

EDITAREA DIGITALĂ A TRADUCERILOR OPERELOR ISTORICE ALE LUI DIMITRIE CANTEMIR

Cristina Vertan*

1. Introducere

Personalitatea culturală multiplă a lui Dimitrie Cantemir, atât de cunoscută publicului românesc (și nu numai celui de specialitate, de exemplu istorici, cercetători în domeniul culturii otomane, muzicologi), este mult mai puțin reliefată peste hotare. În afara specialiștilor, puțini sunt cei care știu că lucrările sale despre istoria Imperiului Otoman sau a Moldovei au fost opere de referință și principala sursă de informație despre aceste regiuni pentru întreaga Europă, până la mijlocul secolului al XIX-lea.

Lucrările lui Dimitrie Cantemir au fost traduse în limbi de circulație europeană (germană, engleză, franceză) în secolul al XVIII-lea. În consecință, traducerile utilizează limba acelor vremuri. În cazul traducerilor în limba germană diferă și tipul de font utilizat pentru tipărire, față de cel din ziua de astăzi. Texte tipărite în „Fraktur” sunt greu accesibile cititorilor din ziua de azi.

Nu numai forma, dar și conținutul ridică anumite probleme de înțelegere. Textul abundă în nume proprii, toponime, nume de obiecte, obiceiuri, meserii, scrise de Cantemir în românește sau turcește și adaptate fonetic de translatorul german sau chiar de Cantemir, atunci când realizează, în cazul limbii române, transliterarea cu caractere latine a scrierii slavone. Pe de altă parte, harta originală atașată în *Descrierea Moldovei* are toate denumirile în limba română și este preluată de translatorul german, dar neadaptată transliterărilor fonetice utilizate în text. Astfel cititorul este confruntat cu cel puțin două denumiri pentru același loc.

Utilizarea noilor tehnologii digitale oferă posibilitatea redării operei lui Dimitrie Cantemir într-un mod mult mai accesibil. Mai mult, este posibilă o mapare a locurilor și numelor evocate pe coordonate actuale, respectiv baze de date de persoane, iar pe textele în formă digitală se pot realiza interogări complexe.

În acest articol vom descrie un proiect aflat în desfășurare la Universitatea din Hamburg și având ca obiectiv principal o ediție digitală, într-o primă fază a traducerii germane a celor două opere de căpătâi ale lui Dimitrie Cantemir, *Istoria creșterii și descreșterii Imperiului Otoman* (în limba germană *Geschichte des osmanischen Reichs nach seinem Anwachsen und Abnehmen*) (Cantemir, Dimitrie,

* Grupul de cercetare „Computerphilologie”, Universitatea din Hamburg.

1774) și *Descrierea Moldovei* (în limba germană *Beschreibung der Moldau*) (Cantemir, Dimitrie, 1771). Ediția digitală va fi îmbogățită cu informații legate de numele proprii de persoane, locuri, obiceiuri, de date istorice și calendaristice, astfel încât cititorul din ziua de azi să poată înțelege ușor manuscrisul, chiar fără a avea cunoștințe aprofundate despre Moldova din acea vreme sau despre Imperiul otoman. Menționăm că în paragrafele următoare ne vom referi exclusiv la lucrarea *Descrierea Moldovei*. În secțiunea a doua vom ilustra particularitățile de formă și de conținut ale textului, problemele pe care acestea le pun în procesul de digitizare, precum și soluțiile adoptate de noi. Secțiunea a treia este dedicată unei caracteristici centrale a tuturor textelor istorice, anume prezența unor fapte și/sau afirmații incerte, vagi. Modelarea incertitudinii este pentru moment relativ ignorată de proiectele de digitizare, aici intervenind unul dintre aspectele inovatoare ale proiectului nostru, care își propune includerea incertitudinii în modelarea datelor. În secțiunea a patra vom descrie principalele componente ale sistemului și vom prezenta aspectele pe care ne vom concentra în perioada următoare.

