- **09:15-09:30: Roxana Patras**

Senior Researcher II/ Cercetator stiintific gr. II (CS II)
Institute for Interdisciplinary Research
"Alexandru Ioan Cuza" University of Iasi

**Introduction to Romanian novels/literary context
Relevant to the Action's goal (e.g. focus on previously non preferred periods)**

COST-ELTeC / **ELTeC-rom**

<> Code   ⓘ Issues 0   �ⁿ Pull requests 0   ▥ Projects 0   ▦ Wiki   ◫ Security   �ⅰⅰ Insights

Branch: master ▾   **ELTeC-rom** / **level1** /

Create new file | Upload files | Find file | History

RoxanaPatras Add files via upload   Latest commit a017818 11 days ago

..

📄 ROM001_Anonim_MCP_ROSCAN_HAIDUCUL.xml	Add files via upload	17 days ago
📄 ROM002_BalanescuSimion_BLESTEMUL.xml	reapply fix	17 days ago
📄 ROM003_BalanescuSimion_SFARSITUL_BLESTEMULUI.xml	Add files via upload	17 days ago
📄 ROM004_DumbravaBucura_HAIDUCUL.xml	Add files via upload	17 days ago
📄 ROM005_MacriPanait_GHITA_CATANUTA.xml	Add files via upload	17 days ago
📄 ROM006_PopescuND_IancuJianuZapciu.xml	Add files via upload	17 days ago
📄 ROM007_PopescuND_IancuJianuCapitan.xml	Add files via upload	17 days ago
📄 ROM008_PopescuND_BUJOR_HAIDUCUL.xml	Add files via upload	17 days ago
📄 ROM009_MacedonskiA_Thalassa.xml	Add files via upload	17 daays ago
📄 ROM010_VlahutaA_Dan.xml	Add files via upload	17 days ago
📄 ROM011_GrandeaG_Fulga.xml	Add files via upload	17 days ago
📄 ROM012_GrandeaG_Vlasia.xml	Add files via upload	17 days ago
📄 ROM013_IonescuR_LaGuraSobei.xml	Add files via upload	17 days ago
📄 ROM014_IonescuR_CatastihulAmorului.xml	Add files via upload	17 days ago
📄 ROM015_IonescuR_DonJuaniiBucuresti.xml	Add files via upload	17 days ago
📄 ROM016_BolintineanuD_Manoil.xml	Add files via upload	17 days ago
📄 ROM017_BolintineanuD_Elena.xml	Add files via upload	17 days ago
📄 ROM018_BolintineanuD_DoritoriiNebuni.xml	Add files via upload	17 days ago
📄 ROM019_VDemetrius_MateiDumbarau.xml	Add files via upload	13 days ago
📄 ROM020_VDemetrius_OrasulBucuriei.xml	Add files via upload	13 days ago
📄 ROM021_VDemetrius_PacatulRabinului.xml	Add files via upload	12 days ago
📄 ROM022_PanaitMacri_HaiduculTandura.xml	Add files via upload	11 days ago
📄 ROM023_Ighell_TilharulFulger.xml	Add files via upload	11 days ago
📄 ROM024_DumbravaB_Pandurul.xml	Add files via upload	13 days ago
📄 ROM025_Bujoreanu_MistereDinBucuresti.xml	Add files via upload	12 days ago
📄 ROM026_MacriP_Bostan.xml	Add files via upload	11 days ago

ELTeC Summary Page

As well as the following summary statistics, this page provides links to human-readable versions of each text currently included in the European Literary Text Collection (ELTeC). Click on a language code in the table below to see a list of texts now available in that language. Then click on the identifier of a text to see a simple rendering of the text as produced by CETEIcean. The original source files are stored in a GitHub repository at COST-ELTeC, and may be downloaded freely from there.

Please note : this is a work in progress! Comments and reports of any problems are much appreciated: send them to the WG1 Issue Tracker.

Language	Texts	Words	Male	Female	Short	Medium	Long	1840-59	1860-79	1880-99	1900-20
cze	23	692936	21	2	20	3	0	8	9	6	0
deu	44	8422194	27	17	13	16	15	13	8	14	9
eng	91	11990064	46	45	17	25	49	20	21	25	25
fra	92	6939456	57	35	25	45	22	16	31	32	13
gre	11	42524	10	1	11	0	0	0	1	6	4
hun	102	7641508	86	16	46	33	23	24	24	25	29
ita	34	3328244	32	2	13	10	11	5	12	10	7
nor	27	1114092	22	5	18	9	0	2	2	19	4
por	57	3685825	46	11	27	20	10	8	18	13	18
rom	26	1196258	23	2	17	8	1	1	9	10	6
slv	72	3894267	65	7	37	32	3	0	4	30	38
spa	19	2022041	15	4	4	11	4	5	10	3	1
srp	29	1087455	25	4	20	9	0	0	1	12	16

Last updated: 2019-10-03

```xml
<head>CATASTIHUL AMORULUI</head>
<head>O PREFAŢĂ CARE NU ESTE PREFAŢĂ</head>
<div><p>Scena se petrece în capul servitorelui dumneavoastră.</p>
<p>La dreapta şi la stînga paradoxe, reflecţiuni, proiecte, fragmente începute1, suveniri, speranţe, căinţe.</p>
<p>în scurt nişte mobile de simpli creeri, dar de ajuns.</p>
<p>Aşezat înaintea unei mese, eu privesc cu un aer mulţumit un splendid soare care se joacă printre sticlele ferestrei mele.</p>
<p>În acest timp, următorul dialog se angajează între oaspeţii creerilor mei, al căror scurt inventar avui plăcere a vi-l aşterne mai sus.</p></div>

    <div><label>Lenea</label>
<p>Ce frumoasă, ce veselă zi! Bătrînii arbori din Băneasa cată să aibă în această dimineaţă reflecte de sărbătoare... Ia-ţi pălăria, bastonul şi vino cu mine, tu.</p></div>

<div><label>Raţiunea</label>
<p>N-o asculta! Editorele tău aşteaptă, chiar în această dimineaţă, primele făscioare ale cărţii ce tu i-ai promis. Au să mai fie zile frumoase! N-o asculta!</p></div>

    <div><label>Lenea</label>
<p>Mîine va ploua, — şi aşa două lune necurmat... Ce mai lucru, cînd vei înegri cîteva pagine din o hîr-tie neofensivă; nu-ţi este ruşine să preferi frumuseţile îndoioase ale stilului tău, splendorilor strălucitoare ale naturei. Eu d-aci văd eîmpiile în cari înverzeşte griul nou. Masa e pusă la birt. Cotleta cîntă, frigîndu-se. Hossanah al lăcomiei! Vinul rîde în vesela butelie... Să plecăm, îţi zic, şi dă dracului secătura ta de volum!</p></div>

<div><label>Vanitatea</label>
<p>Secătură! ... secătură! ... Lesne de zis.</p></div>

<div><label>Raţiunea</label>
```

Cultural context (1840-1920): traditional periodization vs. ELTeC sampling principles

1840-1859 (48 revolution)	1860-1866 (post-48 revolution) 1866-1880 (Junimea age)	1880-1900 (Literatorul + Contemporanul)	1900-1918 (Samanatorul + Viata Romaneasca) 1905-1920 (Belle Epoque/Simbolism)
ELTeC t1 = 1840-1859	**ELTeC t2=1860-1879**	**ELTeC t3=1880-1899**	**ELTeC t4=1900-1920**
-National, unionist movement: Wallachia and Moldavia which were semi-autonomous provinces (Ottoman Empire and other empires interested in the geo-strategic position of the 2 principalities) -UNION of 1859 (national culture and literature = Romanian language)	-Romanian prince (1859-1866) -German prince (starting in 1866) parliamentarianism; state modernization; rural to city life -The war of independence in 1877	-proclamation of the Romanian kingdom (German dynasty) -rise of socialism	-the peasants' riot in 1907 -national movement (union with Transylvania)

