

Jihočeská univerzita v Českých Budějovicích

Filozofická fakulta

Ústav bohemistiky

Bakalářská práce

Automatizace přepisu starých českých textů
pro potřebu lingvistických výzkumů

Vedoucí práce: prof. PhDr. Marie Ryantová, CSc.

Autorka práce: Eliška Pěnkavová

Studijní obor: Bohemistika – Archivnictví

Ročník: 3.

2024

Prohlašuji, že jsem autorkou této kvalifikační práce a že jsem ji vypracovala pouze s použitím pramenů a literatury uvedených v seznamu použitých zdrojů.

V Českých Budějovicích dne 3. 5. 2024

.....

Eliška Pěnkavová

Anotace

Práce se zaměřuje na problematiku automatizace transliterace a transkripce historických textů s využitím digitálních aplikací. Tématem práce je navrhnutí možností automatizace těchto dokumentů na základě konkrétních výsledků aplikovaných na biblických materiálech (rukopisu Bible padeřovské a tiskopisu Bible svatováclavské). Práce také obsahuje komparaci transkripčních pravidel a přístupů a analýzu dostupných transkripčních a transliteračních aplikací s přihlédnutím k online rozhraním zaměřujícím se na Bibli. Praktická část práce aplikuje vybranou transkripční metodiku na data Bible padeřovské a svatováclavské a shrnuje možnosti a meze případného univerzálního nástroje pro automatický přepis staré češtiny.

Klíčová slova: Automatizace, Automatická transkripce, Automatická transliterace, Transkribus, Pravopis, Fonologický vývoj, Transkripce, Transliterace, Stará čeština, Střední čeština, Nová čeština, Ediční zásady.

Annotation

The thesis is focused on the issue of automation of transliteration and transcription on historical texts using digital applications. The topic of the thesis is to present the possibilities of automation of these texts on the basis of the concrete results on the biblical materials (manuscript of the Paderov Bible and printed form of the St. Wenceslas Bible). The work also includes a comparison of transcription rules and approaches and an analysis of available transcription and transliteration applications, taking into consideration online Bible-focused interfaces. The practical part of the work applies the selected transcription methodology to the Paderov and St. Wenceslas Bible data and summarizes the possibilities and limits of a potential universal tool for automatic transcription of Old Czech.

Key words: Automatization, Automatic transcription, Automatic transliteration, Transkribus, Orthography, Phonological development, Transcription, Transliteration, Old Czech, Middle Czech, New Czech, Editorial principles.

Obsah

Seznam zkratk	7
Úvod	8
1 Přepis staro- středně- novočeských textů	9
1.1 Transliterace a transkripce	9
2 Transkripční pravidla	12
2.1 Pravopisný systém v dějinách času	12
2.1.1 Primitivní (jednoduchý) pravopis	12
2.1.2 Spřežkový pravopis	14
2.1.3 Diakritický pravopis	16
2.1.4 Bratrský pravopis	17
2.1.5 Obrozenecký (novočeský) pravopis	20
2.2 Komparace fonologických změn ve vývojových etapách českého jazyka	22
3 Automatizace přepisu starých i novodobých textů	26
3.1 Dostupné transliterační a transkripční aplikace	26
3.2 Transkribus	27
3.3 Výhody a nevýhody aplikací sloužících k automatickému přepisu	44
4 Možnost využití transliterace a transkripce	47
4.1 Tradice edičního přepisu starých českých textů	47
5 Praktická část	53
5.1 Model Bible padeřovské a svatováclavské	53
5.2 Transkripční model aneb univerzální model pro starou češtinu	54
5.2.1 Aplikace „Automatická transkripce“	55
Závěr	61
Zdroje	62
Přílohy	68
Rukopis Bible padeřovské	71
Transliterovaný model Bible padeřovské	72
Transkribovaný model (Transkribus – zkratky) Bible padeřovské	74
Transkribovaný model (Transkribus – zjednodušená transkripce) Bible padeřovské	75
Model zjednodušené transkripce Bible padeřovské	76
Výstup modelu zjednodušené automatické transkripce Bible padeřovské	77
Transkribovaný model (z aplikace „Automatická transkripce“) Bible padeřovské	79
Tiskopis Bible svatováclavské	80

