E-Slomšek: elektronska znanstvenokritična izdaja retorske proze 19. stoletja po standardu XML TEI

Tomaž Erjavec,[†] Matija Ogrin[‡]

†Oddelek za tehnologije znanja
Institut "Jožef Stefan"
Jamova 39,
SI-1000 Ljubljana
tomaz.erjavec@ijs.si

‡Inštitut za slovensko literaturo in literarne vede
Znanstvenoraziskovalni center SAZU
Gosposka 13
SI-1000 Ljubljana
matija.ogrin@zrc-sazu.si

E-Slomšek: an XML TEI encoding of a critical edition of 19th century Slovenian rhetoric prose

In this paper we describe the first Slovenian electronic critical edition, "The Three Sermons on Language" by Anton Martin Slomšek (1800-1862), Slovenian bishop, reformer and man of letters. The e-edition comprises digital facsimiles of the manuscripts, diplomatic transcription, critical transcription and apparatus with biblical references and notes. The e-edition is based on XML technology and all of the components are marked-up following the Text Encoding Initiative Guidelines, TEI P4. It was published as a result of cooperation between two Slovenian research institutions, the Scientific Research Centre of the Slovenian Academy of Sciences and Arts and the Jožef Stefan Institute; it is freely available at http://nl.ijs.si/e-zrc/slomsek/.

Keywords: textual criticism, critical editions, XML, Text Encoding Initiative, Anton Martin Slomšek, 19th century Slovenian literature

1. Uvod

1.1. Znanstvenokritične izdaje in elektronski medij

Znanstvenokritične izdaje se v literarnih vedah imenujejo tiste edicije, v katerih so besedila pregledana, prepisana, rekonstruirana, komentirana in naposled objavljena po načelih tekstne kritike ali ekdotike kot pomožne literarnovedne discipline. Temeljni namen tekstne kritike je, da s podrobnim proučevanjem besedila doženemo, kakšna je njegova izvirna, torej zgodovinska vsebinska, pravopisna in glasovna podoba. V ta namen se poslužuje tekstna kritika kompleksnih postopkov analize: temeljnega pomena je vedno rokopisni izvirnik besedila, nato njegove morebitne variante, ki so predmet posebne pozornosti in morajo biti skrbno dokumentirane v znanstvenem aparatu. Od variant napreduje tekstna kritika k dokumentiranju prve objave besedila, k njegovi recepciji in naposled k opombam bodisi genetične, stvarne ali jezikovne narave, ki so vključene v besedilo.

V primeru zahtevnejših, zlasti starejših besedil tekstna kritika prezentira besedilo – v idealnem primeru, če to dopuščajo materialne možnosti – v treh oblikah ali stopnjah. Prva je faksimile, torej grafična reprodukcija rokopisa. Druga stopnja je diplomatični prepis, ki je dosledna tipografska ponovitev besedila z vsemi avtorjevimi napakami, izpusti, poškodovanimi mesti ipd. Tretja stopnja ekdotične prezentacije besedila je kritični prepis. V kritičnem prepisu je besedilo popravljeno po eksplicitno formuliranih načelih. Največkrat gre za pravopisne posege (pisava bohoričica se npr. zamenja s sodobno in univerzalno gajico), posodobi se stava ločil, v nekaterih primerih seže redakcija še globlje v druge jezikovne ravni, največkrat v oblikoslovno. Obseg sprememb v kritičnem prepisu je odvisen od namena izdaje in lahko sega od neznatnih popravkov, ki redigirajo tekst po historičnih standardih njegovega časa, do posodobitve, ki besedilo približa sodobni slovenščini.

Pomen znanstvenokritičnih izdaj je v tem, da prezentirajo besedilo in raznovrstne vire na izpopolnjen, znanstveno utemeljen način, s tem pa odpirajo pota novim raziskavam in spoznanjem. Po drugi strani pa znanstvenokritične izdaje tekstov tudi same nastajajo kot rezultat obsežnih raziskav, s katerimi je bilo mogoče zbrati, urediti in komentirati tam objavljeno gradivo. Ker po eni strani znanje akumulirajo in prezentirajo, po drugi pa odpirajo pot nadaljnjemu raziskovanju, so takšne izdaje pomembne za domala vse humanistične in tudi nekatere družbene vede.

Praktična težava pri objavljanju znanstvenokritičnih izdaj v tradicionalni, tiskani obliki je, da so proizvodni stroški zelo visoki, bralska publika takšnih izdaj pa je majhna, še zlasti v Sloveniji. Elektronska znanstvenokritična izdaja ponuja odlično rešitev ne le tega ekonomskega problema, ampak razreši edicijo še mnogih drugih omejitev: brez težav je mogoče vključiti digitalizirani faksimile, brez prostorskih omejitev je mogoče vključiti v izdajo diplomatični in kritični prepis ter morebitna variantna besedila, nobenih omejitev ni glede obsega kritičnega aparata. In kar je največja prednost, z analitičnimi orodji je mogoče iskati po besedilu na različne načine, kar pri tiskanih izdajah pomeni precejšnjo izgubo časa.