1840-1859 (48 revolution)	1860-1866 (post-48 revolution) 1866-1880 (Junimea age)	1880-1900 (Literatorul + Contemporanul)	1900-1918 (Samanatorul + Viata Romaneasca) 1905-1920 (Belle Epoque/Simbolism)
ELTeC t1 = 1840-1859	ELTeC t2=1860-1879	ELTeC t3=1880-1899	ELTeC t4=1900-1920
Greek influence vs. intelligentsia educated in France and Germany brought in progressist views which culminated in the 1848 Revolution (late Enlightenment + Romanticism)	-Imports from Western literary traditions (forms, styles, manners) -The obsession of the language's *genius* + national literature (original) -aestheticism	-a socially-oriented literature -modernism	-a nationally oriented literature -aesthecization of daily life

1840-1859 (48 revolution)	1860-1866 (post-48 revolution) 1866-1880 (Junimea age)	1880-1900 (Literatorul + Contemporanul)	1900-1918 (Samanatorul + Viata Romaneasca) 1905-1920 (Belle Epoque/Simbolism)
ELTeC t1 = 1840-1859	ELTeC t2=1860-1879	ELTeC t3=1880-1899	ELTeC t4=1900-1920
Print culture: religious and lay texts in Romanian but printed in the Cyrilic alphabet; transition alphabet or Latin alphabet	Print culture: dominated by lay texts in Romanian but printed in the transition alphabet or Latin alphabet	Print culture: lay texts printed in non-standardized variants of Romanian (lack of consistent norms)	Print culture: lay texts printed in non-standardized variants of Romanian (norms)

"Birth certificates" of Romanian novels

INSTALMENT: *Pustnicul* by GRAF Valberg [Mihail KOGĂLNICEANU], 1844, incomplete novel

VOLUME: *Elvira sau amorul fără de sfârşit. Romans original* by D.F.B., 85 p, 1845, complete but too short novel

The Romanian Novel in figures relevant for ELTeC sampling (1840-1920)

- 350 novels published in volumes

- 279 novels published in instalments

Note that current data on Romanian novels also include items written in other languages (French, German) by "Romanian" authors, then imported into Romanian as a new version or translation.

From 350 volume-novels:

- 37 are written by women (10.5%)

- Between 20 and 30 have been subject to constant reprint/ canonization (5.6%-8.5%)

- Around 100 novels are adventure novels (hajduk + city mysteries), thus short novels (28%)

1/3 of the Romanian novels published in volumes should go into ELTeC (100 novels per each collection)

Is that possible?

09:50-10:20: Roxana Patras

Romanian Language collection.
Specific characteristics and challenges for selecting and digitizing novels

How to... Build a National novel collection?
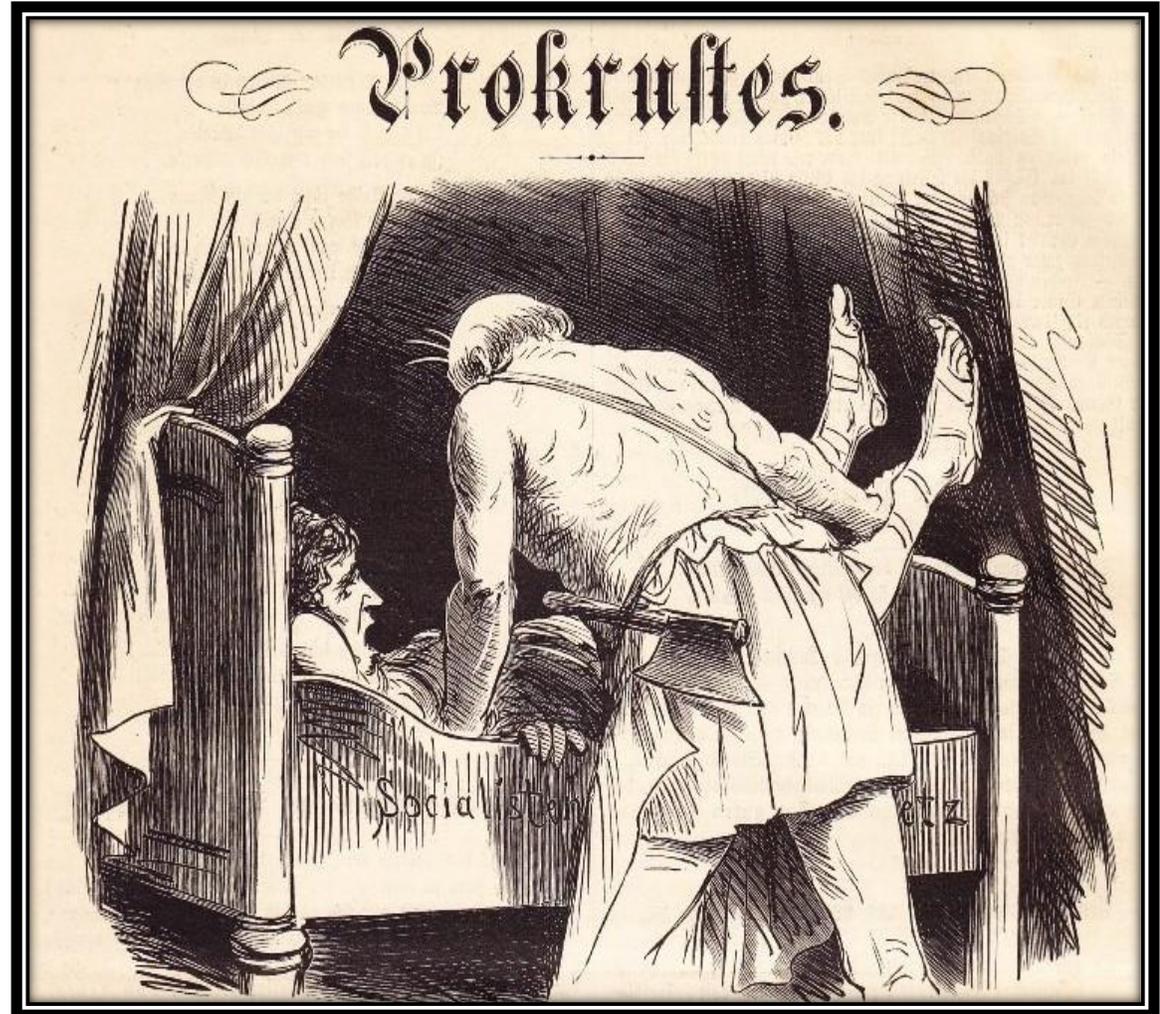
Drawback 1: DIGITIZATION from scratch

The Romanian Collection of ELTeC starts from facsimiles:

➢we scan (very fragile) books

➢OCR (when possible)

➢convert them into xml (automatically when scripts are available, manually when scripts aren't really reliable)

➢A toilsome process of manual cleaning

➢Encode files into TEI

Samples I: printing popular books in 19th-century Romania

Toți cu toții vědură viind in góna mare spre locul de osândă o ființă
omenéscă, călare pe un cal alb care agita mereŭ o batistă și striga în-
tr'una : Stați stați.
Pag. 138.