2. Particularități ale textului și complexitatea prelucrării digitale

Traducerea în limba germană a originalului în limba latină a lucrării *Descriptio Moldaviae* a lui Dimitrie Cantemir a apărut în 1771. Nu există o traducere în germană mai recentă a acestei lucrări. De menționat este faptul că traducerea în limba română din secolul al XIX-lea¹ s-a realizat având ca text sursă traducerea germană și nu originalul latinesc, care la acea vreme era dispărut. Cele două versiuni – germană din 1771 (Cantemir, Dimitrie, 1771) și română (Cantemir, Dimitrie, 1909) – pot fi, așadar, ușor alineate. Pentru tehnologia digitală acest fapt este de o deosebită importanță, fiind unul dintre puținele cazuri în care se poate realiza un corpus bilingv paralel, cu material reflectând limba secolelor trecute.

2.1. Particularități legate de transliterare și caractere/fonturi utilizate

Aproape fiecare pagină a textului conține cel puțin câteva cuvinte, dacă nu propoziții întregi, într-o altă limbă decât cea germană. În funcție de limbă se schimbă uneori și literele și/sau fontul utilizate pentru tipărire astfel:

¹ Vide Dimitrie Cantemir, *Descrierea Moldovei*, ed. Gh. Guțu, București, Editura Academiei, 1973, p. 40: „din inițiativa mitropolitului Veniamin Costachi a fost tradusă în românește de banul Vasile Vârnav, apoi tipărită în 1825 la mănăstirea Neamțu, cu o precuvântare de ieromonahul Gherontie. Alte două ediții ale versiunii Vârnav aveau să urmeze la scurte intervale, una la Iași, prin grija scriitorului Costache Negruzzi, în 1851, alta, tot acolo, inițiată de antiunionistul Teodor Boldur-Lățescu (în 1868). [...] Abia în 1869, Societatea Academică de curând înființată lua cunoștință de existența în biblioteca Academiei Teologice din Moscova a unui manuscris latin al *Descrierii*.” Vide traducerea începută de Al. Papiu Ilarian și încheiată de Iosif Hodoș: *Descrierea Moldovei*, tradusă din textul original latinesc, aflat la Muzeul Asiatic al Academiei Imperiale Științifice de la St. Petersburg, 1874.

- Textul german este tipărit cu litere latine cu font de tip „Fraktur”.
- Termenii și citatele din limba greacă sunt tipărite cu alfabetul grec.
- Citatele din latină sunt tipărite cu caractere latine.

Numele de persoane, locuri, obiceiuri românești sunt tipărite cu caractere latine. Ortografia folosită pentru transliterarea numelor este diferită de la un capitol la altul, acest fenomen având două surse:

1. pe de o parte, Cantemir însuși face una dintre primele încercări de scriere a limbii române cu caractere latine, știut fiind că, în vremea sa, se utiliza alfabetul chirilic. De aceea transliterarea sunetelor specifice limbii române: *ă, î, â, ș, ț* nu este standardizată, la fel ca și a africatelor *ce, ci, che, chi*;
2. pe de altă parte, translatorul german încearcă el însuși să adapteze sunetele limbii române la pronunția germană.

Se ajunge astfel la o multitudine de variante pentru unul și același nume. De exemplu, într-o selecție de 150 de pagini, regăsim 117 nume de localități, numai pentru orașul Iași găsindu-se cinci variante: *Iaschi, Iașii, Jassii, Jașii, Jassy*.

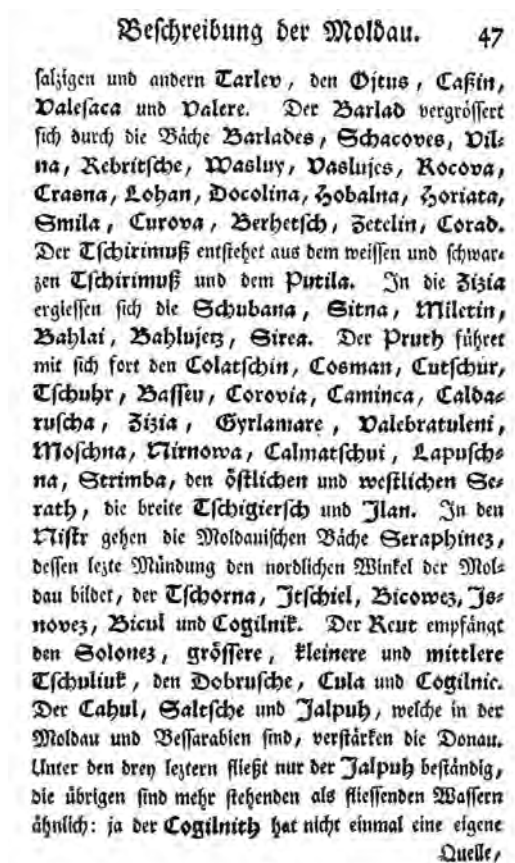


Figura 1. Pagină ilustrativă pentru multitudinea de nume proprii.

