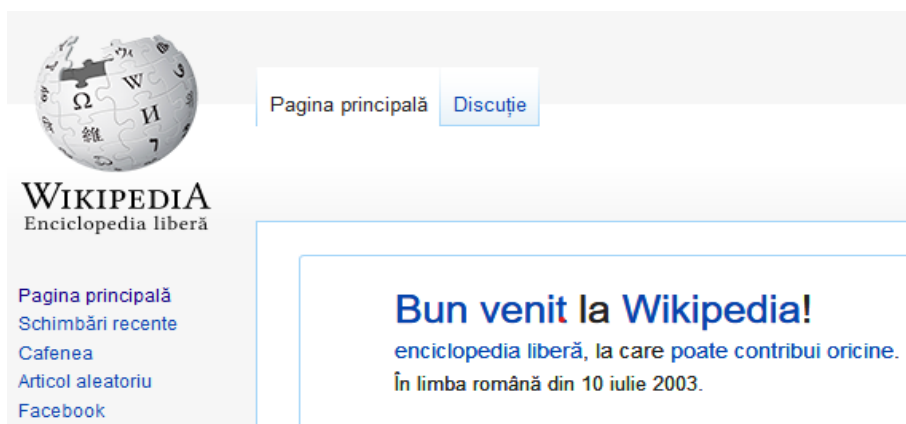


FIGURE 2. Partial screenshot of the Romanian Wikipedia homepage (ro.wikipedia.org)

It is only possible to enter words in Latin script and this will bring the user to the Romanian Wikipedia. Thus, in the sense of Leca-Tsiomis (2006), the encyclopaedic order works by the hypertextual reference structure, but not the alphabetical order in the form of direct look-up. In Wikimoldia it was only possible to call up an article in a targeted manner by manually replacing “ro.wikipedia” with “wikimoldia” in the URL of the Romanian Wikipedia article.

Before discussing the functionality and potentials of Wikimoldia in more detail in the following sections, the question should be allowed whether Wikimoldia could actually represent something like a vitalisation attempt, i.e., in particular whether the Cyrillic transliteration of

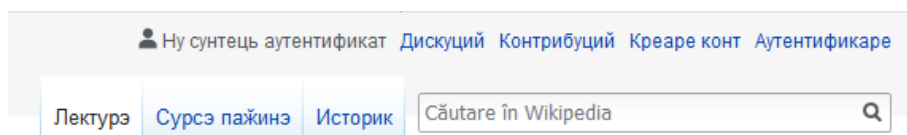


FIGURE 3. Search field “Căutare în Wikipedia” in Wikimoldia

the Romanian Wikipedia was seriously used. In June 2019, the usage data analysis of SimilarWeb (<https://www.similarweb.com>) still provided numbers for the period May to March 2019 (Tab. 2).

TABLE 2. Wikimoldia user numbers according to *SimilarWeb*

| | |
|-------------------------------|----------|
| Total Visits (March-May 2019) | 46,122 |
| Monthly Visits | 15,374 |
| Monthly Unique Visitors | < 5,000 |
| Avg. Visit Duration | 00:01:07 |
| Pages / Visit | 1.53 |

In the three months, the website was accessed 46,122 times, with the number of users estimated at less than 5,000. On the basis of the described diversion via the Romanian Wikipedia articles, one could assume that the number of actual users could be higher, since—unlike in Wikipedia—it was not possible to access concrete articles via the Wikimoldia homepage. *SimilarWeb* also made it possible to determine from which countries users accessed the site (Tab. 3).

TABLE 3. User proportions by country according to *SimilarWeb*

| | Country | Traffic Share | Country Rank |
|---|---------------|---------------|--------------|
| 1 | Ukraine | 33.36% | ·148,316 |
| 2 | Russia | 16.31% | ·722,063 |
| 3 | Moldova | 13.02% | ·42,216 |
| 4 | Turkey | 10.21% | ·1,153,092 |
| 5 | United States | 4.39% | ·1,441,191 |