Transliterovaný model Bible svatováclavské	82
Transkribovaný model (z aplikace „Automatická transkripce“) Bible svatováclavské	85

Seznam zkratek

CER	Character Error Rate
DOCX	Microsoft Word Open XML DOCument
FTP	File Transfer Protocol
HTR	Handwriting Text Recognition
IIIF	International Image Interoperability Framework
JPEG/JPG	Joint Photographic Experts Group
METS	Metadata Encoding and Transmission Standard
OCR	Optical Character Recognition
PDF	Portable Document Format
PNG	Portable Network Graphic
READ-COOP	Recognition and Enrichment of Archival Document
SCE	European Cooperative Society
TEI	Text Encoding Initiative
TXT	Text
URL	Uniform Resource Locator
XML	eXtensible Markup Language

Úvod

Historické texty představují pro lingvisty, archiváře, historiky a jiné odborníky cenný zdroj informací o vývoji jazyka, jeho gramaticy nebo obsahu obecně, avšak studium těchto textů může být náročné a zdouhavé – naráží se například na problematiku obtížnosti textu nebo jeho objemnosti, dále také může nastat problém z důvodu značných fonologických a pravopisných změn, které čeština v průběhu staletí prodělala a se kterými není každý čtenář seznámen –, z tohoto důvodu se práce snaží prosadit ideu automatizace přepisu starých českých textů.

Tato práce se zabývá problematikou automatizace transliterace a transkripce pramenů s využitím digitálních aplikací. Na základě analýzy dostupných transkripčních pravidel a přístupů jsou navrženy možnosti automatizace přepisu staré češtiny. Tyto možnosti jsou posléze testovány na biblických materiálech (rukopisu Bible padeřovské a tiskopisu Bible svatováclavské). Práce vychází z edičních zásad uvedených v dílech Jiřího Daňhelky, Josefa Vintra a Ivana Šťovíčka. Ve fonologické a pravopisné rovině práce vychází z metodologických principů stanovených v díle Jaroslava Poráka, Miroslava Komárka a Dušana Šlosara v kombinaci s elektronickým zdrojem Novým encyklopedickým slovníkem. Kromě výše uvedených literárních pramenů práce využívá i dvou archiválií – Bible svatováclavské a padeřovské. Na tyto texty jsou prakticky aplikovány transkripční pravidla a je na nich otestován navrhovaný transkripční model vytvořený jak v aplikaci Transkribus, tak i ve vlastní naprogramované aplikaci nesoucí název „Automatizace přepisu“.

Cílem práce je zhodnotit možnosti automatizace přepisu starých českých textů a navrhnout varianty, jak automaticky transliterovat prameny, a zároveň představit model pro automatickou transkripci staré češtiny, který by mohl být slibným nástrojem pro odborné výzkumy. Výsledky této práce by mohly být přínosem pro filology, historiky a archiváře, kteří se zabývají studiem pramenů, ale taktéž by práce měla být obohacením pro širokou veřejnost, která se zajímá o historické památky a vývoj českého jazyka obecně.

1 Přepis staro- středně- novočeských textů

Přepis historických textů do elektronické podoby je v dnešní době poměrně známým pojmem. Vzniká mnoho edic, které jsou buďto transliterovaným nebo transkribovaným přepisem historického pramene, ale již méně známý je fakt, že v dnešní době už začínají mezi uživatele pronikat aplikace napomáhající tomuto přepisu. A právě těmto aplikacím se budu v této práci věnovat. Tyto aplikace mají napomoci odborníkům nebo i široké veřejnosti s přečtením historických textů i bez předchozího paleografického vzdělání. Pokud by se archiváři, historici nebo i jiní znalci paleografie podíleli na vytváření modelů potřebných ke správnému chodu automatického přepisu, mohli bychom se v budoucnosti dočkat mnohonásobně rychlejšího automatického přepisu stovek pramenů potřebných například ke tvorbě slovníků, diachronních korpusů, edic a mnoho dalšího. Zatím jsou tyto programy zaměřeny primárně na transliteraci vloženého textu, v praktické části práce však uvedu možnosti a konkrétní příklady, jak by se dané aplikace daly vytěžit i po stránce získání automatické zjednodušené transkripce na použitém českém textu – vytvořím aplikaci a model sloužící k automatické transkripci.

