

PARTICULARITĂȚILE DECHIRILIZĂRII LIMBII ROMÂNE

Valentina DEMIDOVA

Institutul de Matematică și Informatică al AȘM

În lucrare este studiată problema convertirii cuvintelor române din alfabetul chirilic în cel latin. Sunt stabilite și argumentate regulile de transliterare (*dechirilizare*) a cuvintelor din limba română, scrise în grafie chirilică, într-un echivalent scris cu grafie latină. O atenție deosebită este acordată cazurilor care prezintă ambiguități.

Cuvinte-cheie: transliterare, dechirilizare, cuvinte lemă, DEX.

PECULIARITIES OF DECYRILLIZATION FOR ROMANIAN LANGUAGE

The problem of Romanian words transliteration from Cyrillic to the Latin alphabet is studied. The transliteration rules for converting (*decyrillization*) of words in Romanian language, written in Cyrillic, to their equivalent written in Latin script are elaborated and argued. A special attention is paid to cases which present ambiguities.

Keywords: transliteration, decyrillization, canonic words, DEX

Introducere

Republica Moldova dispune de un bogat patrimoniu cultural sub formă de texte scrise cu caractere chirilice. Acest alfabet (sau alfabete) într-o formă sau alta a fost utilizat în diferite perioade istorice, cea mai vastă colecție de texte de diferită natură (literatură artistică, științifică, publicistică, acte oficiale etc.) fiind acumulată în perioada postbelică, adică din 1945 până în 1989, când Legea nr.3462 din 31 august 1989, adoptată de Parlamentul Republicii Moldova, reintroduce grafia latină ca normă oficială pentru scrierea textelor de limbă română în țara noastră.

Aliniindu-se la politicile promovate de Uniunea Europeană prin Agenda Digitală pentru Europa, Republica Moldova a adoptat strategia „Moldova digitală 2020”, planul de acțiuni al căreia prevede digitizarea întregului patrimoniu cultural. Ar fi firesc ca realizarea acestui deziderat să fie însoțită de transliterarea textelor scrise cu grafie chirilică în cea latină, făcând acest strat de valori culturale accesibil pentru un cerc cât mai larg de vorbitori de limbă română.

În conformitate cu definiția din DEX online (<http://dexonline.ro/>) prin **transliterație** (sau **transliterare**) se subînțelege procesul de „transpunere, transcriere dintr-un alfabet în altul a unui text scris, redându-se literele prin echivalentele lor din alfabetul în care se transcrie, fără a se ține seama de eventualele deosebiri de ordin fonetic dintre semnele celor două alfabete”. Dicționarul explicativ Merriam-Webster (<http://www.merriam-webster.com/dictionary/>) operează cu o definiție mai simplă, anume: „Scrierea cuvintelor sau a literelor cu caracterele altui alfabet”. În cazul nostru este important ca în urma aplicării procedurii de transliterare să se obțină un text ortografic corect, deci vom vorbi nu despre o transliterație simplă, ci despre un proces mai complicat, pe care îl vom numi *dechirilizare*. Problema dechirilizării a fost abordată anterior în lucrările [1,2]. În cele ce urmează vom descrie și fundamenta un set de reguli de transliterare, care pot constitui baza unui algoritm ce va realiza acest proces în mod automat. Regulile propuse sunt racordate la alfabetul și ortografia din perioada 1945-1989.

Descrierea regulilor de transliterare (dechirilizare)

La elaborarea și descrierea regulilor de transliterare a cuvintelor din limba română, scrise în grafie chirilică, într-un echivalent scris în grafia latină ne-am condus de normele gramaticale în vigoare. Vom identifica două tipuri de reguli: elementare și complexe.

Alfabetul chirilic contemporan pentru limba română conține 31 litere. În conformitate cu regulile elementare se efectuează transliterarea a 26 litere. Rămân 5 litere care necesită o procesare suplimentară, și anume: “z”, “k”, “q”, “b”, “я”. \ЪХЗЩШГГЪХ=-098

Regulile elementare sunt cele care pun în corespondență independentă de context fiecărei litere chirilice una sau mai multe litere din grafia latină. Următoarele litere sunt traduse în baza regulilor elementare (Fig.1):

Chirilic	<i>а б в д е ж ж з и ѝ л м н о п р с т у ф х ц ш ь э ю</i>
Latin	<i>a b v d e j g z i i l m n o p r s t u f h ț ș i i u</i>

Fig.1. Regulile elementare de transliterare.

Restul conversiilor depind de context, procesul de dechirilizare pentru aceste cazuri îl vom prezenta în tabelele 1-10.

Tabelul 1

Reguli pentru litera "z"

z	gh	înainte de: e, u, я, ю, ь
	g	în restul cazurilor

Exemple: гем – ghem, гиочел – ghiocel, агязмэ – agheazmă, грек – grec.