Praksa objavljanja besedil na medmrežju pa je pokazala tudi šibko stran elektronskih edicij: brez standardiziranega kodiranja se trajnost ter izmenljivost

elektronskih besedil bistveno zmanjšata. Razvoj elektronskega objavljanja besedil je zato naravnan k izdelavi in uporabi standardiziranega kodiranja in označevanja besedil (mark-up), ki naj nevtralizira vpliv spreminjajoče se strojne in programske opreme na elektronske tekste. Takšen standard je označevanje besedil v jeziku XML, ki ga predlaga *Text Encoding Initiative* (TEI) s svojimi obsežnimi in detajliranimi priporočili TEI P4 (Sperberg-McQueen in Burnard, 2002). Ta standard je tudi temelj e-izdaje *Treh pridig o jeziku,* ki smo jo poimenovali e-Slomšek. Nastala je kot plod sodelovanja med Inštitutom za slovensko literaturo ZRC SAZU in Odsekom za tehnologije znanja IJS.

1.2. Slomšek in *Tri pridige o jeziku*

Anton Martin Slomšek (1800 – 1862) je bil Prešernov sodobnik; kot sošolca sta se namreč poznala že iz liceja. Slomšek je šel nato v bogoslovje in postal kot duhovnik, od l. 1846 pa kot lavantinski knezoškof, pomemben reformator slovenskega kulturnega, narodnostnega in verskega življenja. V njegovem času je bil slovenski jezik izpostavljen močnemu pritisku germanizacije, tembolj, ker je bila slovenščina tedaj kulturno in socialno podrejen jezik. Zatiranje slovenščine je bilo posebno hudo na Koroškem in Štajerskem, kar je Slomšek poznal od blizu. Videl je, kako z agresivno germanizacijo izginjajo cele vasi. Uvidel je nujnost, da dobi slovenska mladina osnovne knjige za izobrazbo v domačem jeziku, zato je napisal vrsto učbenikov in beril v slovenščini. Slomšek je podprl ustanavljanje t.i. nedeljskih šol v župniščih, ki so bile edine čisto slovenske šole v tedanjem času. Prek prijateljev Janežiča in Einspielerja je l. 1851 ustanovil Mohorjevo družbo, da so tisti, ki so se v nedeljskih šolah naučili brati, po nizkih cenah dobili slovenske knjige. Te in druge Slomškove dejavnosti so omogočile hitro rast pismenosti, bralne kulture in narodne zavesti med Slovenci. Po mnenju nekaterih literarnih zgodovinarjev (Pogačnik, 1991) se je severovzhodna Slovenija rešila pred popolno germanizacijo predvsem po Slomškovih zaslugah.

Prvovrsten dokument te dejavnosti so Slomškovi govori in pridige o jeziku. Tri od njih, med njimi znamenita pridiga *Dolžnost svoj jezik spoštovati* iz l. 1838, so l. 2001 izšle v knjigi *Tri pridige o jeziku* (Faganel, 2001). Knjiga vsebuje faksimile vseh treh pridig (rokopisa za prvi dve, prvega natisa za tretjo) in njihov diplomatični prepis (za prvi dve) ter kritični prepis (za vse tri), ki jih je pripravil Jože Faganel. Ideja te izdaje je bila, naj omogoči različne raziskovalne pristope k besedilu: za kulturnozgodovinske raziskave bo zanimiv le kritični prepis, za raziskave slovenske historične slovnice bo zanimiv predvsem diplomatični prepis, za zgodovino retorike pa tudi faksimile s členjenjem besedila za govorni nastop. Pomembne so tudi avtorjeve opombe ob robu listov, t.i. marginalije, ki razkrivajo vire, na katere se je Slomšek opiral. S temi marginalijami se odpira zlasti zanimiv pogled na Slomškovo citiranje in prevajanje iz Svetega pisma ter na Slomškov odnos do tedanjih prevodov bibličnih besedil. Vse takšne raziskovalne vidike smo z izdajo e-Slomška želeli omogočiti in olajšati.

V nadaljevanju opišemo proces dela od tiskane izdaje *Treh pridig o jeziku* do pilotske e-izdaje, dodajanja kritičnega aparata in končnih redakcij.

2. Konverzija v elektronsko obliko in dodajanje znanstvenega aparata

Priprava e-izdaje *Treh pridig o jeziku* je potekala v več stopnjah dodelave. Na vsaki od točno definiranih stopenj je bilo gradivo obogateno z določenimi oznakami. V konverziji gradiva od tekstne oblike do končne edicije, pripravljene po smernicah TEI P4, ločimo naslednje korake:

1. priprava gradiva v urejevalniku MS Word,
2. konverzija v XML,
3. konverzija v osnovno inačico z oznakami TEI,
4. prva ročna redakcija,
5. pilotska izdaja,
6. druga ročna redakcija,
7. dodajanje faksimilov,
8. dodajanje uredniških opomb, literature in hipertekstnih povezav,
9. »javna« izdaja

V nadaljevanju si bomo ogledali najpomembnejše korake tega procesa.