Groasnicii Banditii
Lache și Bâbean

Voinea întinse mânele, o apucă repede cu o mână peste mijloc, și o
aruncă cu totul leșinată d'acurmezișul pe cal.
Pag. 32.
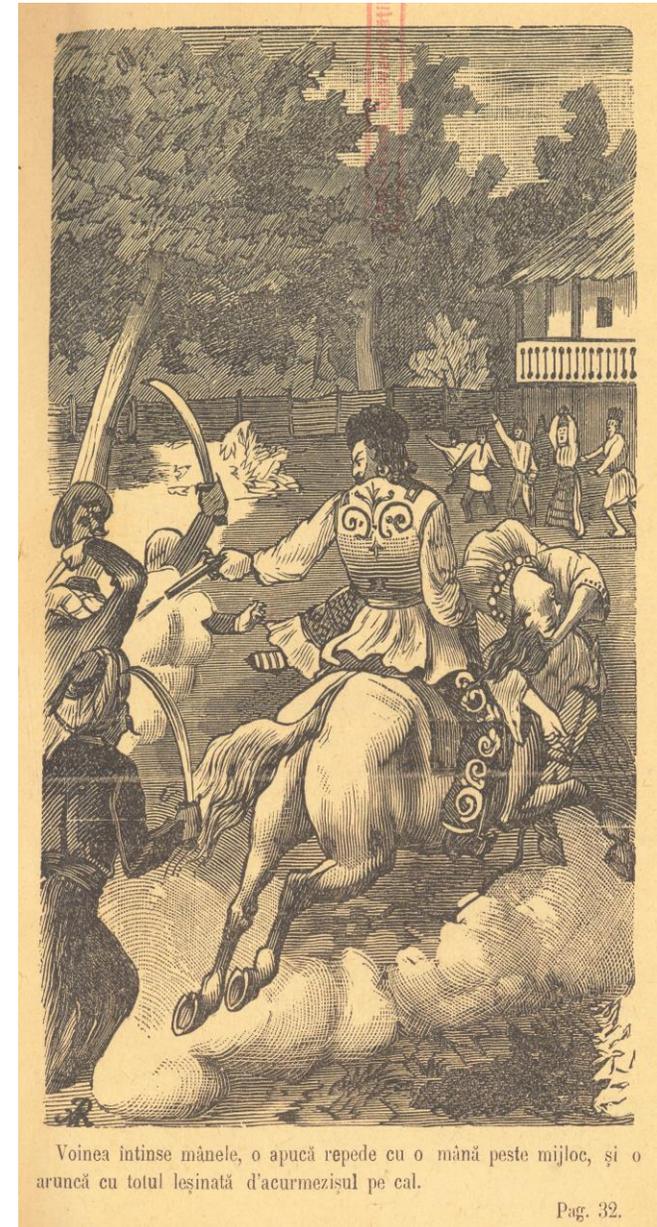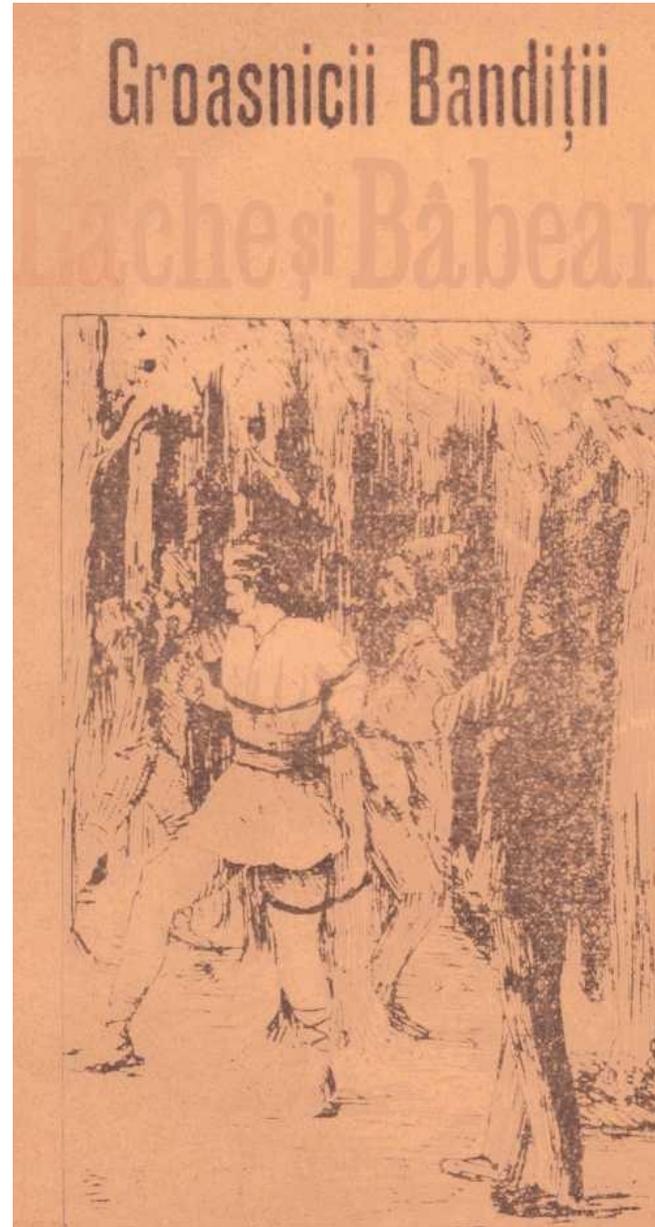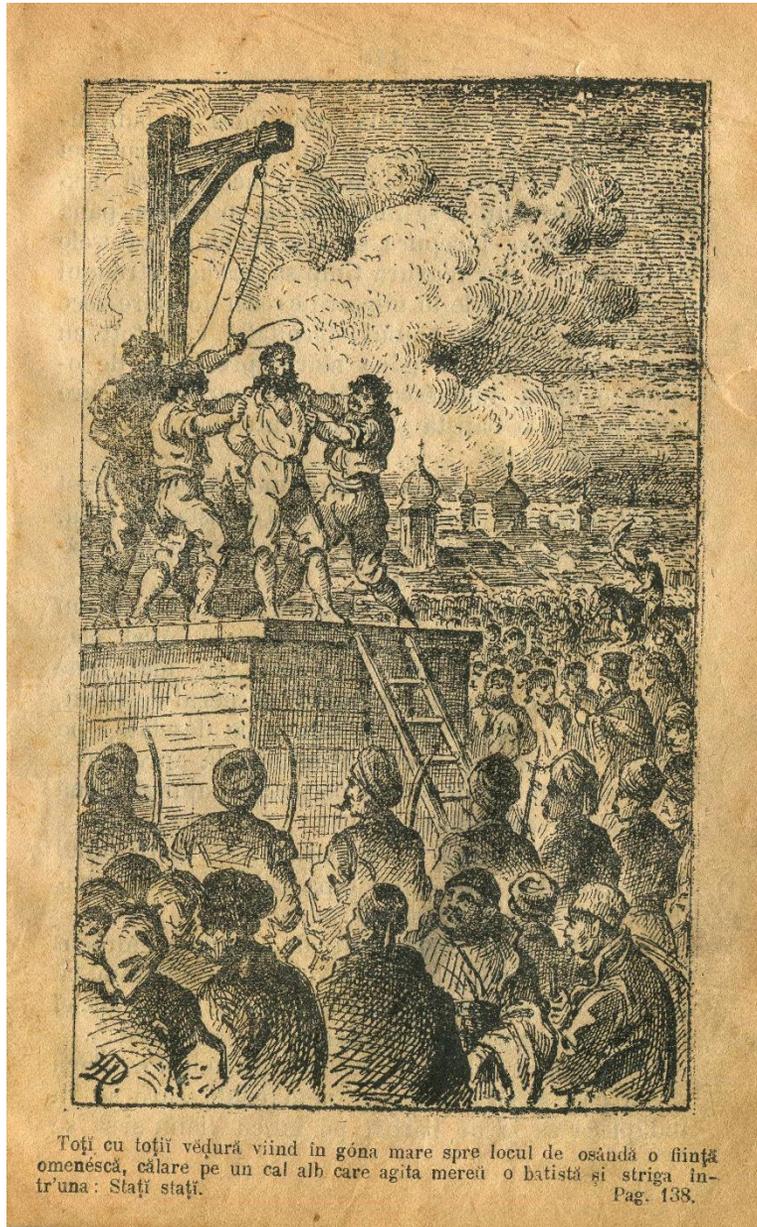
Samples II: take a close look at the glyphs
How does OCR perform on this type of print?