Tabelul 2

Reguli pentru litera "κ"

κ	ch	înainte de: e, u, ю, я
κс	x	
κ	c	în restul cazurilor

Exemple: кел – chel, химие – chimie, кяр – chiar, комун – comun, ксерокс → xerox.

Tabelul 3

Reguli pentru litera "ч"

		context
	c	înainte de: e, u, ь
	ce	înainte de a
ч	ci	înainte de o, y
	ci	la sfârșit de cuvânt
	ci	înainte de consoană

Exemple: чай – ceai, черк – cerc, чуботэ – ciubotă, чинчъ – cinci, мичман – miciman

Reguli de dezambiguizare

Un caz mai complicat prezintă transliterarea literei "ы", care trece în "î" sau în "â" în dependență de mai mulți factori. În expunerea noastră ne-am racordat la recomandările Academiei Române pentru scrierea literelor "î" și "â".

1. Se substituie "ы" prin "î" în cazurile poziționării lui:

- la început de cuvânt: *în, împăca, înainte, încă, înflori, inger* etc.;
- la sfârșit de cuvânt: *coborî, hotărî, ocărî, urî* etc.; *imambaialdî, îhî*.
- în cuvintele care sunt formate cu prefixele *re-, ne-, pre-, supra-* plus cuvântul care începe cu litera î: *neînsemnat, preîncălzire, supraîncălca* etc.

2. Se scrie "â" în următoarele cazuri:

- în interiorul cuvântului: *bâlci, când, fân, gând, lână, mânca și mâncând, pâlc, râu, sân, tângui, vânt* etc.;
- în numele propriu România și în cuvintele din familia lui: *român, românecă, româncuță, românesc, românește, Românie, românime, românism, românist, românistică, româniza, românofil; aromân, dacoromân, istroromân, meglenoromân*.

Un caz deosebit de complicat prezintă procesul de transliterare a literei "я". Tabelul 4 ilustrează trei cazuri posibile de transliterare a acestei litere:

Tabelul 4

Reguli pentru litera "я"

я	ia
я	ea
ия	ia

În pofida faptului că pentru această literă avem un număr mic de variante de transliterare, determinarea alegerii acestor variante prezintă o problemă neașteptat de complicată, deoarece este dificil a stabili algoritmic regula de transliterare. În special această problemă apare în ocurența lui "я" în interiorul cuvântului. Printr-o analiză a cuvintelor lemă din DEX am constatat că în interiorul cuvântului putem avea atât *я – ea* (5460 rezultate), cât și *я – ia* (6004 rezultate).

Exemple:

agheasmă | agoniseală | albineală | ariceală | armean | amorteală | aseamăna | seară | asudeală
агызмэ агонисеалэ албинялэ аричялэ армян аморцялэ асямэнэ сярэ асудялэ
aiială | baiian | băiat | boială | chiag | cian | chișinăuian | chiamăt | chiang | boiamă |
ауялэ баян бэят боялэ кяг чян кишинэуян кямэт кянг боямэ

Așadar, exemplele de mai sus demonstrează că problema transliterării literei "я" nu este deloc trivială. A fost efectuat un studiu folosind DEX online pentru a evidenția anumite reguli de deschirire a literei "я". În rezultatul acestui studiu s-a constatat că totuși putem evidenția anumite legități, și anume:

- La început de cuvânt putem avea atât **я – ea** (18 rezultate), cât și **я – ia** (281 rezultate).

Tabelul 5

Frecvența îmbinărilor „ea” și „ia” la început de cuvânt

combinație	num. de cuvinte canonice din DEX	regulă de transliterare
„ea”	18	я – ea
„ia”	281	я – ia

Prin urmare, dacă cuvântul începe cu litera "я", se aplică regula **я – ia**, cu excepția listei din cele 18 cuvinte de mai jos, în care se aplică regula **я – ea**:

eac | Eac | eacidul | eagra | Eakins | Eam | Eames | ean | Eanes | Earhart | earl | ească | East | Eastman | eastmancolor | Eastbourne | Eastwood | eax

Din câte se poate de observat, aceste cuvinte în majoritatea cazurilor sunt împrumutate din alte limbi.