Jiří Daňhelka ve své *Směrnici pro vydávání starších českých textů* uvádí, že samotné vydávání starších textů je výsledkem složitého a rozsáhlého přípravného úsilí, které je nutné, neboť samotný výsledek poté slouží jako prostředek poznání a tím se stává textem majícím předpoklad sloužit k dalším úvahám nebo bádáním – vydání staršího textu dá památce tzv. „nový život“. Výstup (edice a přepsaný pramen) se posléze stává památkou té doby, zpřístupněným dokumentem pro uživatele a pro editora vyvrcholením a uzavřením filologického přístupu k textu.¹ Práce a úsilí odborníků při manuálním přepisu historické památky je ocenění hodná. Přepis historických textů je časově náročný a zdlouhavý. Badatelé, kteří se zaměřují na obsah textu – například filologové –, tak tráví velkou část svého profesního života přepisy, které nejsou přímo spjaty s jejich oborem. Je důležité si uvědomit, že mnoho z nich nemělo možnost získat paleografické vzdělání, z toho důvodu se nedostanou ke složitějším a náročnějším druhům písemností. Věřím však, že existuje efektivnější způsob, jak historické texty kvalitně přepisovat i bez dlouhého manuálního přepisu, a právě tomu se budu věnovat v této práci. Tyto nástroje sice nejsou prozatím dokonalé, ale dokáží v poměrně rychlém čase kvantitativně, a hlavně i kvalitativně zpřístupnit možnost filologům, archivářům a jiným odborníkům pracovat přímo s automaticky přepsaným textem – mnohonásobně by se tak tím zrychlilo tvoření odborných příruček, edic, slovníků a jiných databází, které potřebují transliterovaný nebo transkribovaný text ke svému chodu a vzniku a jejichž vývoj je v dnešní době velmi zdlouhavý a namáhavý.

1.1 Transliterace a transkripce

Již v úvodu této kapitoly jsem využívala slov transliterace a transkripce – dvou druhů přepisů, které se využívají při zaznamenávání znění historického pramene. Pro účely této práce budu využívat termín transkripce pouze v definici jako přepis textu, tzn. převod slov

¹ DAÑHELKA, Jiří. Směrnice pro vydávání starších českých textů. In: Husitský Tábor 8, 1985, s. 285–301. (Dále jen: J. Daňhelka: Směrnice.)

z původního znění do dnešní češtiny a termín transliterace jako doslovný bezztrátový přepis pramene do latinky.²

Transliterace je mnohdy využívána ve filologických výzkumech. Jedná se totiž o doslovné zachycení vybraného dokumentu s použitím dnešních platných grafických znaků, čímž se badateli zpřístupní text zachycující původní tvar slov. Chápeme to tedy jako přepis pramene z jednoho typu fonografického písma do písma jiného. V českém prostředí je transliterace nejčastěji chápána ve smyslu převodu textu mezi cyrilicí³ a latinkou, mezi abecedou⁴ a latinkou, popřípadě mezi starším tvarem latinky do dnešního tvaru. Jedná se o bezztrátový přepis, který umožní jednoznačně identifikovat původní tvar zápisu. Soupis zásad pro náhradu některých u nás neobvyklých grafických znaků psaných cyrilicí se nachází v Pravidlech českého pravopisu,⁵ stejně tak jako i zásady pro náhradu některých u nás neobvyklých grafických znaků starší latinky.⁶ Taktéž tento soubor pravidel můžeme najít ve *Směrnici pro vydávání starších českých textů* od Jiřího Daňhelky.⁷ Normovaná pravidla pro přepis z cyrilice a abecedy do latinky jsou popsána v ISO 9 (Transliterace cyrilice do latinky – slovanské a neslovanské jazyky)⁸ a v ISO 843 (Konverze řeckých znaků do latinky).⁹ Samotné komparaci transliteračních a transkripčních pravidel se budu více věnovat ve stejnojmenné kapitole později v této práci. Dle publikace od Ivana Šťovíčka *Zásady vydávání novověkých historických pramenů z období od počátku 16. století do současnosti* se transliterací chápe přesný kritický přepis všech znaků textu za užití současné grafické soustavy písma, přičemž se zachovávají původní interpunkční znaky a všechny grafické, gramatické a pravopisné zvláštnosti textu, taktéž se mají rozepisovat všechny zkratky.¹⁰