Crima misteriósă din calea Mogoșóiei

Siluirea

E nópte.

Să ningă încetase și cerul ca și pămêntul părea înghețat.

E nópte, târḑiŭ de tot și cu tóte astea într'una din casele din calea Mogoșóiei luminarea ardea încă.

Casa are un aspect frumos, ea e văpsita cu galben și zidăria de și e destul de veche, cu tóte astea o reparație făcută de curênd îi dă o aparență solidă și frumósă.

Luminarea ardea într'o cameră din etagiul al doilea; afară de acéstă camere tot restul apartamentuluĭ era în întuneric și tăcere.

Afară domnea un frig teribil și zăpada era mare, căcĭ tótă ḑiua ninsese. In timpul acesta

Măndria luĭ îl stăpânea și la cazarmă și nu putea să sufere cea maĭ mică observație a căpitanuluĭ Lambru, ci respundea cu obrăsnicie, din care causă era rěŭ persecutat. Persecuțiile deveniră atât de amare pentru Dragoș, în cât îșĭ jura să deserteze și să'șĭ rěsbune maĭ târḑiŭ pe căpitanul Lambru.

Cât timp a stat în cazarmă, nu legase prietenie cu nimenea de cât cu un óre-care Ión Emanoil, supranumit Mână lungă, care se asemăna cu Dragoș în privința caracteruluĭ, ca douě picăturĭ de apă. Intre ómenĭ de acelaș caracter, învoeala se face în doue vorbe. Se hotărâsera decĭ să dezerteze amândoĭ de odată, și de óre-ce eraŭ istețĭ, după cum am maĭ spus, nu fuse trébă mare să găséscă un mijloc bun și iata'ĭ decĭ pe amândoĭ liberĭ, înarmațĭ până'n dințĭ și pregătițĭ cu haine și rufe pe carĭ avuseseră grije să și le cumpere maĭ din'nainte și luând drumul către pădurea Nicorescĭ.

Finele Prologuluĭ

OCR output: really untidy

File Edit Search Markers Folding View Utilities Macros Plugins Help

PanaitP_Dragos2.htm (%USERPROFILE%\Desktop\CORPUSURI\HAJDUKS_PDFsearchable\de convertit cu abby\)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"
    "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head><meta http-equiv="content-type" content="text/html; charset=UTF-8"/><meta name="generator" content="ABBYY FineReader 12"/><link rel="stylesheet" href="Banditul Dragoş_file
</head>
<body><img src="Banditul Dragoş_files/Banditul Drago19-1.jpg" style="width:330pt;height:483pt;"/>
<p><span class="font14">PRBTU I- </span><span class="font9" style="font-weight:bold;">l</span><span class="font14"> LEU</span></p><div><img src="Banditul Dragoş_files/Banditul D
<p><span class="font9" style="font-style:italic;">s+iţK</span><span class="font14">    ,rv*v sr^ .s*^ .s*^s</span></p></div><br clear="all"/><div><img src="Bandit
<p><span class="font14">BUCURESCI</span></p>
<p><span class="font14">&lt; 'oneurenţa Sir. De^el^;</span></p></div><br clear="all"/><div>
<p><span class="font14">K^e&lt;-,\&lt;</span></p>
<p><span class="font9" style="font-style:italic;">vi'.Kjf.</span><span class="font14">    W^ ^    C^i\^T C^A^f Ov' ,</span></p></div><br clear
<p><span class="font14">Satul Nicoresci. este unul din satele cele mai prospere din tuta România, şi de aceea ţfiranii din el nu duc lipsa, ci au tot-d'a-una cele nece
<p><span class="font14">Acum, când Dragoş avea 22 ani, toate apucăturile rele din copilărie luaseră o des-voltare enormă şi bătrînul seu tată se întreba adesea dând cu neîncrede
<p><span class="fonţ14">A trecut un ăn. Bătrînul tată al lu Dragoş s'a dus să'şl regăsească nevasta în ceruri. Dragoş e acum în miliţie, l-şise la sorţi. II trecură întâiu la că
<p><span class="font14s">Mândria lut il stăpânea şi la cazarmă şi nu putea să sufere cea mal mică observaţie a căpitanului Lambru, ci respundea cuobrăs-nicie, din care causă
<p><span class="font14s">Cât timp a stat în cazarma, nu legase prietenie cu nimenea de cât cu un 6re-care I6n Rtnanoil, supranumit Mână lungă, care se asemăna cu Dragoş
<p><span class="font10" style="font-variant:small-caps;">capitolui î.</span></p><h5><a name="bookmark2"></a><span class="font11" style="font-weight:bold;">Jefuirea arendaşului</
<p><span class="font14">*</span></p>
<p><span class="font14">La marginea pădure! NieorescT se afla un han ţinut de un bulgar. Deşi mic, hanul nu era mai nici odatA ocupat pe deplin,, căci călătorii în par
<p><span class="font14">Dragoş gîise tovarăşului seu, Ion </span><span class="font9" style="font-weight:bold;">l</span><span class="font14"> inanoil, arătându'i hanul :</span></
<p><span class="font14">Mi se pare că vom avea aci de lumi. L timpul culesului viilor şi hanul treime să fie îndopat, de calatori. Lână una alta mc duc să v&lt; ghez la &lt
<p><span class="font14">cunoscuta numai de el, fiind c.i el singur chiar o făcuse din copilăria lui şi care scurta cu atâta drumul, în cât după cate-va minute se găsea
<p><span class="font14">Trecu cinci minute de aşteptare, apoi dece, şi nici un om nu se darea.</span></p>
<p><span class="font14">Dragoş începu sa se plictisescă şi se hotară chiar sa amae lucrarea pe a doua d'&gt; cu atât ca şi începuse a se însera. Eşi deci cu băgare de
<p><span class="font14">Dragoş </span><span class="font9" style="font-weight:bold;">¡l</span><span class="font14"> examina cu atenţiune. Intune-recul facea sa nu i se vadă faţa.
<p><span class="font14" style="font-weight:bold;">— Iată </span><span class="font14">pe jupan Dinţa. Vine le sigur</span></p>
<p><span class="font14">din oraş..... de la bancher.... ce tot are omul</span></p>
<p><span class="font14">asta a face cu bancherii, nu sein.... ml-a'şi pune capul contra unui chilo de vin, ca</span></p>
<p><span class="font14">are cu el câte-va h.irtii de acele..... pc cari</span></p>
<p><span class="font14">i le voiţi visita îndată, căci ori cât de grăbit pare a fi, eu voi., ajunge înaintea lui la han.</span></p>
<p><span class="font14">Şi in adever Dragoş o luase pe poteca lui cea strimta şi ajunse la han cu mult înaintea arendaşului</span></p>
<p><span class="font14">El intri pe porta.</span></p>
```

ALDO ȘI AMINTA

sau

BANDIȚII

PARTEA I.

Les actions des hommes sont une se-
mence fécondérépandue sur les champs
obscurs de l'avenir, confiée avec espé-
rance aux divinités fatales.

Schiller.

I.

Timpulŭ nŭ a grămăditŭ atîțea anĭ în memoria ace-
lŭia ce sŭsținŭ, de kînd fie-kare vale a țiranților Kar-
pațĭ era o forterețŭ mŭnitŭ de natŭrŭ a xaĭdŭcilor noștri;
de kînd fie-kare čimŭ skotea ŭnŭ eko ka sŭ rŭspŭnzŭ la
vočea vravilor, če înkinaŭ în sŭnŭtatea kŭpitanŭlŭĭ lor,
ши kŭlmea večinŭ, revĭvpînds-o, o repŭrta kŭ dragŭ ast-
felŭ pŭnŭ la Bŭčečĭ.

Totŭ se maĭ ține 'n leĝende vĭptŭtea moшilor no-
штрĭ, lokŭlŭ tpĭŭmfŭlŭĭ lop, ши onoapea a maĭ mŭltop spe-
lŭnče, lokŭ de skŭpape alŭ vp'ŭnĭĭ epoŭ;

Dap ŭĭtapea din zi în zi kpeшte, ши memopia tpe-

КАПІТОЛ I.

ЕЛЕФТЕРІКА.

Балул де Жоі сеара, ін зиоа де сінтул Георгіе,
ера унъ раіу пентру mine, ера унъ извору де веселіе,
пе каре inima mea o симте атіт де мултъ, дап вор-
беле меле сінт слабе ка s'o апате індестул ши s'o
поатъ еспіма. Ам възут Феріčіреа съпізінд din 'na-
intea mea, стрълуčind ка ін черу інцеріі, авеа таліа
свелтъ ши оки албастрі ка сенінул черулуі, інвестітъ
ка о колумбъ, Фъръ 'ніčі о патъ пе dînsa. Еа ера
чеа маі граціоасъ ши чеа маі біне Фъкутъ din кîте
а възут оки mеĭ. Дар, еу, ваі! ерам уну копілу, пе
лînгъ dînsa, къčĭ, опĭ кînd дам s'o пpинзъ, опĭ кînd
вреам съ'ĭ ворбеск, пpинтp'ун съпîс ал окілор еĭ,
пpинтp'о къутътуръ де Фокъ, Фъчеа пептул съ'mĭ салте
ши ворбеле меле ерау інъбъшіте de бътъіле inimeĭ.
Еу, рътъčiam, ку toate ačestea, ін хаосул бъкуріі ши
ъртъреам Феріčіреа че ам зърit'o deskpiindu-se din
'naintea окілор meĭ.

OCR output: absolutely unreliable

C:\Users\roxana\Documents\proiecte - finantari\PARTHENOS\aldo aminta.htm (CORPUS_MISTERE_19_output) - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

untitled | aldo aminta.htm

FOLDERS
CORPUS_MISTERE