2.1. Pretvorba gradiva v zapis TEI

Najprej smo vse tekstne datoteke združili v en sam Wordov dokument, ki je služil kot digitalni vir za izdajo po smernicah TEI. Tega je bilo treba pretvoriti v XML. Med številnimi programi, ki to omogočajo, smo se odločili za odprtokodni program Open Office (http://www.openoffice.org/), ki emulira MS Office, a shranjuje datoteke v XML, pri čemer pa uporablja svojo lastno definicijo tipa dokumentov (DTD). Pretvorjeno datoteko smo shranili v tej primarni obliki XML.

V osnovni zapis TEI smo jo pretvorili s pomočjo kombiniranih informacij, ki smo jih dobili iz elementov (zlasti <par> in) in njihovih atributov, pa tudi vzorcev, ki se nahajajo v besedilu. Za pretvorbo je bilo potrebno najprej definirati ciljni zapis, t.j. TEI DTD za naš projekt, kar smo dosegli s parametrizacijo TEI P4 (podrobneje v 3. delu).

Parametriziran TEI DTD še vedno dopušča veliko opcij v izbiri elementov XML, mdr. številne elemente, ki jih pri e-Slomšku nismo potrebovali. Zato smo napisali »mali«, striktni DTD, ki je specializiral in omejil TEI DTD za potrebe naše izdaje in je bil pripraven za delo z urejevalnikom XML. Končna javna izdaja pa spet uporablja integralni TEI DTD. Konverzijo v »mali« DTD smo napravili s kombinacijo filtrov v jezikih XSLT (za strukturne informacije) in Perl (za vzorce), ki so odstranili tudi vse označevanje, ki ni bilo interpretirano kot oznake TEI, denimo tabulatorje, spremembo tipografije ipd. Ker Word ni enoznačen urejevalnik (isti izgled je mogoče doseči na več načinov), je s tem prišlo do določene izgube informacij. Filter Perl je vrsticam dokumentov dodal tudi številčenje; kot bomo videli, je bila natančna identifikacija vsake vrstice pomembna za povezavo obeh prepisov pridig.

Ker je avtomatsko številčenje – kakor vsaka stopnja avtomatske konverzije – pravilno označilo le večino, ne pa vseh vrstic, je bila potrebna ročna redakcija, da smo dobili striktno in konsistentno kodirano verzijo zapisa XML TEI. Pri tem smo se morali najprej odločiti za urejevalnik XML. Preizkusili smo več možnosti (Emacs, JEdit, XMLspy) in se naposled odločili za Oxygen (http://www.oxygenxml.com/), poleg ekonomskih razlogov tudi zato, ker je vanj že vgrajenih več

parametrizacij TEI, kar omogoča enostavno kreacijo novih dokumentov XML v zapisu TEI. Kot zadnjo stopnjo te priprave smo napisali še stil (stylesheet) v XSLT, s katerim se pretvori zapis XML TEI v HTML.

Ko smo imeli urejevalnik, DTD in stil, smo nadaljevali ročno redakcijo ter popravili in dopolnili avtomatsko pridobljeni zapis TEI. Najprej smo pravilno oštevilčili vse vrstice; pri tem se je številčenje knjižne izdaje *Treh pridig* na več mestih spremenilo, težavna mesta pa smo označili s <sic> za nadaljnjo diskusijo. Te težave smo rešili v več pogovorih in z nasveti, dobljenimi v debatni skupini TEI (tei-l@listserv.brown.edu). Rešitve, ki smo jih sprejeli, so zahtevale več sistemskih sprememb označevanja besedil, vendar upamo, da smo s tem dosegli konsistentno in uporabno metodo kodiranja za tovrstna besedila. V drugi ročni redakciji besedil smo te spremembe označevanja uvedli, razrešili težavna mesta <sic>, v marsičem pa tudi izpopolnili diplomatični in kritični prepis knjižne izdaje, saj vsak poglobljen stik s historičnim besedilom odpira nove vsebinske in jezikovne neznanke. Končno smo avtomatsko povezali vrstice diplomatičnega in kritičnega prepisa ter tako prišli do pilotske izdaje v zapisu XML TEI.

2.2. Težave s posebnimi znaki

Problem, na katerega bo naletela vsaka e-izdaja besedil starejšega slovenskega slovstva, so znaki za posebne črke starejših pisav, kakršna je bohoričica, ki jo je tudi Slomšek uporabljal do l. 1846, ko se je med prvimi oprijel gajice. Še pred nekaj leti bi bila predstavitev teh znakov na način, ki bi bil neodvisen od računalniške platforme, nemogoča. S široko uveljavitvijo nabora znakov Unicode pa je to postalo izvedljivo. V diplomatičnem prepisu je bilo troje takih znakov, in sicer:

- Za znak ſ, ki je bil v latiničnih pisavah v splošni rabi vse do konca 18. stoletja in še dlje, izgovarjamo pa ga kot sodobni 's', obstaja nedvoumna rešitev: v Unicodu je opisan kot LATIN SMALL LETTER LONG S s kodno številko 017F, ki pripada bloku Latin Extended-A. Gre torej za natanko ustrezen znak, ki ga tudi podpira večina računalniških platform.