Oproti transliteraci se častěji při přepisu textu využívá transkripce, přestože v ní dochází k úpravě původního pramene dle platných zásad pravopisné normy a grafické soustavy písma.¹¹ Nejznámější definice transkripce v lingvistice je tzv. fonetická transkripce, která pojednává

² Latinka neboli latinské písmo je druh abecedy sloužící pro zápis mnoha jazyků na celém světě. Jedná se o hláskové písmo, tzn. obsahuje znaky jak pro souhlásky, tak pro samohlásky. Převzato z: Nový encyklopedický slovník: Latinka [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/LATINKA>.

³ Cyrilice je druh slovanského písma vzniklého patrně v Bulharsku na přelomu 9. a 10. stol. za cara Symeona. Definice: Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/CYRILICE>.

⁴ Alfabeta je druh řecké abecedy. Jedná se o ustálené pořadí grafických znaků, které označují jednotlivé hlásky řečtiny. Svůj název získala od prvních dvou znaků řecké abecedy (alfa a beta). Stala se základem pro vznik dalších abeced, mezi nimi hlaholice, cyrilice a společně s vlivy etruskými také abecedy latinské. Definice: RICHTEROVÁ, Alena. Alfabeta. In: KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online]. Praha: Národní knihovna ČR, 2003- [cit. 2024-02-23]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000002275&local_base=KTD.

⁵ V sekci *Přidavek* se nachází tři přehledy zahrnující zásady transkripce (běžného přepisu) – jedná se o přepis z vybraných jazyků užívajících cyrilice, dále z vybraných jazyků užívajících zvukového písma nebo písma slabičné abecedy a posledním je kapitola z vybraných jazyků užívajících latinky. Viz HLAVSA, Zdeněk. Pravidla českého pravopisu: s dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky. Praha: Academia, 1993.

⁶ Nový encyklopedický slovník: Transliterace [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/TRANSLITERACE>.

⁷ J. Daňhelka: *Směrnice*.

⁸ ČSN ISO 9 (010185) Transliterace cyrilice do latinky – slovanské a neslovanské jazyky (platná od r. 2002).

⁹ ČSN ISO 843 (010145) Konverze řeckých znaků do latinky (platná od r. 2001).

¹⁰ ŠŤOVÍČEK, Ivan. *Zásady vydávání novověkých historických pramenů z období od počátku 16. století do současnosti: příprava vědeckých edic dokumentů ze 16.-20. století pro potřeby historiografie*. Praha: Archivní správa Ministerstva vnitra ČR, 2002, s. 50-51. (Dále jen: I. Šťovíček: *Zásady*.)

¹¹ Týž, s. 51.

o písemném záznamu zvukové řeči pomocí vhodně zvolených grafických značek (grafémů) tak, aby se výstupem stala zvuková rovina.¹² Pokud opět nahlédneme do Šťovíčkovy publikace, transkripce se rozumí přesný kritický přepis textu za užití současné grafické soustavy písma a současné pravopisné normy, užívá se současná interpunkce, ale zvláštnosti se zachovávají a grafické znaky se nahrazují znaky foneticky platnými. Používá se hlavně u písemností vzniklých po r. 1500, které se vydávají pro potřeby historické vědy. Provádí se podle přesně stanovených pravidel, z nichž všeobecná jsou společná pro všechny texty, bez ohledu na jazyk, a speciální vycházejí ze specifik jednotlivých jazyků.¹³ Vedle těchto definic termínu existuje ještě řada dalších, ty však nejsou pro účely této práce relevantní. O dalších definicích se můžeme dočíst například v Novém encyklopedickém slovníku na stránkách CzechEncy.¹⁴ Pro účely práce budu pracovat s termínem transkripce jako s přepisem z původního znění slova do tvaru dnešního,¹⁵ čemuž nejvíce odpovídá definice od Šťovíčka, popřípadě definice z Encyklopedického slovníku češtiny, která uvádí, že transkribovaný text zachycuje předpokládanou zvukovou podobu starších českých památek pomocí novočeských grafických prostředků.¹⁶

