

**Program *Transkribus* i njegova upotreba u muzikologiji:
primjer III. sveska Kuhačeve korespondencije**

Sara Ries
Odsjek za povijest hrvatske glazbe HAZU
HMD, 21. godišnji susret
Zagreb, 2020.

Transkribus

Unutar projekta READ („Recognition and Enrichment of Archival Documents“), koji je financiran iz fondova Europske unije, razvio se program Transkribus koji je izrađen za optičko prepoznavanje rukopisnih znakova. Program je nastao kako bi se pronašao način za automatsko optičko prepoznavanje znakova povijesnih dokumenata. Softver pokušava automatski prepoznati slova rukopisa na temelju unesenoga skupa podatka kojima se koristi za trening te dodatno omogućuje pretraživanje sadržaja digitaliziranih rukopisa. Ono što je bitno istaknuti, program ne radi na principu OCR-a, odnosno optičkog prepoznavanja znakova (koji se obično koristi za tiskovine ili strojopise) nego prepoznaje retke i grupacije slova. Podržava i stare rukopise poput hebrejskoga, ćirilice i arapskoga, tj. svaki jezik i rukopis. Program se može besplatno preuzeti, a za besplatan pristup potrebno je imati službeni mail preko za registraciju. Nakon instalacije programa može se početi s radom. Upute za rad s programom napisane su na wiki stranicama.

Da bi se napravio vlastiti model (tzv. „training model“) za optičko prepoznavanje znakova određenog rukopisa korisnik mora za najmanje 50-100 (ovisno o rukopisu) skeniranih dokumenata i gotove prijepise. Nakon toga može se kontaktirati proizvođače programa da stvore modul za prepoznavanje znakova na temelju postojećega uzorka. Prema najnovijim istraživanjima, kada se program nauči prepoznavati rukopis, pogreške u prepoznavanju znakova su oko 5%.

Prije govora o samoj transliteraciji korespondencije, valjalo bi razjasniti razliku između transkripcije i transliteracije.

TRANSKRIPCIIJA

U lingvistici, postupak zamjene znakova jednoga pismovnog sustava znakovima drugoga pismovnog sustava. Pri tom se nužno mora voditi računa o izgovoru znakova koje treba transkribirati. Stoga, sustav transkripcije ne pruža obavijesti o tome kako je što bilo napisano u izvorniku, ali pruža obavijesti o tome kako se što izgovaralo u jeziku izvorniku. Transkripciju treba razlikovati od transliteracije, u kojoj se znakovi jednoga pismovnog sustava zamjenjuju znakovima drugoga pismovnog sustava po načelu znak za znak. Pokuša li se transkribirati npr. staro(crkveni) slavenski tekst pisan glagoljicom ili ćirilicom, iz jednoga se pismovnog

sustava (glagoljice, ćirilice) tekst prenosi u drugi pismovni sustav (latinicu), i to tako da se pokušava bilježiti pretpostavljeni izgovor.

Jednom transkribirani tekst ne može se više mehanički vratiti u izvorni tekst jer se više različitih slova može izgovarati na isti način pa se tu ne radi nužno o obostrano jednoznačnom preslikavanju. I po tom se transkripcija razlikuje od transliteracije koja čuva sve razlike među slovima pa se iz transliteriranoga teksta u pravilu ponovno može dobiti izvorni tekst.

TRANSLITERACIJA

Transliteracija, s druge strane, doslovno znači (preslovljavanje). To je postupak zamjene znakova jednoga pismovnog sustava znakovima drugoga pismovnog sustava po načelu »slovo za slovo« ne vodeći pritom računa o izgovoru. Iz transliteracije se zna kako je što bilo napisano u izvorniku, ali se ne zna kako se što izgovaralo. Svako se slovo (grafem) jednoga pisma zamjenjuje odgovarajućim ili najslbličnijim slovom (grafemom) drugoga pisma. Tako se npr. mogu transliterirati grčka ili glagoljična slova na latinicu. Taj način prenošenja znakova iz jednoga pismovnog sustava u drugi slabo pomaže nestručnjacima jer oni ne znaju kako su se pojedina slova čitala. Stoga će se transkripcija primijeniti kada se za širu publiku pripremaju stari tekstovi, a transliteracija kada se pripremaju kritička izdanja, namijenjena stručnjacima, koja trebaju poslužiti za znanstvena istraživanja. Transliteracija je u načelu jednoznačno preslikavanje po načelu znak za znak pa se iz transliteriranoga teksta može ponovno dobiti izvorni tekst. Ipak se ni kod transliteracije često ne može postići potpuna istovjetnost transliteriranoga teksta i izvornika jer su stari tekstovi često pisani u neprekinutome slijedu bez bjelina između riječi. Ako se bjeline i nalaze u tekstu, one su različite od uobičajenih očekivanja. Katkad se u transliteraciji i zanemaruju razlike među slovima ako se komu učine nebitnima. Dakle, transliteracija može dosljednije ili manje dosljedno prenositi razlike u pismu kao što transkripcija može dosljednije ili manje dosljedno prenositi razlike u izgovoru.