- Manj idealna rešitev je dostopna za znak, ki so ga v Slomškovih časih uporabljali kot deljaj. Unicode nima povsem ustreznega znaka, vendar pa je znak, opisan kot LOW DOUBLE PRIME QUOTATION MARK s kodno številko 301F (CJK Symbols and Punctuation) zelo podoben historičnemu deljaju. Zanj smo se odločili, čeprav ga ne podpirajo vsi sistemi, ki smo jih testirali, saj pripada vzhodno azijskemu repertoarju znakov. Zato je ta znak v končni verziji HTML prikazan kot '=’.

- Težave je povzročal tudi znak LATIN CAPITAL LETTER LONG S, ki pa v Unicodu *ne obstaja,* čeprav je bil v slovenski literaturi in še kje standardni par malega dolgega ſ. Težave s tem znakom so očitno nastopale že v tiskani izdaji, kjer ga je predstavljal diagraf ‘,S’. Ko smo preizkusili vrsto možnosti, ki naj bi posnemale izgled znaka v tiskani izdaji in bi bile tudi pomensko neoporečne, smo se odločili, da si izposodimo znak ∫ z Unicodovo kodo 222B, INTEGRAL (matematični operatorji). Znak po vsem sodeč podpira večina sistemov.

XML sicer podpira definicijo entitet, ki enopomensko definirajo znak (denimo &Slong; za LATIN CAPITAL LETTER LONG S), vendar ima ta rešitev slabost, da so entitete definirane v DTD, kar dokumentom onemogoča, da bi jih distribuirali neodvisno od DTD. XML namreč razlikuje dve stopnji urejenosti dokumentov: *well-formed XML document,* ki upošteva določila glede oblike XML (pravilno zapisane in gnezdene oznake in atributi), ter *valid XML document*, ki pa mu mora dodatno biti pripisan še DTD. Ta, druga stopnja je koristna za razvoj, za samo distribucijo pa je dodtana obremenitev z DTD pogosto nepotrebna; prav zato smo posebne znake raje shranili neposredno v Unicodu.

2.3. Pilotska e-izdaja

Ko smo izoblikovali telo celotnega gradiva, smo morali pripraviti še glavo TEI, element, ki ga zahteva vsak veljaven dokument v zapisu TEI. Glava TEI vsebuje meta podatke, s katerimi je izdaja opisana analogno kolofonu v klasični knjigi, vendar s pozornostjo na elektronsko gradivo. Ko smo napisali glavo, smo jo dodali telesu in s tem je bila pilotska izdaja *Treh pridig o jeziku* končana. Poskusno smo jo objavili na medmrežju, da bi preverili odzive glede uporabnosti in preskusili konsistentnost označevanja.

Za pilotsko e-izdajo na medmrežju smo dodelali stil XSLT, tako da izdela tudi kazalo, in kar je bistveno, da omogoči vzporeden prikaz diplomatičnega in kritičnega prepisa pridig po vrsticah. Ta vzporedni prikaz obeh prepisov je – v primerjavi s tiskano izdajo – prva povsem nova pridobitev elektronske izdaje, zanimiva za strokovno bolj zainteresiranega bralca.

Na tej stopnji smo napisali tudi obsežno Poročilo o izdelavi in oznakah naše e-izdaje in izdelali vhodno spletno stran. Gradivo v XML in HTML, dokumentacija, slike faksimilov in vhodna stran so bile potem objavljene na spletu na ne-javnem URL, in ta izdaja je bila osnova za končno inačico.

2.4. Končna e-izdaja

Tretje branje celotnega gradiva je razkrilo še prenekatere napake, nastale tako v procesu označevanja kot že prisotne v knjižni izdaji. Te smo odpravili in hkrati identificirali Slomškove reference na citate iz Svetega pisma. V kritičnem prepisu smo vse te reference označili, napačne popravili (in označili kot popravljene), vse pa s hiperpovezavami usmerili k ustreznemu biblijskemu odlomku na straneh http://www.biblija.net, kjer je na voljo elektronska izdaja slovenskih bibličnih besedil. V tem procesu so bile napisane tudi uredniške opombe o nastanku pridig, njihovem kulturnem kontekstu in stvarnih problemih.

Na tej stopnji je bil v TEI zapis vključen faksimile vseh treh pridig, stil XSLT pa smo razvili tako, da prezentira tudi faksimile in glavo TEI. Po vseh teh izpopolnitvah smo celotno gradivo objavili na javnem naslovu http://nl.ijs.si/e-zrc/slomsek/.