```html
1  <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"
2      "http://www.w3.org/TR/html4/loose.dtd">
3  <html>
4  <head><meta http-equiv="content-type" content="text/html; charset=UTF-8"/><meta name="generator" content="
   ABBYY FineReader 12"/><meta name="author" content="roxana patras"/><link rel="stylesheet" href="aldo
   aminta_files/aldo aminta.css" type="text/css"/>
5  </head>
6  <body>
7  <p><span class="font3" style="font-weight:bold;">SAU</span></p><img src="aldo aminta_files/aldo aminta-1.jpg
   " style="width:227pt;height:57pt;"/>
8  <p><span class="font3" style="font-weight:bold;">—&lt;&amp;&amp;&amp;$&amp;&amp;&amp;&gt;—</span></p><a name
   ="caption1"></a><h1><a name="bookmark0"></a><span class="font1">PABTEA I.</span></h1>
9  <p><span class="font3" style="font-weight:bold;">Les actions des hommes sont une semence fécondejépandue
   sur les champs obscurs de l'avenir, confiée avec espérance aux divinités fatales.</span></p>
10 <p><span class="font3" style="font-weight:bold;font-style:italic;">Schiller.</span></p><h2><a name="
   bookmark1"></a><span class="font0">I.</span></h2>
11 <p><span class="font4">Timnblb ns a rpT.im.dit8 atîijea anï în memopia aqe-l»ia qe sbsnim., de kînd
   fie-kape Baie a ijirani^ilop Kap-naijï epa o foptepegi» nibnih. de natsp't a xaldbqilop nompi; de kînd
   fie-kape qiim» skotea bris eko ka si. pisnbnzT. la Boqea Epaoilop, qe înkinas în s^nitatea kinitanblbl
   lop, mi kblmea Beqim», peBir.pîndb-o, o penbpta kb dparb ast-felb mm la Bbqeqï.</span></p>
12 <p><span class="font4" style="font-weight:bold;">Totb se mai yine 'n leidende Biptbtea momilop no-mtpi,
   lokblb tpibrnfblbï lop, mi onoapea a mal rnbltop sne-lsnqe, lokb de skinape alb </span><span class="font2"
   style="font-weight:bold;">Bp'biibî </span><span class="font4" style="font-weight:bold;">epob;</span></p>
13 <p><span class="font4" style="font-weight:bold;">Dap bïtapea din zi în zi kpeuite, mi memopia tpe-</span></p
   >
14 </body>
15 </html>
```
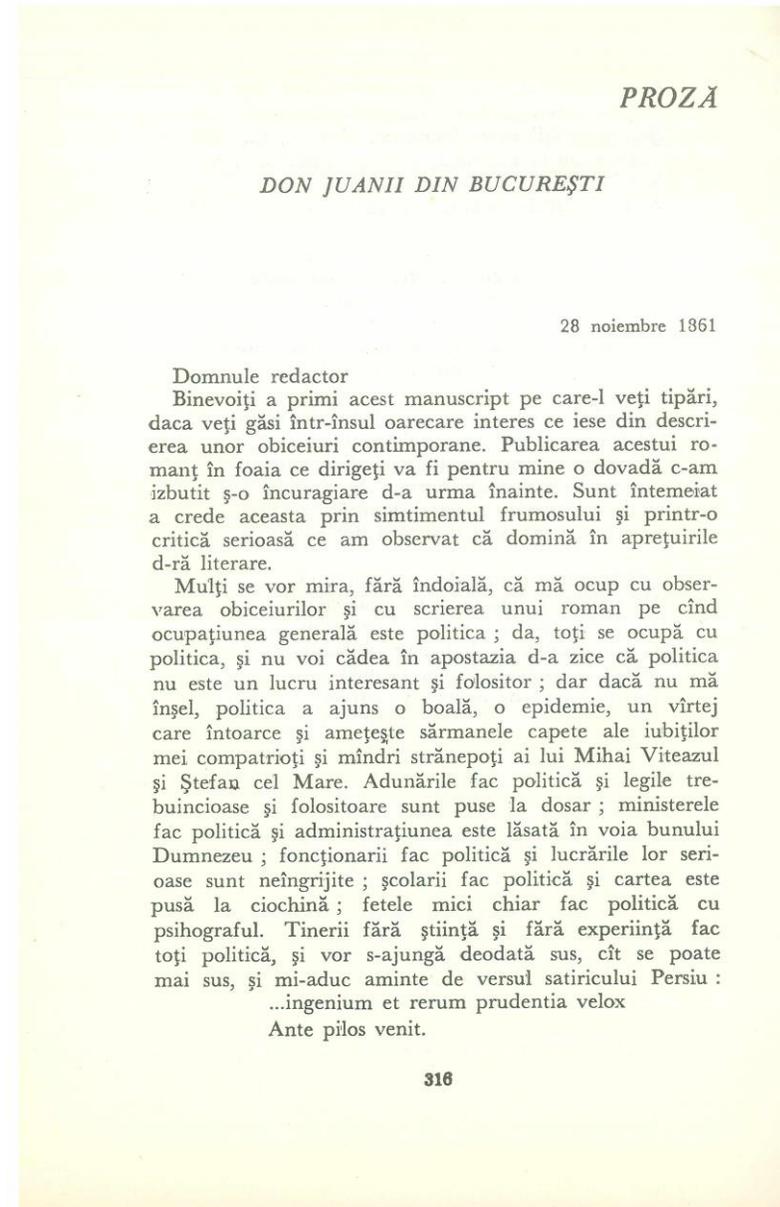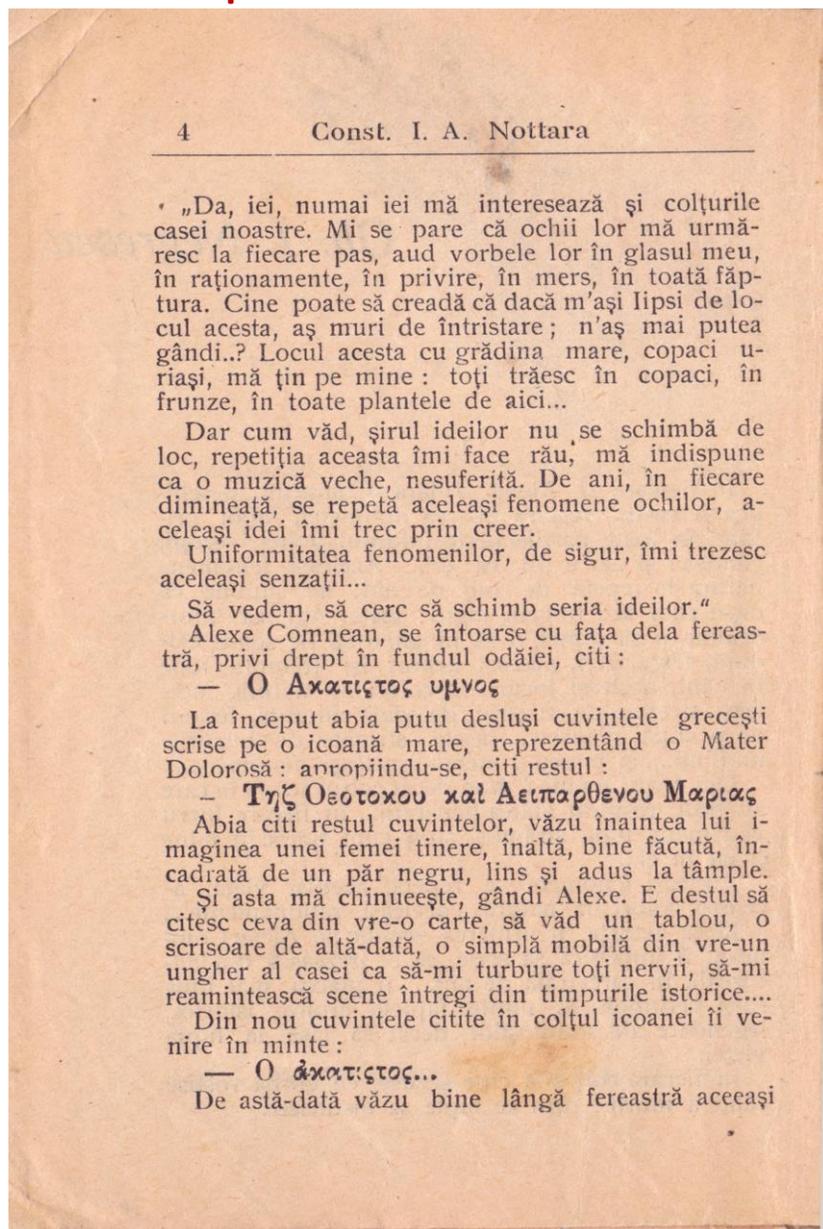
Activate Windows
Go to Settings to activate Windows.

Line 1, Column 1   Tab Size: 4   HTML

Some normal and, of course, normalized texts
Books published between 1920-1945 & scholarly editions published between 1950-1989

· „Da, iei, numai iei mă interesează şi colţurile casei noastre. Mi se pare că ochii lor mă urmăresc la fiecare pas, aud vorbele lor în glasul meu, în raţionamente, în privire, în mers, în toată făptura. Cine poate să creadă că dacă m'aşi lipsi de locul acesta, aş muri de întristare ; n'aş mai putea gândi..? Locul acesta cu grădina mare, copaci uriaşi, mă ţin pe mine : toţi trăesc în copaci, în frunze, în toate plantele de aici...

Dar cum văd, şirul ideilor nu se schimbă de loc, repetiţia aceasta îmi face rău, mă indispune ca o muzică veche, nesuferită. De ani, în fiecare dimineaţă, se repetă aceleaşi fenomene ochilor, aceleaşi idei îmi trec prin creer.

Uniformitatea fenomenilor, de sigur, îmi trezesc aceleaşi senzaţii...

Să vedem, să cerc să schimb seria ideilor.“

Alexe Comnean, se întoarse cu faţa dela fereastră, privi drept în fundul odăiei, citi :

— Ο Αχατιϛτος υμνος

La început abia putu desluşi cuvintele greceşti scrise pe o icoană mare, reprezentând o Mater Dolorosă : apropiindu-se, citi restul :

— Τηϛ Οϭοτοχου χαλ Αϭιπαρθϭνου Μαριαϛ

Abia citi restul cuvintelor, văzu înaintea lui imaginea unei femei tinere, înaltă, bine făcută, încadrată de un păr negru, lins şi adus la tâmple.

Şi asta mă chinueşte, gândi Alexe. E destul să citesc ceva din vre-o carte, să văd un tablou, o scrisoare de altă-dată, o simplă mobilă din vre-un unghier al casei ca să-mi turbure toţi nervii, să-mi reamintească scene întregi din timpurile istorice....

Din nou cuvintele citite în colţul icoanei îi venire în minte :

— Ο ἀχατιϛτος...

De astă-dată văzu bine lângă fereastră aceeaşi

PROZĂ

DON JUANII DIN BUCUREŞTI

28 noiembre 1861

Domnule redactor

Binevoiţi a primi acest manuscript pe care-l veţi tipări, daca veţi găsi într-însul oarecare interes ce iese din descrierea unor obiceiuri contimporane. Publicarea acestui romanţ în foaia ce dirigeţi va fi pentru mine o dovadă c-am izbutit ş-o încuragiare d-a urma înainte. Sunt întemeiat a crede aceasta prin simtimentul frumosului şi printr-o critică serioasă ce am observat că domină în apreţuirile d-ră literare.

Mulţi se vor mira, fără îndoială, că mă ocup cu observarea obiceiurilor şi cu scrierea unui roman pe cînd ocupaţiunea generală este politica ; da, toţi se ocupă cu politica, şi nu voi cădea în apostazia d-a zice că politica nu este un lucru interesant şi folositor ; dar dacă nu mă înşel, politica a ajuns o boală, o epidemie, un vîrtej care întoarce şi ameţeşte sărmanele capete ale iubiţilor mei compatrioţi şi mîndri strănepoţi ai lui Mihai Viteazul şi Ştefan cel Mare. Adunările fac politică şi legile trebuincioase şi folositoare sunt puse la dosar ; ministerele fac politică şi administraţiunea este lăsată în voia bunului Dumnezeu ; fonçţionarii fac politică şi lucrările lor serioase sunt neîngrijite ; şcolarii fac politică şi cartea este pusă la ciochină ; fetele mici chiar fac politică cu psihograful. Tinerii fără ştiinţă şi fără experiinţă fac toţi politică, şi vor s-ajungă deodată sus, cît se poate mai sus, şi mi-aduc aminte de versul satiricului Persiu :

...ingenium et rerum prudentia velox

Ante pilos venit.

316

Drawback II: TEI HEADER – particularities
Referencing facsimiles on zenodo https://zenodo.org/record/3229284

● ROM014_IonescuR_CatastihulAmorului.xml ✕   ● **ROM015_IonescuR_DonJuaniiBucuresti.xml** ✕

TEI   teiHeader   fileDesc   extent   measure