V této práci budu nejvíce pracovat s aplikací Transkribus, která slouží k automatickému čtení textů.¹⁷ Tato aplikace umožňuje za poplatek pomocí existujících modelů, popřípadě modelů uživatelem vytvořených, automaticky transliterovat jakýkoli text, který je do aplikace uživatelem nahrán. Nachází se zde však i funkce, která umožňuje aplikaci Transkribus naučit se rozpoznávat a rozepisovat zkratky, a právě to bude jedna z variant, kterou budu popisovat při nastínění možností, které umožní vytvořit modely pro automatickou transkripci. Dalšími možnostmi automatizace transkripce se budu zabývat, stejně tak jako komplexním vysvětlením aplikace Transkribus, v dalších kapitolách.

¹² Nový encyklopedický slovník: Transkripce [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/TRANSKRIPCE>.

¹³ I. Šťovíček, *Zásady*, s. 51.

¹⁴ Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/>.

¹⁵ Tzn. pokud budu využívat slova transliterace, v praxi tím myslím například slovo „kterzij“. V transkribované podobě se jedná o slovo „kteří“.

¹⁶ Na stránkách slovníku také můžeme najít srovnání pojmů transliterace a transkripce na ukázce z Alexandreidy – transliterace: „Czo gye wspachal, łamy wyzte, gyez protywno gyełt zagylte“, transkripce: „Co je vzpáchal, sami vizte, jež protivno jest zajisté“, přepis do dnešní podoby: „Co spáchal, sami vizte, jež protivno jest zajisté“. Ukázka viz Nový encyklopedický slovník: Transkripce [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/TRANSKRIPCE>.

¹⁷ Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-23]. Dostupné z: <https://www.transkribus.org/>.

2 Transkripční pravidla

2.1 Pravopisný systém v dějinách času

Historie vývoje češtiny se datuje již do 5. století n. l. příchodem slovanských kmenů na české a slovenské území. Podoba jazyka v této době však není pramenně doložena, jedná se o historickosrovnávací konstrukt. Zlomem ve vývoji jazyka bylo období Velkomoravské říše – příchod Konstantina a Metoděje (r. 863), kteří na naše území přinesli slovanský překlad základních náboženských textů a kteří sestavili pro zachycení těchto textů speciální druh abecedy – hlaholici –, čímž dali vzniknout tzv. staroslověnštině. Významné také je, že danému slovanskému dialektu, přesněji bulharsko-makedonskému, dali pevnější normu, která napomohla k vytvoření spisovného českého jazyka, neboť jej tím ustálili po stránce syntaktické, stylistické i lexikální, a tak tento čin dal vzniknout nejen zmíněné staroslověnštině, ale i české církevní slovanštině. Původní a prvotní tvar češtiny můžeme najít například v osobních a místních českých jménech v jinojazyčných textech již na přelomu 8. a 9. století. Jedním takovým bohemikem je například slovo „Zpuytignev“ – „Spytigněv“. Dle Nového encyklopedického slovníku¹⁸ je dochován tvar slova „Spytihněv“ ve tvaru „Spytihneus“ nacházející se na latinských listinách vyhotovených v 11. století. Nejstarší bohemika můžeme nalézt již v *Annales Francorum* k roku 872, dále z opisů nejstarších českých mincí z 11. století a posléze i v listinách z 11. až 13. století. Nelze také opomenout *Chronica Boemorum* od Kosmy z počátku 12. století.¹⁹ Většímu rozvoji výše zmíněných slov se dostalo při feudalizaci z důvodu umocnění síly byrokratizace, čímž došlo k rozvoji právní agendy a kronik.