În Figura 1 prezentăm un exemplu de pagină care include toponime numeroase, greutatea citirii provenind și din faptul că transliterarea este redată cu caractere de tip „Fraktur”.

Este evident că această multitudine de variante reprezintă o problemă pentru orice sistem automat, care trebuie să recunoască și să mapeze toate aceste nume la o singură coordonată geografică.

Totodată, această conglomerare de fonturi și alfabetice face imposibilă utilizarea în exclusivitate a unui program OCR², care ar permite automat transformarea imaginilor fiecărei pagini într-un text care poate fi mai apoi prelucrat. În proiectul de față am optat pentru o soluție semiautomată, ceea ce înseamnă utilizarea programului ABBY cu pachetul special pentru fontul de tip „Fraktur”³ și incluzând ca limbi ce ar trebui recunoscute: româna, latina și greaca. Rata de eroare pe o pagină încărcată mediu cu cuvinte dintr-o altă limbă decât cea germană este de aproximativ 25%. Procesul automat de recunoaștere a caracterelor este urmat de o corectare manuală. Acest pas elimină o primă dificultate în citirea textului și anume descifrarea caracterelor în „Fraktur”.

2.2. Particularități lingvistice

Traducerea a fost realizată în anul 1771 și oglindește limba germană a acelei perioadă. Cu toate că deja vorbim de limba germană modernă (Neuhochdeutsch), există totuși diferențe morfologice, sintactice și semantice față de germana vorbită astăzi. Aceste diferențe influențează pe de o parte inteligibilitatea textului de către cititor (mai ales cele semantice), cât și procesabilitatea sa de către calculator. Progresele în domeniul lingvisticii computaționale au făcut ca pentru limba germană actuală să existe numeroase programe software care să producă o analiză automată morfologică și sintactică a unui text. Din păcate aceste programe nu știu să proceseze direct texte în care apar deviații de morfologie și/sau sintaxă de la gramatica actuală. De aceea este nevoie de o preprocesare a textului.

Cele mai frecvente abateri de la gramatica actuală sunt în domeniul morfologic, de exemplu:

- folosirea diftongului *ey* în loc de *ei*: (*bey* în loc de *bei* (rom. *la, pe lângă*), *dem zweytem* în loc de *dem zweitem* (rom. *al doilea*).
- intercalarea unui *h* între *t* și *e* (*Theil* în loc de *Teil* (rom. *parte*).

Aceste abateri de la normele actuale vor fi rezolvate printr-o transformare a textului utilizând un set de reguli, de tipul celor de mai sus.

Sintaxa pare a se fi schimbat cel mai puțin în ultimii trei sute de ani, neîntâlnindu-se fenomene complet diferite de textele moderne. În schimb, există numeroase cuvinte care au suferit o modificare de semantică și care, pentru înțelegerea textului de către cititorul din ziua de azi, trebuie explicate.

² OCR = Optical Character Recognition.

³ <http://www.frakturschrift.com/de:start>

De exemplu, în propoziția „*welche den flüchtigen Ochsen im Wasser aufgesucht hatte*” (rom. *Care i-a găsit pe boii fugăriți în apă*), adjectivul „*flüchtigen*” este utilizat cu sensul de *fugărit*. În textele moderne, acest sens este aproape inexistent (spre deosebire de substantivul „*der Flüchtiger*”, rom. *fugarul*, care este în continuare folosit). În schimb, a fost adăugat un sens nou: „*volatil*”.

O altă problemă este ridicată de anumite referințe de timp și spațiu care, dacă nu sunt explicate suplimentar, pot introduce erori sau ambiguități de percepere. De exemplu, în expresia „*von dem heutigen Hungarn*” (din Ungaria actuală), D. Cantemir face referire la Ungaria secolului al XVIII-lea, parte a Imperiului Austro-Ungar. Fără explicații suplimentare, cititorul neavizat se va gândi desigur la Ungaria așa cum o cunoaștem astăzi.