3. Struktura zapisa TEI

Izdaja e-Slomšek je kodirana z oznakami smernic TEI P4 (Sperberg-McQueen in Burnard, 2002), ki je najbolj obsežna in široko uporabljana shema za označevanje elektronskih besedil, v celoti dostopna na medmrežju na http://www.tei-c.org/P4X/. TEI P4 sestoji iz smernic, tj. proznega opisa oznak, in formalnega dela, tj. vrste posameznih modulov (fragmentov XML DTD), ki jih je mogoče kombinirati, da bi ustvarili zaželeni DTD,

ustrezen namenom določene e-izdaje. Za e-Slomška smo uporabili naslednje module TEI:

- TEI.prose, osnovni modul za prozna besedila, ki prevzema elemente TEI.core, ta pa sestoji iz glave TEI in elementov za osnovno besedilno strukturo, kakor so razdelki, odstavki, stavki, opombe itn.
- TEI.transcr, dodatni modul oznak, potrebnih pri transkripciji primarnih virov. Definira elemente za popravljanje besedil, s katerimi označujemo avtorjeve in urednikove posege v besedilo.
- TEI.linking, dodatni modul, ki ga uporabljamo za povezave med faksimilom in prepisi.
- TEI.figures, dodatni modul, ki ga uporabljamo za zapis faksimilov.
- TEI.extensions, poljubni modul, v katerem so podane specifične razširitve ali modulacije oznak TEI, nastale za posamezno e-izdajo. Za e-Slomška smo napravili nekaj manjših sprememb: nekaterim elementom smo dodali atribut 'url', atribut 'rend' pa smo omejili na fiksni niz vrednosti. In kar je pomembneje, uvedli smo dva nova elementa, <page> in <line>, ki ju podrobneje razložimo spodaj.

S to izbiro smo dobili DTD naše e-izdaje. Parametrizacijo je mogoče določiti direktno v internem podnizu DTD v dokumentu; s pomočjo spletnega generatorja, imenovanega »TEI Pizza Chef« (http://www.tei-c.org/pizza.html), pa smo naredili tudi DTD v eni datoteki, ki se distribuira skupaj z gradivom. Kot smo omenili, smo v razvojni fazi naredili tudi pomanjšan, strogi DTD, ki je specializiral uradnega le za razvoj te izdaje.

Celotna struktura zapisa TEI je preveč kompleksna, da bi jo lahko tu predstavili, zato se bomo omejili le na najbolj zanimive vidike označevanja.

3.1. Glava TEI

Namen glave TEI je opisati označeno elektronsko besedilo, tako da so tekst sam, njegov vir, kodiranje in revizije temeljito dokumentirani. Glava TEI, njena oznaka je <teiHeader>, ima štiri glavne dele:

- Opis datoteke <fileDesc> vsebuje poln bibliografski opis same računalniške datoteke. Vsebuje tudi podatke o viru, iz katerega je bil elektronski dokument pridobljen.
- Opis označevanja <encodingDesc> opiše odnos med elektronskim tekstom in njegovim virom. Podrobno je povedano, če in kako je bilo besedilo redigirano v procesu prepisovanja, katere ravni označevanja so bile uporabljene itn.
- Profil besedila <profileDesc> vsebuje klasifikacijske in kontekstualne informacije o besedilu samem, kot so njegov predmet, osebe, ki so sodelovale pri njegovem nastanku itn. Takšen profil besedila je še posebno uporaben pri zelo strukturiranih besedilnih zbirkah, kjer je zaželeno jasno opisno izrazje.
- Zgodovina redakcij ali revizij e-izdaje <revisionDesc> poda pregled sprememb v razvoju ali nastajanju e-izdaje.

Za ilustracijo informacij, zbranih v glavi TEI, je v Sliki 1 na voljo opis profila besedila, ki opisuje uporabljene jezike in »roke«, ki so pisale ali posegale v besedilo.

3.2. Besedilo, razdelki in strani

Besedilo naše izdaje vsebuje uvodno gradivo <front>, telo besedila <body> in končno gradivo <back> z opombami in njihovimi viri.

Telo besedila obsega tri razdelke , po enega za vsako pridigo. Vsak tak razdelek nadalje vsebuje svoje , po enega za faksimile, za diplomatični prepis, za kritični prepis in za uredniške opombe. Vsak razdelek je preko atributov označen s tipom, številko pridige, poravnavo ter identifikatorjem. Vsaka pridiga vsebuje tudi »generirani razdelek« <divGen/>, ki je prazen element in služi kot sidro za avtomatsko generirani razdelek HTML z vzporednima prepisoma.

Razdelki so nato sestavljeni iz glav <head>, ki niso del pridig, pač pa uredniški naslovi iz tiskane izdaje, in posameznih strani. Struktura strani se razlikuje glede na to, ali so to strani faksimila (razložene kasneje) ali pa prepisov. Pri slednjih je vsaka stran zapisana v na novo definiranem elementu <page>, sestavljenem iz vrstic, <line>, katerih definicija je ravno tako del TEI.extensions. Oba elementa nosita atribute za povezovanje (corresp) in identifikacijo (id). Struktura dela je ilustrirana v Sliki 2, kjer podamo prvih nekaj vrstic diplomatičnega prepisa prve pridige.

Potrebno je opozoriti, da je takšno kodiranje, ki jemlje kot osnovo stran oz. vrstico, problematično, če bi hoteli označiti tudi retorično (členitev na razdelke znotraj (prepisa) ene pridige) oz. jezikovno (stavki, besede) zgradbo besedila, saj bi se srečali s problemom navzkrižnih hierarhij (Thompson in McKelvie 1997; Durusau in O'Donnell, 2001). Vendar pa alternativa, da zapisujemo strani in vrstice s praznimi elementi TEI za prelom strani in vrstic (<pb/>, <lb/>), vnaša probleme v takšno procesiranje, ki, kot pri nas, temelji na vrsticah.