Proces transliteracije na Kuhačevoj korespondenciji

Istraživanje Kuhačevih pisama proces je koji je započet prije mnogo godina te je korespondenciju do sada istraživalo više istraživača. Međutim, zbog različitih okolnosti i golemog opsega materijala građa nije u potpunosti obrađena. Jedan od glavnih razloga zbog čega istraživanje nije u potpunosti provedeno je taj da su pisma na njemačkom jeziku pisana rukopisnom goticom, koje se danas smatra izumrlim pismom. Stoga je čitanje građe otežano i rezervirano za mali broj osoba koje poznaju pismo. Da bi se omogućio daljnji rad (analiza i interpretacija) na građi, bilo je potrebno sva pisma pisana spomenutim pismom transliterirati i prevesti.

Postoje transliteracije pisama iz 1950-ih godina koje je napravio Vinko Žganec. Žganec je napravio i registar Kuhačeve korespondencije kako bi istraživačima olakšao pronalazak pojedinih osoba, a te transliteracije i spomenuti registar čuvaju se u strojopisu u Odsjeku za povijest hrvatske glazbe HAZU pod signaturom Pb – 15/1-14. Iako su vrlo koristan materijal za konzultaciju, Žgančeve transliteracije sadrže mnoge pogreške (krivo pročitane ili

izostavljene riječi, ovisno o svesku, tj. neke osobe nisu bile pedantne u transliteraciji, kao što sam imala prilike vidjeti u trećem svesku), a pisma nisu doslovno transliterirana, odnosno nije se uzela u obzir ortografija braće Grimm koju je Kuhač koristio, a koja se bitno razlikuje od ortografije suvremenog njemačkog jezika. Zbog navedenog dotične transliteracije nisu sasvim pouzdan izvor za daljnje istraživanje te je za potrebe objavljivanja kritičkih izdanja bilo potrebno napraviti transliteracije iz originalnog Kuhačevog rukopisa.

Sljedeće sistematizirano istraživanje korespondencije dogodilo se tridesetak godina kasnije. Prepoznavši iznimnu vrijednost korespondencije kao svjedočanstva vremena, Ladislav Šaban je 1980-ih godina pokrenuo projekt u sklopu kojega su ponovno načinjene transliteracije pisama iz prvog sveska korespondencije (obuhvaća godine 1860–1863). Transliteracije je napravila Vjera Katalinić te je svezak tiskan u dva dijela 1989. i 1992. godine (urednici su L. Šaban i K. Kos), a izdanje je uključivalo transliteracije i prijevode pisama koja su popraćena komentarima. Autori komentara su Vjera Katalinić, Dubravka Franković, Zdravko Blažeković i Ladislav Šaban.

Organizirano istraživanje korespondencije nastavilo se tek 2017. godine, i to u sklopu četverogodišnjeg projekta Hrvatske zaklade za znanost „Umrežavanje glazbom: promjene paradigmi u 'dugom 19. stoljeću' – od Luke Sorkočevića do Franje Ksavera Kuhača“ (NETMUS19), čija je voditeljica Vjera Katalinić. Prema istom principu, dakle transliteracije pisama, prijevodi, te komentari obradio se drugi svezak korespondencije te je pripremljen za tisak (urednice Vjera Katalinić i Sara Ries). Treći svezak koji obuhvaća godine 1870–1874. je transliteriran (uz pomoć programa *Transkribus*) te će sada biti preveden.

Važnost pravilne pripreme dokumenta

Nakon uvoza dokumenta potrebno je napraviti tzv. layout analysis. To znači pokrenuti program da automatski prepozna retke s tekstom u dokumentu. Međutim, prije pokretanja radnje valja obratiti pozornost na margine dokumenta – ako su one ispisane ispisane – program će prepoznati i retke ispisane na marginama te će doći do greške pri transliteraciji. Stoga se moraju (ručno) jasno odvojiti ta područja ili prije uvoza dokumenta ukloniti margine (ako tekst na marginama nije bitan za transliteraciju).

Primjer prikazuje urednu analizu i transliteraciju dokumenta. Program automatski podebljava red na kojem se trenutačno kako bi se korisnik lakše snašao u tekstu.

Izvoz dokumenta

Obrađene rukopise može se izvesti u različitim formatima poput datoteka Word, .txt, .xml te PDF-a. U PDF dokumentu na slikama rukopisa moguće je čak i kopirati tekst (tekst će biti kopiran na latinici).

Transkribus je odličan za istraživanja koja koriste velike količine rukom pisanih dokumenata jer pretvaranjem slike u tekst ti dokumenti postaju lako pretraživi i time pogodniji za raznovrsna istraživanja i analize. Dodatnu pažnju na pogreške u transliteraciji mora se obratiti jedino ako se programom služi za izradu kritičkih izdanja.