According to this data, Moldova is only in third place after Ukraine and Russia. This may be simply because Ukrainian and Russian IP addresses are used in Transnistria. But it can also be linked to the Romanian-Moldovan minorities mentioned above, especially in Ukraine. Moreover, the higher number of users from these coun-

tries could also have something to do with the Cyrillic script, which builds a bridge to Romanian here. In August 2019, for example, <https://news.ru> quotes Wikimoldia as the source for a photo of the former Romanian Education Minister Ecaterina Andronescu (<https://news.ru/europe/slova-ob-iznasilovannoj-devochke-stoili-rumynskomu-ministru-dolzhnosti/>). It is probable that a Russian reporter in the search for a picture of “Екатерина Андронеску” has just found a relevant result via Wikimoldia. This example, as well as the user figures mentioned above, provide rough indications of a certain vitality of use for the period during which Wikimoldia was online.

4. Automatic Transliteration in Wikimoldia

4.1. Operating Modes

For the automatic transliteration of Romanian into Cyrillic letters according to the principles described above, the PHP script `slava37md2` was published in *GitHub* in August 2018 (<https://github.com/slava37md2/wikimoldia>). It contains 361 paragraphs or 9,807 characters with so-called assignment operators for all letters of the Romanian alphabet, as well as for <k>, <q>, <w>, <x> and <y>, each separated into upper- and lowercase letters—the rules for upper- and lowercase are identical for the Latin and Cyrillic writing of Romanian. A very short readme file contains an explanatory description in English and Russian: “Script translits romanian (latin) to moldovan (cyr) characters Скрипт переводит румынские буквы в кириллицу. Можно переводить интернет-страницы. Например Википедию.”⁹

Simple assignments can be made where a Romanian grapheme in Latin script has exactly one Cyrillic equivalent, e.g., <J> → <Ж> or <d> → <д>. Accordingly, the assignment operator consists of only one command (Fig. 4).

```
case "J":  
    echo "Ж";  
    break;
```

FIGURE 4. Script of a simple assignment operator

The graphemes of Romanian which have different phonetic realisations depending on their position—in particular <g> and <c>—must

9. Translation of the Russian sentences: The script translates Romanian letters into Cyrillic. You can translate websites. For example Wikipedia.

each be assigned to different Cyrillic letters, because the grapheme-phoneme-correspondences in Cyrillic Romanian or “Moldovan” are almost completely unambiguous.¹⁰ Like in Italian, the Romanian <c> when preceding the vowels <e> and <i> is articulated as a voiceless prepalatal affricate [tʃ]. This articulatory rule can be resolved by the digraph <ch> which, like <c> in all other positions, leads to velar articulation as [k]. In automatic transliteration, a complex assignment operator generates the letters <ч> and <к> depending on the position (Fig. 5).

```
case "c":
{
  if ($str[$i+1]=="e" or $str[$i+1]=="i"
      or $str[$i+1]=="E" or $str[$i+1]=="I")
  { echo "ч"; break; }
  if ($str[$i+1]=="h" or $str[$i+1]=="H"
      and ($str[$i+2]=="e" or $str[$i+2]=="i"
          or $str[$i+1]=="E" or $str[$i+2]=="I"))
  { $i=$i+1; echo "к"; break; }
  echo "к";
  break;
}
```

FIGURE 5. Script of a complex assignment operator

The following applies to complex assignments: Each larger defined unit of characters has priority over smaller character units or individual characters. So simple assignments can be supplemented by additional condition sets for special cases. Nevertheless, the grapheme <i> (or <I>), on which almost a quarter of the entire script is used, is at the limit of automatic assignment possibilities. As an isolated vowel, <i> should be rendered as <и>. However, especially at the end of a word, <i> is usually not a vowel but indicates palatalisation (<ь>, cf. Tab. 1, Ex. 2). In rising diphthongs, the iotified letters <ю> and <я> mentioned in Tab. 1, Ex. 5, are used. In falling diphthongs, such as <ei>, but also in the double <ii> that is frequent in Romanian, <й> is used, i.e., <ей> and <ий>.

A short text example from the Romanian Wikipedia (<https://ro.wikipedia.org/wiki/Portal:Limbi>) and the resulting transliteration will illustrate some of the complex assignments:

10. The only exceptions to graphophonetic and phonographic unambiguity are the iotified letters <ю> and <я>, which are popular in Cyrillic (see Tab. 1, Ex. 5, and the transliteration of <i> below).