Evident aceasta nu este numai o problemă a traducerii germane, ci și a celei române. Traducerea românească are însă două avantaje: în primul rând, a fost realizată mult mai târziu, când eventual granițele se schimbaseră; o simplă notă de subsol poate rezolva ambiguitatea. Și chiar fără această notă de subsol, cititorul român va ști să facă diferențierea.

O problemă specifică traducerii o reprezintă însă următorul paragraf:

„*Die auf der andern Seite angränzende Polen und Russen nennen die Moldauer Wolochen, d. i. Wälsche oder Italiäner, die Walachen aber, die auf dem Gebirge wohnen, heissen sie die Berg-Walachen, oder die Leute jenseits des Gebirges*”.

Aici se descrie modul în care sunt numiți moldovenii de popoarele vecine. Utilizarea termenului german „*Wälsche*” introduce o ambiguitate, greu de rezolvat de un cititor modern pentru care „*Wälsche*” sunt francezii, adică poporul de la sudul Germaniei.

Toate aceste ambiguități trebuie captate și explicate prin adnotări și eventual conexiuni cu baze de date explicative. În cazul referințelor spațiale (vezi cazul Ungariei evocat mai sus) ne propunem să introducem și elemente multimedia, de pildă hărți.

a.1. Incertitudinea faptelor istorice și tehnologiile digitale

Ca toate operele istorice, lucrarea lui Cantemir abundă de fapte neclare, locuri și date incerte. Pentru unele, Cantemir însuși își exprimă îndoiala, pentru altele există note editoriale ale traducătorului, care pune la îndoială sau chiar contrazice spusele lui Cantemir. O a treia categorie o reprezintă afirmații care din perspectiva cititorului de astăzi nu mai pot fi considerate ca sigure. Pentru acest din urmă tip dăm ca exemplu referirile la împăratul roman numit de Cantemir și preluat fără obiecții de translator Nerva Traian. Or, se știe prea bine astăzi că de fapt a fost vorba de doi împărați Nerva, tatăl adoptiv al lui Traian și Traian însuși.

Tehnologiile actuale, folosite de noua direcție de cercetare „*Digital humanities*”, tind să constrângă mult prea mult datele provenite din științele umaniste, în general și cele istorice, în special: fapte, aserțiuni sunt la acest moment clasificate în „*alb*” și „*negru*”, respectiv o anumită aserțiune în text se referă sau nu

la conceptul X. Faptul că autorul nu este el însuși sigur sau că sursele citate sunt vagi (cum este cazul și în textele lui Cantemir) nu este în general marcat. În ultimii ani, standardul de adnotare TEI, utilizat pe scară largă în procesul de prelucrare digitală a textelor, a introdus mecanisme de marcare a incertitudinii prin elementele <certainty> și <precision>. Aceste elemente dau posibilitatea marcării unui nume propriu, de exemplu în cazul Moldovei atât ca regiune, cât și ca element hidrografic. De asemenea se pot asocia valori de credibilitate pentru variante. Momentan nu există însă nicio posibilitate de:

- marcare a expresiilor care introduc o anumită incertitudine sau o siguranță;
- marcare a gradului de siguranță a surselor pe care se bazează autorul în corelație cu faptele expuse.

Redăm mai jos două citate din *Descrierea Moldovei*, în care Dimitrie Cantemir însuși construiește un adevărat suport de discurs cunoscut în lingvistică sub numele de RST:

Exemplul 1: Fapte menționate ca „sigure” în *Descrierea Moldovei*:

„Pe lângă aceasta, istoricii neamului nostru zic cu toții că moldovenii, după ce s-au întors din Maramureș în patria lor cea veche, au găsit cetăți și castele pustii și astfel *se poate spune fără șovăială* că întemeierea lor trebuie așezată în vremuri mult mai îndepărtate. *Această părere este întărită*, în afară de alte temeuri, și de felul clădirii zidurilor celor mai multe din cetăți care, cu siguranță, nu pot fi decât romane, în afară de câteva, așa cum am arătat mai sus, ce par a fi înălțate în vremuri mai noi împotriva năvălirii tătarilor. *Toate acestea sunt întărite* de mărturiile celor mai de seamă istorici latini, prin care se arată că împăratul roman Traian a adus în Dacia colonii puternice de cetățeni romani, și că urmașul său Adrian, după ce a lăsat barbarilor unele ținuturi din răsărit, n-a părăsit Dacia, numai de teamă ca acestea să nu fie pustiite. Se mai adaugă un monument veșnic al acestui lucru, și anume Valul împăratului Traian”.