```xml
1   <?xml version="1.0" encoding="UTF-8"?>
2   <?xml-model href="eltec-1.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
3 ▽ <TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:e="http://distantreading.net/eltec/ns" xml:id="ROM015_IonescuR_DonJuaniiBucuresti" xml:lang="ro">
4 ▽  <teiHeader>
5 ▽   <fileDesc>
6 ▽    <titleStmt>
7      <title> Don Juanii din București: ediție ELTeC</title>
8      <author> Ionescu, Radu (1834-1872)</author>
9 ▽    <respStmt>
10      <resp>Codare ELTeC</resp>
11      <name>Roxana Patras</name>
12     </respStmt>
13    </titleStmt>
14 ▽  <extent>
15    <measure unit="words">22860</measure>
16    <measure unit="pages">67</measure>
17   </extent>
18 ▽  <publicationStmt>
19    <p>Încorporat în ELTeC <date>2019-08-28</date></p>
20    </publi...
21 ▽   ...sc>
22 ▽  <bibl type="printSource">
23    <author>Radu Ionescu</author>
24    <title>Don Juanii din București</title>
25    <pubPlace>Bucharest</pubPlace>
2    <publisher>Editura Minerva</publisher>
2    <date>1974 (1861-1862)</date>
2    <idno type="DOI">10.5281/zenodo.3229285</idno>
2   </bibl>
```

Text   Grid   Author

RPUSURI\ELTEC transa 1\ROM015_IonescuR_DonJuaniiBucuresti....   ■ Validation successful   U+000A   15 : 40