- (1) O limbă reprezintă un sistem abstract, complex, de comunicare verbală între oameni. În afară de forma orală (**limba vorbită**), bazată pe articularea de sunete, limbile actuale au în general și o formă grafică, **limba scrisă**.
- (2) О лимбэ репрезінтэ ун систем абстракт, комплекс, де комуникаре вербалэ ынтре оамень. Ын афарэ де форма оралэ (**лимба ворбитэ**), базатэ пе артикуларя де сунете, лимбиле актуале ау ын **женерал ши** о формэ графикэ, лимба скрисэ.

There is nothing wrong in the automatic transliteration of the Romanian text. Complex assignments are marked here, where the functionality is clearly visible: the grapheme <g> is transliterated according to the pronunciation once as <r> and once as <ж>. In the transliteration of <i>, the assignment operator distinguishes between palatalisation in “оамень” and vowel realisation in “ши,” whereby a final <i> is in most cases to be read as a palatalisation sign, but not in monosyllabic words such as *și*, where it must be vocal.¹¹ To end, in the transliteration “articularea,” it can be seen that the final vowels are not individually recognised as hiatus, but as a diphthong, reproduced with <я>.

4.2. Limits of the Automatic Transliteration

The text example shows overall that the assignment operators work. However, in the following, some problems that cannot be solved with the script are pointed out. Firstly, a number of special characters are missing, i.e., letters with diacritics and other letters that go beyond the basic Latin alphabet and that are not represented in Romanian. For example, the name *Frédéric Chopin* is transliterated as *Фредѣрик Кхонин*, with the two vowels <é> remaining, and the digraph <Ch>, since it is not placed before <e> or <i>, is interpreted as two single consonants. Moldovan spelling would normally orientate on the Russian variant *Фридерик Шопен*. So, it is not only special characters that cause difficulties, but also foreign-language sound patterns, which would be rendered phonetically when transferred to another writing system. It would be conceivable, however, to make automatic transliteration capable of learning for this purpose, since it follows the principle—as described above—that each larger sequence of letters is given priority over smaller combinations and individual characters. Accordingly, proper names could be continuously included into the script.¹²

11. The final <i> is not recognisable in the script as a vowel in infinitives (e.g., *a veni*). This is discussed below (Tab. 5).

12. This alone, however, cannot solve the problem that when transliterating proper names, the original spelling would be added as a parenthesis. Particularly problematic is the retransliteration of proper names previously transcribed into the Latin alphabet, which are written in Cyrillic in the original.

Another problem are symbol letters and abbreviations. In chemistry, for example, element symbols are represented by Latin letters independently of the writing system of a language. In Wikimoldia, however, “H₂O” changes to “X₂O”. For Roman numerals at least, a number of instructions has been preserved to prevent nonsensical transliteration. An incorrect abbreviation is pointed out in the forum for the script in *GitHub* (<https://github.com/slava37md2/wikimoldia/issues/2>): due to the positional phonetic realisation of the Romanian grapheme <c> (cf. Fig. 5), an error occurs in the abbreviation of ‘centimetre’ (Tab. 4).

TABLE 4. Incorrect automatic transliteration of Rom. *cm*

| Romanian | correctly transliterated | automatically transliterated |
|-----------|--------------------------|------------------------------|
| <i>cm</i> | <ЧМ> | <КМ> |
| <i>km</i> | <КМ> | <КМ> |

Another special case of Romanian, which cannot be fully resolved in automatic transliteration, is homography, i.e., words that are spelt the same but pronounced differently. These include infinitives with the vocal ending *-i* (Tab. 5).