Exemplul 2: Fapte declarate „nesigure” de Cantemir însuși în *Descrierea Moldovei*:

„*Mulți cred, dar nu toți*, că aici trebuie să fi descălecat la început pământeni noștri, după ce s-au întors. Căci nu departe de aici, pe malul de răsărit al Siretului, se arată un alt loc, numit și astăzi de către locuitori Smedorova și *se zice* că aici ar fi fost cea dintâi și cea mai mare cetate. *Este sigur* că, multă vreme după aceea, Ștefan cel Mare a înnoit și adus iarăși această cetate la vechea ei faimă”.

„*Am citit odată într-un manuscris al istoriei lui Herodot* că pe Prut, cale de trei zile de la Dunăre, locuia neamul războinic al taifalilor, care își zidise o cetate foarte mare”.

„În apropiere de acest loc se vede o movilă mare, ridicată de mână omenească, numită de tătari Han Țepeși, adică movila hanului, iar de către locuitori Movila Râbâi. *Asupra ivirii ei sunt mai multe păreri. Unii zic* că un han

tătărăsc a fost nimicit cu întreaga lui oaste de către moldoveni și întru pomenire s-ar fi ridicat această movilă; *alții povestesc* că o regină scită, numită Rabie, a fost ucisă în acest loc, pe când ieșise cu oastea împotriva sciților așezați în Moldova și a fost îngropată aici de oamenii săi. *Ce e adevărat sau minciună în aceasta nu cutez să scot din întunericul atât de adânc al acestei legende*”.

Proiectul de față își propune marcarea elementelor de discurs ce indică gradul de credibilitate a faptelor și punerea în legătură prin marcaje specifice a acestor expresii cu numele de locuri, persoane sau datele pentru care există mai multe interpretări.

În acest fel se deschide calea spre o evaluare automată a textului, spre investigarea elementelor de limbaj utilizate de Cantemir pentru exprimarea certitudinii și, într-un pas următor, spre compararea operelor sale cu alte documente ale vremii sau cu comentariile inserate de editori sau traducători.

a.2. Implementare actuală și perspective

O primă etapă de dezvoltare a sistemului a fost reprezentată de identificarea automată a numelor proprii, construirea unei ontologii și asocierea primară între aceste două seturi de date.

Ontologia este realizată în OWL și conține un nucleu de 30 de clase construite manual și legate apoi pentru a asigura consistența de DOLCE – o ontologie de nivel superior (upper-ontology) (Gangemi, Guarino *et alii*: 2002). În figura 2 prezentăm nucleul de concepte, pe care au fost mapate toate numele proprii întâlnite în text.

Numele proprii au fost identificate prin următoarele euristici:

1. Extragerea tuturor cuvintelor marcate bold sau italic, în interiorul unui paragraf.
2. Selectarea celor scrise cu majuscule.
3. Utilizarea unui algoritm de lematizare pentru gruparea tuturor derivatelor morfologice ale unui cuvânt.
4. Extragerea contextului gramatical pentru a se sugera conceptul sau conceptele pe care cuvântul poate fi mapat.
5. Algoritmul a fost implementat cu adaptările specifice pentru limba română și limba germană. În cazul limbii germane, s-a procedat mai întâi la o normalizare morfologică, așa cum a fost prezentată în secțiunea 2.2.

În figura 3 prezentăm rezultatul acestei etape pentru textul în limba română.

Este evident că acest proces introduce o serie de ambiguități care pot fi rezolvate numai prin adnotări suplimentare și, de asemenea, prin introducerea gradului de certitudine cu care una sau alta dintre posibilități poate fi admisă. Această adnotare va fi realizată într-o etapă ulterioară și va urmări introducerea de informații suplimentare din surse variate, precum: structuri sintactice asociate contextului, coordonate geografice, legături cu baze de date de tip Wikipedia, ca în figura 4.