- La sfârșit de cuvânt de asemenea distingem mai multe cazuri, și anume:

1. **ия – ia** (1760 rezultate)

Exemple:

academia | Armenia | Bacovia | devia | filia | întârzia | media | mafia | Magnezia | maia
академия Армения Баковия девия филия ынтырзия медия мафия Магnezия мая

2. **ея – eia** (57 rezultate)

Exemple:

a femeia | Alazeia | ambreia | ambuteia | aquileia | ardeia | ateia | baleia | bureia | căreia | cărceia | ceia |
cheia | chirieia | choreia | cîrceia | codeia | correia | daseia | debleia | debreia | depareia | descheia |
descleia | dezambreia | echeia | filofteia | Filofteia | forbeia | Galateia | Gamaleia | graseia | încârceia |
încheia | încîrceia | înclieia | înclieia | încondeia | îndumnezeia | întemeia | miheia | Nedeia | oleia | osteia |
ovreia | paideia | Peneia | pereia | phorbeia | rambleia | redescheia | reîncheia | reîntemeia | scânteia |
temeia | uleia | usteia

3. **я – ea** ("я" după consoană) (1625 rezultate)

Exemple:

abea | abîinea | acătărea | acolea | adesea | adicălea | adicăteia | adineaurea | aevea | a finea |
абя абциня акэтэря аcoled адеся адикэля адикэтея адиняуря аевя афиня

Astfel, dacă litera "я" se află la sfârșit de cuvânt, în dependență de context se aplică una din cele trei reguli descrise mai sus.

De asemenea a fost efectuată o analiză a cuvintelor lemă din DEX online, care conțin combinațiile „ea” și „ia” în interiorul cuvântului, exceptând cazurile descrise mai sus. Au fost revizuite îmbinările cu toate literele alfabetului român. În rezultatul acestei analize s-a constatat că pentru anumite litere **a, e, ф, y, ж, к (a, e, f, u, j, k)**, care precedă litera "я", putem găsi unele legități ce ne permit aplicarea anumitor reguli de transliterare, și anume:

Combinația „aea” se întâlnește în 33 de intrări din DEX, iar combinația „aia” – în 509 cuvinte:

Tabelul 6

Frecvența îmbinărilor „ea” și „ia”, precedate de „a”

combinație	num. de cuvinte canonice din DEX	regulă de transliterare
„aea”	33	ая – aia
„aia”	509	ая – aia

Așadar, se aplică regula **ая – aia**, cu excepția listei din 33 de cuvinte de mai jos, pentru care se aplică regula **ая – aia**:

actaea | aea | aeacus | aeaea | althaea | astraea | capraea | cărăeală | ciorovăeală | circaea | ytheraea | dionaea | erithraea | europaea | gaea | gărăeală | graeae | hydralernaea | idaea | jacobaea | lathraea | lăscăeață | lernaea | maeander | maeandrus | măeastră | nymphea | pangaea | pospăeală | puhăeală | răgăeală | răsgăeală | spirae

Combi-nația „**fea**” se întâlnește în 16 cuvinte lemă din DEX, iar combinația „**fia**” – în 197 de cuvinte:

Tabelul 7

Frecvența îmbinărilor „ea” și „ia”, precedate de „f”

„ fea ”	16	фя – fea
„ fia ”	197	фя – fia

Regula aplicată este **фя – fia**, cu excepția listei din 16 cuvinte, în care se aplică regula **фя – fea**:

bârfeală | cafea espresso | corcofeală | corcofeală | feaci | feacieni | feamen | feară | feată | filozofeală | fofează | fofează | fufează | nimfeacee | sfeat | tufean

Combi-nația „**uea**” se întâlnește în 317 cuvinte din DEX și „**uia**” – în 39 de cuvinte:

Tabelul 8

Frecvența îmbinărilor „ea” și „ia”, precedate de „u”

„ uea ”	317	уя – uia
„ uia ”	39	уя – uea

În acest caz putem aplica regula **уя – uea**. Excepție constituie lista din 39 de cuvinte, în care se aplică regula **уя – uia**:

aueală | bănueală | bântueală | bruftueală | căpătueală | cerueală | cheltueală | chibzueală | ciurmueală | făgădueală | fălălueală | futueală | gerueală | hăitueală | hărțueală | hurueală | îngădueală | înghesueală | mântueală | murueală | mustueală | neorândueală | orândueală | părădueală | părueală | păsueală | pârğueală | răbueală | răfueală | rândueală | sfânțueală | sghihueală | târgueală | târnueală | tehueală | tencueală | ulueală | urlueală | zeciueală

De asemenea, au fost analizate intrările din DEX online care conțin combinațiile „ea” și „ia” urmate de o literă. S-a constatat că pentru literele **x, u, κ, n, y (h, i, k, p, u)**, care urmează după litera „я”, la fel putem stabili anumite reguli de transliterare. Un caz deosebit de interesant prezintă combinația „яn”.

- Combi-nația „**eap**” se întâlnește în 37 de cuvinte din DEX, iar combinația „**iap**” – în 129 de cuvinte.