```xml
2   <?xml-model href="eltec-1.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
3   <TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:e="http://distantreading.net/eltec/ns" xml:id="ROM015_IonescuR_DonJuaniiBucuresti" xml:lang="ro">
4     <teiHeader>
5       <fileDesc>
6         <titleStmt>
7           <title> Don Juanii din București: ediție ELTeC</title>
8           <author> Ionescu, Radu (1834-1872)</author>
9           <respStmt>
10            <resp>Codare ELTeC</resp>
11            <name>Roxana Patras</name>
12          </respStmt>
13        </titleStmt>
14        <extent>
15          <measure unit="words">22860</measure>
16          <measure unit="pages">67</measure>
17        </extent>
18        <publicationStmt>
19          <p>                    00-28</date></p>
20          <publicationStmt>
21          <sourceDesc>
22          <bibl type="printSource">
23            <author>Radu Ionescu</author>
24            <title>Don Juanii din București</title>
25            <pubPlace>Bucharest</pubPlace>
26            <publisher>Editura Minerva</publisher>
27            <date>1974 (1861-1862)</date>
28            <ref target="https://zenodo.org/record/3229286">Facsimile on zenodo</ref>
29          </bibl>
```

TEI teiHeader fileDesc extent measure

ROM014_IonescuR_CatastihulAmorului.xml   ROM015_IonescuR_DonJuaniiBucuresti.xml

Text   Grid   Author

Next   Previous   All   ☑ Incremental   ☐ Case sensitive   Close

ORPUSURI\ELTEC transa 1\ROM015_IonescuR_DonJuaniiBucuresti....   ■ Validation successful   U+000A   16 : 37

Activate Windows
Go to Settings to activate Windows.

Drawback III: The Splendors and Miseries of Digital Literary Studies in Romania in 2019
or
A Harlot High and Low

- **treatment of metadata rather than data:**
a plethora of recent studies strongly rely on a basic counting of metadata provided by the entries in various dictionaries and lexicons (e.g. Transylvanian Review, vol XXXVIII, supplement no. 1, 2019,
  - *Romanian Literature in the Digital Age*)

- **strong focus on post-45 Romanian literature and on non-literary corpora:**
Corpora such as ROMBAC, RODICA, and so forth

- low digitization of both resources and library metadata: e.g. partially digitized catalogue of the Library of the Romanian Academy;

- dacoromanica platform not really functional;

- shy initiatives started by university libraries, but no interoperable formats (xml, txt, epub)

- incomplete references of this literary span (1840-1920) in Worldcat & others

- Romanian corpora haven't included enough literary texts and if they have, texts were selected from those published after 1945 (standardized Romanian)

- books indexed in older printed catalogues may not have survived (books to be scanned are fragile and not suited for regular scanning)

- no lemmatizers and POS taggers for diachronic varieties of Romanian (thus automatic analysis functions suboptimally)

Briefly...
Creating such a collection feels like writing fixed verse: allign resources to sampling principles

- at least 10%-50% have been written by female authors for the language subcollection.

- 9 to 11 authors are represented with exact three novels.

- at least 30% are highly canonized novels, at least 30% should be non-canonized novels, based on the following reprint groups: reprinted not at all, reprinted once, reprinted more than once within the period 1980-2000

- at least 20% are short novels (10-50k word tokens), at least 20% are long novels (>100k word tokens).

Date : 1840 to 1920 (first iteration)

We will divide into four groups

- group A (1840-1859): code T1

- group B (1860-1879): code T2

- group C (1880-1899): code T3

- group D (1900-1920): code T4

Output of Romanian novels vs. ELTeC groups (Time Slots): culturally determined unbalance of a literary corpus

	T1 (1840-1859)	T2 (1860-1879)	T3 (1880-1899)	T4 (1900-1920)
Total no	11	38	120	181
Transition Alphabet	4	4	0	0
Incomplete	2	2	1	-
Under 10,000 words (around 100p)	4	5	19	27
Posthumous		2	1	-
Other languages		1 (French)	6 (French and German)	2 (German)
CANDIDATES for ELTeC	**1**	**24**	**94**	**152**

How many of them actually available in Romanian libraries after

- **2 World Wars**
- **Communist expurgation of undesirable books published during the interbellum period or even during the 19th-century**
- **The 89 Revolution**

Female authors: bettering the scores

From 37 novels authored by females:

-5 are written in French and German

-4 are around 100p (possibly under 10,000 words)

If the 9 problematic items are taken out, in real terms we've got only 8% female-authored novels

Among the 28 novels that would fit as candidates:

2 authored 3 novels (I.G.Lecca, S. Nadejde)

2 authored 2 novels (Smara, E.I. De Reus, Olteo)

6 authored 1 novel (E. Bacaloglu, M. Miller Verghi, C. Hodos, A. Xenopol, S. Cassvan, V. Ermali)

1 authored 3 novels but 2 of them are too short (E. Tailler is left with only 1)

Sampling 9-11 authors with exactly 3 books

Who published at least 3 books between
1840-1920?

1. D. Bolintineanu
2. R. Ionescu
3. N.D. Popescu
4. P. Macri
5. C. Sandu-Aldea
6. N. Rădulescu-Niger
7. I. Pop-Florantin
8. A. Theochari
9. M. Sadoveanu
10. Duiliu Zamfirescu
11. I. Slavici
12. R. Rossetti
13. V. Demetrius

14. G. Baronzi
15. Al. Pelimon
16. C. D. Aricescu
17. S. Nadejde
18. I.G. Lecca
19. A. I. Alexandrescu
20. C. Oeconomu
21. D.C. Moruzi

Which one would you chose?

Debatable choices from the viewpoint of established Romanian literary historians and theorists

How can we get more novels in a specific group?

- Pioneer editing of problematic novels (e.g. novels in the transition alphabet can be recovered by training a HTR model on Transkribus)
- Open discussion about the status of translations or versions (into national languages) that have been authorized by the authors themselves

What is the best way to get/ produce useful resources?

- customized, project-oriented libraries (https://zenodo.org/login/?next=%2Fdeposit%3Fpage%3D1%26size%3D20)
- larger digitization projects with a strong and well defined research focus: e.g. literary genres (novel, poetry, theatre), periods, themes, and other types of topical interests

- 12:00-13:00　　　Roxana Patras

Experiece with building a Romanian Language collection

Focus on typefaces, special characteristics on the text marterial etc.

Hands-on session (transliteration)

Introduction to Transkribus

TRANSKRIBUS enables you to:

➢register in the platform

➢download an expert tool specifically designed for your needs

➢upload your own documents/images

➢manage your own private collection (no one else has access to your documents!)

➢segment the images into blocks, lines/baselines and words with the support of layout analysis tools