3.3. Zgradba vrstic

Vrstice nosijo osnovno strukturo prepisov, saj so elementi, ki vsebujejo dejansko besedilo pridig kot tudi opombe in popravke, poleg tega pa služijo kot sidra, ki medsebojno vežejo diplomatični in kritični prepis (glej Sliko 2).

Slomškove pridige vsebujejo številne opombe na robu, ki označujejo bodisi številko razdelka znotraj pridige ali pa referenco na Sveto pismo, tj. knjigo in verz, od koder je Slomšek prevzel določen citat. Te opombe (<note>) so v e-izdaji označene z atributrom rend, ki določi njihovo pozicijo na strani. Vrstice vsebujejo tudi oznake za poudarjeno besedilo <emph>, kjer je Slomšek besedilo podčrtal, in za vrzeli <gap/>, kjer je faksimile neberljiv.

3.4. Popravki

Prepisi pridig vsebujejo tudi popravke, ki se pojavljajo kot dodatki ali izbrisi v besedilu. Popravki v diplomatičnem prepisu sledijo tistim iz faksimila, torej tistim, ki jih je naredil Slomšek sam, bodisi, da je prečrtal del besedila, ali pa je dopisal novo besedilo nad črto. Popravki v kritičnem prepisu pa so bili narejeni s strani urednika in popravljajo napake iz originala. Za obe vrsti popravkov uporabljamo enake elemente, in sicer <add>, <del> in <corr>, vendar pa jih med seboj ločimo z vrednostjo atributa hand (za definirane vrednosti glej Sliko 1). Dva primera, prvi iz diplomatičnega in drugi iz kritičnega prepisa, sta podana v Sliki 3.

3.5. Faksimile

Originali prvih dveh (rokopisnih) pridig (vsaka po 8 strani) so bili najprej fotografirani na diapozitive velikosti 6 x 9 cm, ti nato digitalizirani v resoluciji 300 dpi in shranjeni v formatu TIFF, medtem ko je bila tretja (tiskana) pridiga digitizirana neposredno v format JPEG. Zaradi lažjega prenosa smo najprej vse slike shranili v formatu JPEG, pri čemer so dimenzije približno 800 x 1400 točk in velikost približno 300 kB. Slike so bile tudi pomanjšane na približno tretjino te velikosti, da jih lahko na zaslonu prikažemo vzporedno z besedilom.

Kot ilustriramo v Sliki 4, je v zapisu TEI faksimile vsake pridige zapisan v svojem razdelku, ki vsebuje seznam <list>, čigar elementi združijo kazalce na slike posamezne strani faksimila v različnih dimenzijah. Povezave na datoteke s slikami so zapisane kot vrednost (nestandardnega) atributa url na elementu <figure>, ki tudi definira tip slike.

4. Prikaz gradiva

Čeprav smo bili v tej prvi fazi projekta osredotočeni predvsem na digitalni zapis materialov na standardiziran način, smo seveda morali tudi prikazati končni rezultat. Kot smo že omenili, je to narejeno s pomočjo stila XSLT, ki pretvori obiko XML/TEI v HTML. Poleg pretvobe v obliko, ki je čimbolj podobna knjižni, prikaže stil tudi glavo TEI (v slovenskem jeziku), izpiše kazalo, prikaže pomanjšani faksimile poleg prepisov, generira razdelke z vzporednim diplomatičnim in kritičnim prepisom in poveže pomanjšane faksimile s slikami v polni velikosti. Stran v takem izpisu HTML je podana v Dodatku I.

Zaradi "didaktične" narave projekta smo posebno pozornost namenili dokumentaciji, ki je razmeroma obširna, in, upamo, jasna. Tudi dokumentacija je zapisana po standardu TEI (TEI Lite) in prikazana v HTML s pomočjo stila XSLT, ki ga ponuja TEI. Tudi celotna priporočila TEI P4 so dodana dokumentaciji kot lokalna kopija in vsaka omemba določenega elementa TEI je povezana z razlago tega elementa v Priporočilih. Dokumentacija je tako celovita in se lahko skupaj s samim gradivom distribuira samostojno na CD-ROMu. Končno vsebuje dokumentacija tudi arhiv pisnih diskusij in komentarjev, ki dodatno pojasnjujejo nekatere odločitve v zvezi z našim označevanjem.

Prezentacija pridig v HTML je zaenkrat statična, saj je izvorna oblika v XML pretvorjena v HTML bodisi predhodno bodisi z brskalnikom, ki podpira XML (kot npr. IE Explorer) s pomočjo enega samega stila XSLT, poleg tega pa je celoten e-Slomšek – v TEI ali HTML obliki – shranjen v eni sami datoteki. V prihodnosti načrtujemo prehod na dinamičen prikaz, kjer si lahko uporabnik sam določi spletno obliko in obseg gradiva.