TABLE 5. Incorrect automatic transliteration of Rom. *dormi*

| Romanian | pronunciation | correctly transliterated | automatically transliterated |
|---------------------|---------------|--------------------------|------------------------------|
| <i>tu dormi</i> | [ˈdormʲ] | <дормь> | <дормь> |
| <i>tu vei dormi</i> | [dorˈmi] | <дорми> | <дормь> |

In principle, the palatalising function of <i> at the end of a word is much more frequent, so that the special case of the infinitive is less significant. Homographs sometimes occur between the infinitive and the conjugated form for second person singular in present tense. But numerous verbs of the *i*-group have stem extensions (e.g., *a citi—tu citești*) or irregularities (e.g., *a veni—tu vii*). In these cases, there are no homographs and the infinitives could be assigned to the correct transliteration as individual lexemes. For the remaining verbs with homograph forms (*a dormi, a fugi, a ieși...*) a fully automatic software should be able to distinguish between infinitives and conjugated verb forms. In view of the (Balkan-typical) restrictive use of the infinitive, which only allows its use in a few constructions with auxiliary verbs or the preposition *a*, it would be conceivable to program with so-called regular expressions,

which can formally map the syntactic embedding of infinitives beyond the level of lexemes.

The problems with automatic transliteration and their solutions presented in this section are purely technical. From a sociolinguistic perspective, however, the question may be asked concerning the interest of an error-free transliteration of Romanian into Cyrillic script. This will be the subject of the following section.

5. The Potentials of Wikimoldia

The previous considerations about Wikimoldia are based on insights into the website when it was still active, on the analysis of the PHP script in GitHub, and on the analysis of usage data. The research for further background information on the creation and motivation of Wikimoldia remained fruitless. Thus, we can only speculate about the intended function of the website. The possible authorship, which is particularly important with regard to a conceivable political motivation, will also be discussed in this context.

Wikimoldia can be understood as an access to the Romanian Wikipedia for a Cyrillic socialised audience. This is particularly important for the population of Transnistria, but also for Romanian minorities in Ukraine. Although most Romanian speakers should be able to read the language in Latin script, in terms of literacy the possibility of reading in Cyrillic-script represents an additional benefit, provided that the readers do not switch completely to the Russian or Ukrainian Wikipedia. The above example, in which Wikimoldia was quoted on a Russian news site, also shows that Moldova's neighbours—potentially also Romania's Cyrillic-writing neighbours (Serbia, Bulgaria, Northern Macedonia)—benefit from the Cyrillic online presence as a bridge to Romanian-speaking culture.

Looking at the Wikipedia versions of different minority languages, it is often found that there is only a sparse number of rather short articles, so the value may be in the perception of the languages, but not in providing a useful encyclopaedia. In the case of Moldovan in the status described above, i.e., as a variety whose distance from standard Romanian is primarily defined by its script, the path of automatic transliteration provides the possibility of generating a comprehensive encyclopaedia on an *ad hoc* basis. While the previous section has explained how to improve the transliteration performance by computer, for a real encyclopaedia the suggestion could be made to freeze the transliterated version, to repair it manually, and to enrich it with individual content—although ethical and legal concerns may well be raised about the complete transfer of Wikipedia to Wikimoldia. Although Wikipedia content can theoretically be used freely as so-called open content and al-

though even the elaborately developed Wiki structure is freely available for the programming of independent variants, the transfer of the content without identifying the source is a plagiarism, and the replacement of *Wikipedia* by *Викимолдія* distorts the origin. But here as well, the genesis of the encyclopaedia could be made more transparent in a revision.

Finally, the question of the *cui bono* and thus the authorship of the Wikimoldia project can be raised. In the absence of precise indications, two quite contrary hypotheses on political motivation can be put forward: On the one hand, as a contribution to the vitalisation of the controversial Moldovan language, we could assume that the Transnistrian government and its post-Soviet continuation of the ideology of Moldovan “creat în laboratoarele Moscovei”¹³ (Cimpoeșu and Musteață, 2018, p. 236) is the responsible agent behind this. On the other hand, the exact opposite can be assumed, namely Wikimoldia as a liberally oriented project that creates access to Romanian-influenced content and thus a rapprochement with the rest of Moldova, Romania and Europe, which, through the direct transmission of standard Romanian, ultimately even expresses itself linguistically in the aforementioned *limbă pășărească* that can also be understood as a protest against socialism (cf. Gabinskij, 2002, p. 135). These are, however, free speculations that would be conceivable alongside an apolitical interpretation of Wikimoldia, too.