➢link the text with the image which increases the value of your transcription significantly (actually you cannot transcribe text without linking!)

➢transcribe text in any language and with any character set (load your own virtual keyboard)

➢export your documents at every time in several formats such as TEI, RTF, PDF, XML.

This sounds interesting, but it starts to get really exciting if you consider that

- once you have properly transcribed e.g. 100 images you may inform us and we will train an HTR engine from the Computational Intelligence Technology Lab (CITlab) of the University of Rostock on your documents and

- you will be able to transcribe further pages of your documents with the support of automatically produced handwritten text.

Installation

- Register

- Download Transkribus

- Try out some test documents with automatically produced full-text in the TranskribusCloud collection

Or

- Try to do it intuitively by starting from the documents available in the google drive

Samples from 4 Romanian typesets

1. C.A. Aricescu, Misterele Casatoriei. Barbatul desilusionat, vol 3, 1866

2. G. Baronzi, Fontana zanelor, 1896

3. Ilie Ighel, Banditul Simion Licinski, 1890

4. Th. A. Myller, In Iassy, 1871

5. G. Baronzi, Misterele Bucurestilor, vol 1, 1862

People that are not familiar with language and contents are better in transliteration:

What they see is what they retrieve

So

You won't learn Romanian in 1h!

But at least you'll train your eyes and patience for other situations when you need to treat a print as a handwritten text

References

Dicționarul cronologic al romanului românesc (2003). Vol. 1. Bucharest: Editura Academiei Române.

Dicționarul romanului românesc tradus (2005). Vol. 1. Bucharest: Editura Academiei Române.

Bibliografia românească modernă (1831-1918), https://biblacad.ro/bnr/brm.php

Drăgan, I. (2001). Romanul popular in România. Literar si paraliterar. Cluj: Casa Cartii de Stiinta.

Barbu, M. (2003). Romanul de mistere în literatura română. Craiova: Fundatia Scrisul românesc.

Transylvanian Review, vol XXXVIII, supplement no. 1, 2019, Romanian Literature in the Digital Age

***, New Literary History. Theorizing Genres, Vol. 34, no. 2, spring 2003.

Arthur, P & Bode, Katherine, Advancing Digital Humanities: Research, Methods, Theories, Pagrave Macmillan, 2014.

Barbu Mititelu, V. et al., Corpus of Contemporary Romanian. Architecture, Annotation Levels and Analysis Tools, in Helga Bogdan Oprea et alii (eds.), Lingvistică românească, lingvistică Romanică, Editura Universității din București, 2017, pp. 13-20.

Bernard, M.; Bohet B. (2017). Littérométrie. Outils numériques pour l'analyse des textes littéraires, Paris, Presses de la Sorbonne-Nouvelle.

Bode, Katherine, Reading by numbers: recalibrating the literary field, Anthem Press, 2012.

Bode, Katherine, The equivalence of "close" and "distant" reading; or, toward a new object for data-rich literary history, in Modern Language Quarterly, 78 (1), 2017, pp. 77-106.

Burnard, Lou, Qu'est-ce que la Text Encoding Initiative ?, Nouvelle édition [en ligne], Marseille, OpenEdition Press, 2015.

Chivu, Gh. et al. (éd.), Studii de istorie a limbii române. Morfosintaxa limbii literare în secolele al XIX-lea şi al XX-lea, Bucureşti, Editura Academiei Române, 2015

Eder, M. ; Rybicki, J. ; Kestemont, M., Stylometry with R: a package for computational text analysis, in R Journal, 8 (1), 2015, pp. 107–121.

Galleron, I, Conceptualisation of theatrical characters in the digital paradigm: needs, problems and foreseen solutions, Human and Social studies, De Gruyter, vol. 6, issue 1, 2017.

Gheţie, I. ; Mareş, Al., De când se scrie româneşte?, Bucureşti, Univers Enciclopedic, 2001.

Heiden, S., The TXM Platform : Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme, In K. I. Ryo Otoguro (Ed.), 24th Pacific Asia Conference on Language, Information and Computation (p. 389-398). Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010.

Hettinger, Lena et alii, Significance Testing for the Classification of Literary Subgenres, DH conference paper, 2016, http://www.dhd2016.de/abstracts/vortr%C3%A4ge-049.html.

Hoorn, Johan F, How is a Genre Created? Five Combinatory Hypotheses, CLCWeb: Comparative Literature and Culture 2.2, 2000, https://pdfs.semanticscholar.org/0dea/af47fa52207388fbef4bf9f769d2e26a2a9b.pdf.

Ivănescu, G., Istoria limbii române, Iaşi, Junimea, 2000.

Jockers, M., Macronanalysis. Digital methods and literary history, Urbana, Chicago and Springfield, University of Illinois Press, 2013.

Jockers, M. L., Detecting and Characterizing National Style in the 19th Century Novel, Paper presented at Digital Humanities Conference 2011, Stanford University, 2011, pp. 159-160.

Jockers, Matthew L., Computing and Visualizing the 19th Century Literary Genome, Paper presented at Digital Humanities Conference 2012, University of Hamburg, 2012, p. 242-244.

Juola, P., Authorship Attribution, in Foundations and Trends in Information Retrieval. 1 (3): 3, 2006.

Kenny, A., The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities, Oxford: Pergamon Press, 1982.

Mancaş, M, Limbajul artistic românesc modern. Schiţă de evoluţie, Editura Universităţii din Bucureşti, 2005.

Moretti, F., Distant reading, Londres et New York, Verso, 2013.

Novakova, Iva, Siepmann, Dirk (Eds.), Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives, Palgrave-Macmillan, 2019.

Pană Dindelegan, Gabriela (ed.)., The Syntax of Old Romanian, Oxford, Oxford University Press, 2016.

Rhody, L., Topic Modeling and Figurative Language, in Journal of Digital Humanities, 1/ 2012.

Schöch, Christof, Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama, in Digital Humanities Quarterly, 11, no. 2, 2017, p. 1-53.

Tufiş, D., CoRoLa Primul corpus computaţional de referinţă pentru limba română contemporană, in Market Watch, no. 205, 2018, p. 28-29;

Van Holland, S.; Hengchen, S.; Gillet, F.; De Wilde, M., Introduction aux humanités numériques, De Boeck, 2016.

Villalva, A.; Williams, G. (eds.), The Landscape of Lexicography, John Benjamins, 2019.

Williams, G., Collocational Networks: Interlocking Patterns of Lexis in a Corpusof Plant Biology Research Articles, in International Journal of Corpus Linguistics, vol. 3, issue 1, 1998, p. 151-171.

Zafiu, R., Criterii estetice în normarea limbii române, în C. Stan, R. Zafiu, Al. Nicolae (coord.), Studii lingvistice. Omagiu profesoarei Gabriela Pană Dindelegan, la aniversare, Bucureşti, Editura Universităţii din Bucureşti, 2007, p. 467-473

If you have suggestions for ELTeC in general, for the national collections of ELTeC or anything else related to this project

please drop me a line at

CS II dr. Roxana Patras

roxana.patras@uaic.ro

https://proiectulbrancusihairo.wordpress.com/

`<p> <emph>`THANKS!`</emph></p>`

`</body>`
`<back></back>`
`</text>`