Tabelul 9

Frecvența îmbinărilor „ea” și „ia”, urmate de „p”

„ eap ”	37	яп – eap
„ iap ”	129	яп – iap

Regula de bază este **яп – eap**, iar 37 de presupuse “excepții” pentru **яп – iap** sunt:

antiapartheid | antiapoplectic | biapiculat | diapauză | diapazon | diapedeză | diapente | diapeutică | diampir | diampirism | diaplegie | diaporamă | diapozitiv | diapozitivare | diaproiector | diaproiecție | disdiapazon | disdiapazon | epidiapente | epidiaproiector | hemidiapente | hipodiapente | Лиарупнов | miniaparat | miniaparatură | miriapod | periapendicită | periapendicular | periapical | priapee | priapesc | priapeu | priapic | priapism | subdiapente

Analizând această listă observăm că practic toate aceste excepții sunt false, deoarece în scrierea cu alfabetul chirilic nu intră litera “я”, ci îmbinarea de litere “**ua**”, precum în cuvântul “*диазон*”. Unica excepție reală este substantivul propriu “*Лиарупнов*” (*Liарупнов*).

- Combi-nația „**eau**” se întâlnește în 112 cuvinte din DEX, iar combinația „**iau**” o găsim în 33 de cuvinte.

Tabelul 10

Frecvența îmbinărilor „ea” și „ia”, urmate de „u”

„ <i>eau</i> ”	112	<i>яу – eau</i>
„ <i>iau</i> ”	33	<i>яу – iau</i>

Deci, putem aplica regula *яу – iau*, excepție constituind lista din 33 de cuvinte de mai jos, în care se aplică regula *яу – eau*:

chasseriau | chiau | chiau-chiau | chiaun | diaulă | diaulos | fabliu | Gaboriau | ghiaur | Ghiaurov | miau | miau-miau | miaul | miaun | miauna | miaut | miniautobuz | miniautodrom | miniautomobil | miniautoturism | nemaiauzit | piaui | riau | schiaun | schiauna | schiaunare | semiautomat | semiautomatic | semiautopropulsat | semiautopropulsie | semiauxiliar | Siauliai | șchiau

De asemenea, putem evidenția anumite reguli de transliterare a literii “я”, atunci când “я” face parte din unele sufixe, și anume: *ялэ (eală), яцэ (eață), яскэ (ească)*:

Se aplică regula *я – ea*, după consoane, în sufixele:

- *-ea* (f. sg. de la -el): *фрумушя (frumușea), кэця (cățea), виця (vițea)*;
- *-eală*: *бетежялэ (betejeală), мынжялэ (mînjeală), обложялэ (oblojeală), прăжеалă; căpтуșeală, greșeală, învălmășeală, moleșeală, sfirșeală, зăпуșeală; agoniseală, oboseală, procopseală; umezeală, urzeală; ațejeală, iuțeală; amăreală, mohoreală, oțăreală;*
- *-eață*: *рошыцэ (roșeață)*;
- *-ească* (f. sg. de la -esc): *витежяскэ (vitejească); остэшыяскэ (ostășească), рэзэшыяскэ (răzeșească), стрэмошыяскэ (strătoșească); rusească, săsească; englezească, franțuzească; băiețească, frățească etc.*
- De asemenea, se aplică regula *я – ea* în sufixul de apartenență locală – *ян (ean)* după majoritatea consoanelor, inclusiv *ж, ш, ц, з, в (j, ș ț, z, v)*: *clujean, doljean, gorjean, lugojean; argeșean, crișean, făgărășean, hușean, ieșean, maramureșean, orășean, oșean; bănățean, cetățean, gălățean, județean; hurezean, vârlizean; brașovean, lipovean, moldovean, prahovean.*

Concluzii

Regulile stabilite pot servi drept bază pentru elaborarea algoritmului de dechirilizare a limbii române. Constatăm totuși că caracterul specific al limbii nu permite formalizarea completă a procesului de transliterare.

În rezultatul acestui studiu s-a constatat că, în pofida faptului că există foarte puține litere în alfabetul chirilic (“*ы*” și în special litera “*я*”) care prezintă ambiguități la transliterare și a numărului relativ mic al variantelor de transliterare, determinarea alegerii acestor variante prezintă o problemă neașteptat de complicată, chiar și pentru cunoscători de limbă română.

Bibliografie:

1. COJOCARU, S., BOIAN, E., CIUBOTARU, C., COLESNICOV, A., DEMIDOVA, V., MALAHOV, L. *Regeneration of printed cultural heritage: challenges and technologies*. Chișinău: The Third Conference of Matematical Society of the Republic of Moldova, 19-23 August, 2014, p.481-489.
2. DEMIDOVA, V. *Particular Aspects of the Cyrillization Problem*. Chișinău: The Third Conference of Matematical Society of the Republic of Moldova, 19-23 August, 2014, p.493-498.

Prezentat la 16.08.2015