5. Sklep

V članku smo predstavili prvi rezultat skupnega projekta ZRC SAZU in IJS, izdelavo e-Slomška. Poudarek te prve faze skupnega dela je bil v razvoju metodologije in sheme za označevanje, ki omogoča izdelavo standardnega digitalnega tekstnokritičnega zapisa slovenske literature. Seveda pa smo se trudili zagotoviti tudi zanimiv in uporaben rezultat.

Naš trenutni rezultat je mogoče izboljšati na več načinov. Dodatno označevanje bi lahko poseglo v retorično ali jezikoslovno strukturo besedila. Zanimiva je posebej druga možnost, saj omogoča implementacijo mrežnega konkordančnika nad besedilom, kot je to že storjeno nad našim slovensko-angleškim vzporednim korpusom (Erjavec, 2002), kot tudi luščenje vzporednega (diplomatično / kritičnega) slovarja iz besedila.

Trenutno naši knjižnici "e-ZRC" http://nl.ijs.si/e-zrc/ dodajamo nova dela - zaenkrat smo izdelali še dve drugi pilotski ediciji. *Korespondenca Žige Zoisa* vsebuje petindvajset pisem, zlasti med Zoisom in Jernejem Kopitarjem. Ta edicija vsebuje faksimile (v nemščini), diplomatični prepis (transliteracija iz gotice v latinico), prevod v slovenski jezik, uredniške opombe in imensko kazalo. V eni poskusnih inačic smo na manjšem korpusu pisem v besedilu označili vsa lastna osebna imena (vsega skupaj 712 pojavnic), kar odpira možnost študija onomastike in kulturne zgodovine. Druga pilotska edicija je zbirka *Pesmi o Maji* Alojza Gradnika. Edicija vsebuje pesmi v številnih variantnih zapisih, npr. prva objava, objava v zbranem delu, t.i. krtačni odtisi, rokopisi in tipkopisi itn. ter uredniške opombe. Ta izdaja omogoča raziskavo odnosov med variantnimi zapisi.

Naši nadaljnji načrti za e-knjižnico vključujejo *Škofjeloški pasijon* in *Brižinske spomenike,* pri čemer obstajata za *Pasijon* tudi dva video posnetka, *Spomeniki* pa imajo izredno kompleksen kritični aparat.

Zahvala

Delo, predstavljeno v tem prispevku, je finančno podprlo Ministrstvo za šolstvo, znanost in šport Republike Slovenije in Slovenska akademija znanosti in umetnosti v okviru aplikativnega projekta MŠZŠ L6-3083: Znanstvenokritične izdaje v elektronskem mediju in delno v okviru raziskovalnega programa MŠZŠ PO-0542-0106: Inteligentna analiza podatkov, računalniška logika in jezikoslovje.

Literatura

Durusau, P. Brook O'Donnell, M.: Implementing Concurrent Markup in XML. Extreme Markup Languages, Conference Proceedings, Montréal, 2001.

Erjavec, T.: The IJS-ELAN Slovene-English Parallel Corpus. International Journal of Corpus Linguistics, 7(1), pp.1-20, 2002.

Faganel, J. Anton Martin Slomšek: Tri pridige o jeziku. Mohorjeva družba, Celje, 2001.

Pogačnik, J.: Kulturni pomen Slomškovega dela. Obzorja, Maribor, 1991.

Sperberg-McQueen, C. M., Burnard, L. (ur.) Text Encoding Initiative: Guidelines for Electronic Text Encoding and Interchange, TEI P4, XML-compatible edition. TEI Consortium, 2002.

Thompson, H. S., McKelvie, D.: Hyperlink Semantics for Standoff Mark-up of Read-only Documents. In: Proceedings of SGML Europe '97, Barcelona, Spain, 1997.

Viscomi, J.: Digital Facsimiles: Reading the William Blake Archive. Computers and the Humanities, 36, p. 27-48, 2002.

TRI PRIDIGE O JEZIKU. Elektronska znanstvenokritična izdaja v zapisu XML - TEI (2004-08-11) II. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

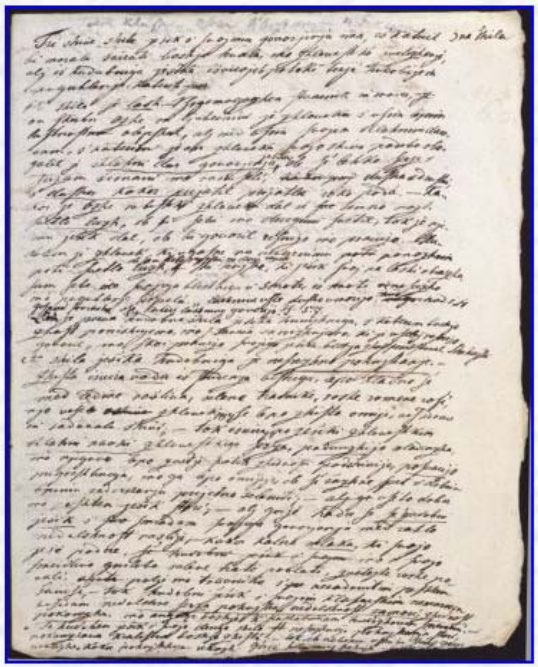
Address http://nl.ijs.si/e-zrc/slomsek/data/eSlomsek-sl2.html#sl2p Go