6. Conclusion

In this paper, the curiosity of a Moldovan phantom page of the Romanian Wikipedia has been treated in its technical functioning and potential use spotting the sociolinguistic background of the Moldovan language and its controversial interpretation. The most striking element of Moldovan, and thus Wikimoldia, is the use of the Cyrillic alphabet according to the standard developed in the Soviet Union in the 1920s, which provides Romanian with an orthography that is largely harmonious with Russian, but phonographically particularly flat.

Since Wikimoldia is no longer online, we could dismiss as idle thoughts the pronounced ideas about improving and elaborating the project into a functioning encyclopaedia. In fact, the foundations that have been laid remain available in the form of the PHP script `s1ava37md2`, which could be used to reconstruct the site. However, because it has not become clear whether and what political ideology was behind Wikimoldia, and also because of the aforementioned concerns about the theft of data and ideas, the continuation and expansion of Wikimoldia should

13. Translation (with reference to Moldovan): created in the laboratories of Moscow.

perhaps rather be avoided. Irrespective of this, the programming of Wikimoldia can serve as a lesson for other multialphabetic languages, such as Serbian and other Serbo-Croatian or BCMS varieties, or even Hindi and Urdu. Within the Romance language family, Jewish Spanish is particularly worthy of mention, because it is maintained in its own Wikipedia (<https://lad.wikipedia.org/>) in two languages with the Latin and Hebrew alphabets, where the inequality in the expansion of the encyclopaedia is immediately apparent: There are much fewer and mostly only shorter articles written in Hebrew. The option of an automatic transliteration could also be a helpful support in this context.

References

- Born, Joachim (2007). "Wikipedia. Darstellung und Chancen minoritärer romanischer Varietäten in einer virtuellen Enzyklopädie." In: *Sprachliche Diversität: Praktiken—Repräsentationen—Identitäten*. Ed. by Martin Doring, Dietmar Osthus, and Claudia Polzin-Haumann. Bonn: Romanistischer Verlag, pp. 173–189.
- Cimpoesu, Dorin and Sergiu Musteață (2018). *Basarabia la un secol de la Marea Unire. O istorie politică a Republicii Moldova. 1991–2018*. Târgoviște: Cetatea de Scaun.
- Coulmas, Florian (2018). *An Introduction to Multilingualism. Language in a Changing World*. Oxford: Oxford University Press.
- Dahmen, Wolfgang (2018). "Les frontières linguistiques extérieures du dacoroumain." In: *Manuel des frontières linguistiques dans la Roumanie*. Ed. by Christina Ossenkop and Otto Winkelmann. Berlin/Boston: De Gruyter, pp. 338–357.
- Gabinskij, Mark A. (2002). "Moldawisch." In: *Enzyklopädie des europäischen Ostens*. Ed. by Miloš Okuka. Klagenfurt: Wieser, pp. 133–143.
- Koch, Christian (forthcoming). "Wikimoldia – Викимолдия. Rekonstruktion eines Projekts zur Vitalisierung der moldauischen Sprache." In: *Korpus im Text* 10.
- Kramer, Johannes (1989). "Rumänisch: Graphetik und Graphemik." In: *Lexikon der Romanistischen Linguistik*. Ed. by Günter Holtus, Michael Metzeltin, and Christian Schmitt. Vol. III. Tübingen: Niemeyer, pp. 14–18.
- Leca-Tsiomis, Marie (2006). "Une tentative de conciliation entre ordre alphabétique et ordre encyclopédique." In: *Recherches sur Diderot et sur l'Encyclopédie* 40.41, pp. 55–66.
- Olariu, Florin-Teodor (2017). *Variație și varietăți în limba română. Studii de dialectologie și sociolingvistică*. Iași: Institutul European.
- Onu, Liviu (1989). "Rumänisch: Langue et écriture." In: *Lexikon der Romanistischen Linguistik*. Ed. by Günter Holtus, Michael Metzeltin, and Christian Schmitt. Vol. III. Tübingen: Niemeyer, pp. 305–324.

Zikmund, Hans (1996). "Transliteration." In: *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch*. Ed. by Hartmut Günther and Herbert Ernst Wiegand. Vol. 2. Berlin: De Gruyter Mouton, pp. 1591–1604.