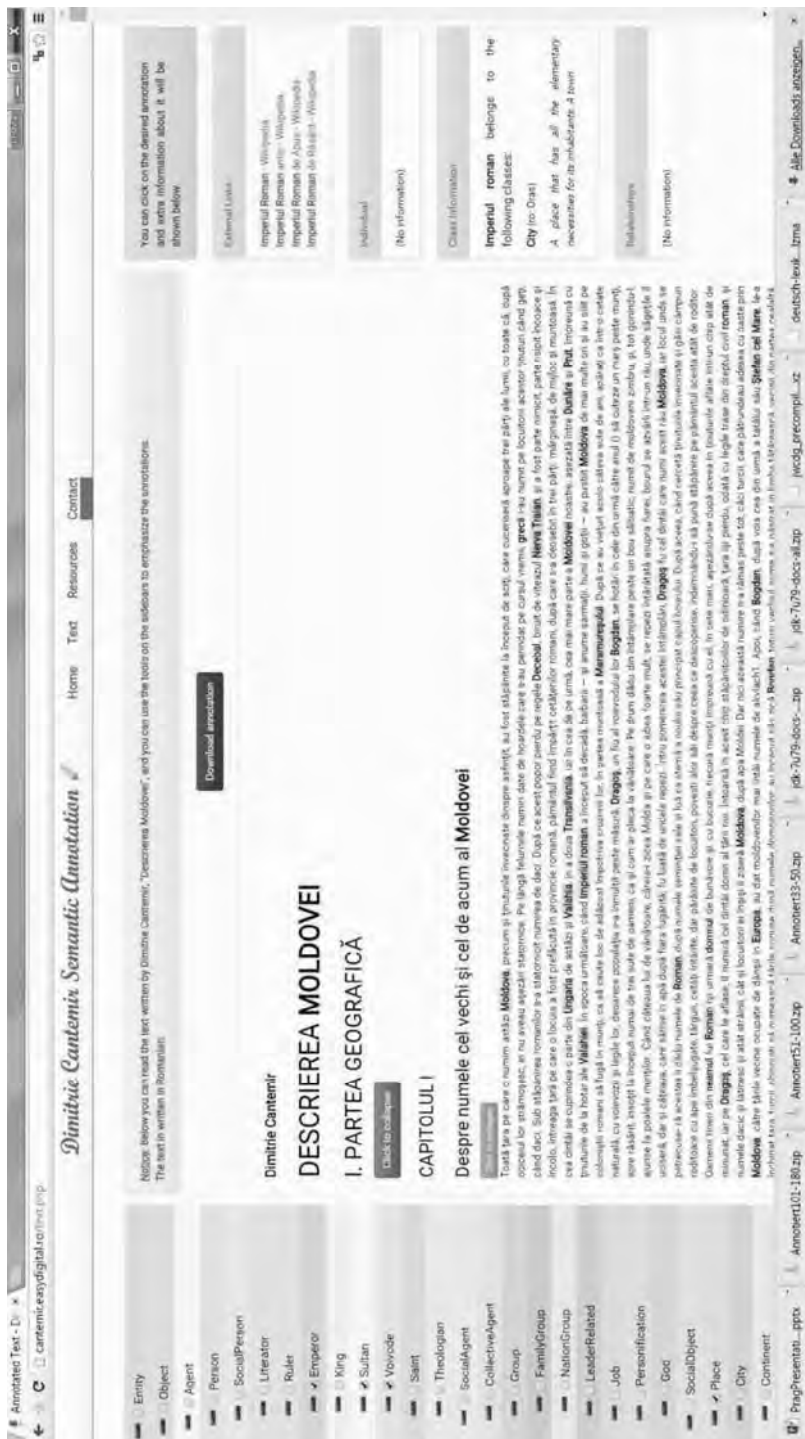


Figura 3. Conectarea numelor proprii cu ontologii.