"jes'ik ogenj vſe krivize, kateri veſ tek naſhiga shivlenja s'ashge, ino od ſamiga 2.90

jezik ogenj vse krivice, kateri ves tek našega živl[j]enja zažge, ino od samega 2.90

"pekla prishgan. *Jak. III.5*

pekla prižgan." *Jak. [3,5-6]* 2.91 *op. ur. -22]*



Tri shive shile jes'ik v ſvojimu govorjenju ima, is'katerih *3na shila*

Tri žive žile jezik v svojemu govorjenju ima, iz katerih *3na žila*

bi morala s'virati boshja hvala, ino zhloveſhko s'velizhanje,

bi morala izvirati božja hvala ino človeško zveličanje,

alj is'hudobniga jes'ika is'virajo le potoki terji hudobije in

al' iz hudobnega jezika izvirajo le potoki trijé hudobije in 2.94 *op. ur. [23]*

pogublenja, katerih ~~per~~ 2.95

pogubljenja, katerih 2.95

1va shila je lash. – Vſigamogozhen ſtvarnik in moder, ~~je~~

... [prva] žila je laž. Vsegamogočen Stvarnik in moder

on ſkerbni Ozhe in ljubes'nivi je zhloveka s'vſim lepim

on skrbni Oče in ljubeznivi je človeka z vsem lepim

laſtnoſtim olepſhal, alj med vſim ſvojim shlahnim daro=

lastnostmi olepšal, al' med vsem svojim žlah[t]nim daro-

vam, s'katerim je on zhloveka ſvojo shivo podobo obo=

vam, s katerim je on človeka, svojo živo podobo obo-

gatil je shlahni dar govorjenja jes'ika, de ſe' lehko ſerze ſ' 2.100

gatil, je žlah[t]ni dar govorjenja jezika, da se lehko srce s 2.100

ſerzam ses'nani ino ras'veſeli, ~~kakor prej~~ duſha s'drushi,

srcem seznani ino razveseli, duša združi

s'duſho, kakor prijatel prijatlu roko poda. – Ka=

z dušo, kakor prijatelj prijatlu roko poda. – Ka-

kor je Ozhe nebeſhki zhloveku dal s'a ~~fve~~ temno nozh

kor je Oče nebeški človeku dal za temno noč

ſvetlo luzh, de bi ſebi ino drugim ſvetil, tok je nje=

My (Internet

```
    <profileDesc>
      <langUsage>
          <language id="en">angleščina</language>
          <language id="sl">slovenščina</language>
      </langUsage>
      <handList>
        <hand id="AMS" scribe="Anton Martin Slomšek" first="yes"/>
        <hand id="JFA" scribe="Jože Faganel" />
        <hand id="MOG" scribe="Matija Ogrin" />
        <hand id="TER" scribe="Tomaž Erjavec" />
      </handList>
    </profileDesc>
```

Slika 1. Primer iz glave TEI

```
 <div id="sl1d" corresp="sl1k" n="1" type="dipl">
  <head>Diplomatični prepis</head>
   <page id="sl1d-f.1" corresp="sl1f.1" n="1">
     <line id="sl1d.1" corresp="sl1k.1" n="1"
rend="right">1825. XIII</line>
     <line id="sl1d.2" corresp="sl1k.2" n="2"
        rend="center">Na 16 nedelo po Binkuſhtih.</line>
     <line id="sl1d.3" corresp="sl1k.3" n="3"
        rend="center">K’kerſhanſkimu govorjenju</line>
     <line id="sl1d.4" corresp="sl1k.4" n="4"
        rend="center">nagovor.</line>
     <line id="sl1d.5" corresp="sl1k.5" n="5">Takrat bode tebi zheſt,
kader tijiſti, kateri je tebe po&#301F;</line>
     <line id="sl1d.6" corresp="sl1k.6" n="6">vabil, tebi porezhe:
Prijatel, pomekni ſe gori.
       <note place="right">Luk. 14.</note>
     </line>
   ...
```

Slika 2. Začetek diplomatičnega prepisa prve pridige

```
<line id="sl1d.156" corresp="sl1k.156" n="156"><del hand="AMS">tudi</del>
dobro to ſvoje, in per temu <del hand="AMS">pos’abijo</del> <add
hand="AMS">blishenju ſpregledajo</add> ſvojo laſtno </line>
...
<line id="sl1k.18" corresp="sl1d.18" n="18">tok bo tebi čest v pričo
povabl<add hand="JFA">j</add>enih vseh... </line>
```

Slika 3. Primera popravkov

```
<list>
  <item corresp="sl3k-f.1" id="sl3f.1" n="1">
    <note place="inline">Faksimile, pridiga III, stran 1</note>
    <figure type="jpeg" url="img/sl3-01.jpg" />
    <figure type="thmb" url="img/sl3-01_t.jpg" />
  </item>
```

Slika 4. Zapis faksimila