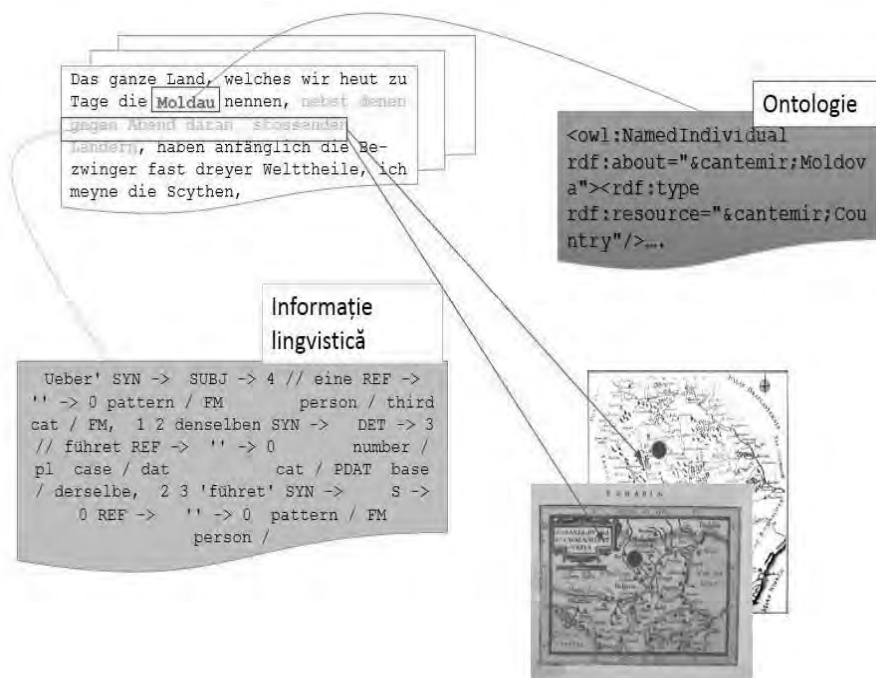


Figura 4. Anotare cu surse de informații multiple.

Concluzii

În acest articol am descris un proiect în desfășurare, pe parcursul căruia se urmărește o adnotare amănunțită a textelor lui Dimitrie Cantemir, în special traducerile originale în limba germană, cu scopul de a le face mai accesibile cititorului modern. Având în vedere că traducerile au fost realizate în secolul al XVIII-lea, sunt prezente anumite elemente de ortografie, morfologie și semantică care trebuie să fie normalizate în unele cazuri sau explicate în alte cazuri. Mai mult decât atât, proiectul urmărește să pună la dispoziția utilizatorului un sistem de adnotare a incertitudinii și a surselor acestei incertitudini.

Analiza lingvistică a textului german a fost realizată de prof. dr. Walther v. Hahn, de la Universitatea din Hamburg. Determinarea automată a numelor proprii în textul în limba română a fost implementată de Alexandra Beldica de la Universitatea din București.

BIBLIOGRAFIE

- Cantemir, Dimitrie, 1771: *Beschreibung der Moldau nebst dem Leben des Verfassers und eine Landkarte*, Frankfurt und Leipzig.
- Cantemir, Dimitrie, 1774: *Geschichte des osmanischen Reichs nach seinem Anwachse und Abnehmen*, Hamburg.
- Cantemir, Dimitrie, 1909: *Descrierea Moldovei*, Editura Librăriei Socec.
- Gangemi, Aldo; Guarino, Nicola; Masolo, Claudio; Oltramari, Alessandro; Schneider, Luc, 2002: *Sweetening ontologies with DOLCE*, în *Knowledge engineering and knowledge management: Ontologies and the semantic Web*, Springer Berlin Heidelberg, p. 166–181.

ABSTRACT

Dimitrie Cantemir was one of the most prominent figures at the end of XVIIth and beginning of the XVIIIth century in the cultural space of Central Eastern Europe. His works about the history of the Ottoman Empire and the Description of Moldavia, were translated in English, German and French and remained capital works until the end of XIXth century. However nowadays, outside Romania his works are known only by specialists.

New digital technologies offer the possibility to annotate and display his texts in a much readable manner. Names of places, persons, dates can be mapped on existent databases and/or maps and linked with other knowledge bases. On such annotations complex queries are possible.

In this article we will describe an ongoing project at the University of Hamburg, aiming to realise a digital edition of the German translations of Dimitrie Cantemir's *Descriptio Moldaviae*. The digital version will be enriched with annotations so that the reader can understand the text without being familiarized with the described region.

Key-words: digital technologies, annotation, German translation *Descriptio Moldaviae*, intelligent search, text understanding.