2.1.1 Primitivní (jednoduchý) pravopis

Primitivní pravopis (dle Gebauera jednoduchý pravopis)²⁰ je nejstarší českou formou pravopisu. Využíval se v nejstarším období českého jazyka, tedy v období pračeštiny (konec 10. stol. až pol. 12. stol.) – tj. v nejstarším vývojovém stadiu češtiny jako samotného jazyka –, dále v rané češtině, ale jeho přesahy můžeme najít dokonce i později. Samotné pojmenování „primitivní pravopis“ není mezi odborníky oblíbené, neboť tento pravopis neměl ustálenou normu pravopisného systému – byl nejednoznačný a nesystematický –, a proto se dle některých odborníků nedá hovořit o pravopisu jako takovém. Zápis hlásek se v tuto dobu řídil převážně sluchem písařů, z toho důvodu byly výsledky rozdílné – k tomu došlo zejména kvůli nestejnému počtu latinských liter, jejichž počet byl nižší, než bylo českých fonémů. Nejednoznačnost můžeme vidět například ve slově „Mikulčice“, které by se pomocí primitivního pravopisu zapisovaly jako „Miculcice“, neboť písmeno „c“ zastupovalo hlásky „c“, „č“ i „k“, podobně u hlásky „k“ zapisovanou pomocí „c“, „k“ a „g“ (Janek – Janek, Petrec – Petrek, Vneg – Uněk).²¹ Od 12. století se začaly postupně využívat spřežky, které částečně tuto nesystematičnost odstranily.

Dle Nového encyklopedického slovníku rozlišujeme dva typy primitivního pravopisu – první užívaný k zápisu apelativ (obecných jmen), glos (neboli pokusů o překlad cizojazyčného,

¹⁸ Nový encyklopedický slovník: Primitivní pravopis [online]. [cit. 2024-03-02]. Dostupné z:

https://www.czechency.org/slovník/PRIMITIVNÍ_PRAVOPIS. (Dále jen: NES: Primitivní pravopis.)

¹⁹ ŠLOSAR, Dušan. Spisovný jazyk v dějinách české společnosti. Vyd. 3., opr. a dopl. Brno: Host, 2009, s. 12-32. (Dále jen: D. Šlosar, Spisovný jazyk.)

²⁰ GEBAUER, Jan. Příruční mluvnice jazyka českého pro učitele a studium soukromé, 3. vyd., upravil F. Trávníček, 1925, s. 444-446.

²¹ NES: Primitivní pravopis [cit. 2024-03-02].

nejčastěji latinského textu),²² přípisků a souvislých textů. Byl více nejednoznačný, ale přesto stereotypní a snadno si lze u něj domyslet dnešní tvar. Druhý typ pravopisu se využíval při zápisu proprií (částí mší, zpěvů) a byl mnohem obsáhlejší. Disponoval mnoha variantami, jak zapisovat jednu literu – například pro písmeno „c“ existovalo 15 možností, jak jej zapsat.²³

První českou dochovanou větou nalezneme na zakládací listině litoměřické kapituly, samotný vznik přípisku na listině se datuje na počátek 13. století, avšak stáří listiny je prozatím nejasné.²⁴ Vedle této památky také můžeme najít dochované bohemismy v glosách. Čeština v této době plnila funkci primárně církevní, bohoslužebnou, ale již na sklonku 13. století bylo zapotřebí, aby se mohla používat i v dalších rovinách – vyjadřování abstrakce, činitelských jmen, verbálních substantiv či adjektivizovaných přechodníků přítomných. Taktéž bylo zapotřebí v době feudalizace vytvořit nová pojmenování termínů sloužících k popisu osob a právních termínů potřebných v této době, a to se stalo buďto vytvořením zcela nového pojmu, nebo převzetím již původního slova. Často se proto přejímaly termíny z cizích jazyků, především z německého jazyka (hrabie, markrabie, rytieř, lán apod.), nebo se vytvářela nová slova, jak jsem již zmiňovala výše (dušník z latinského *proanimatus*, zeměnin aj.), popřípadě se využily starší výrazy (kmet, vladyka). Důkazem, že se čeština začala ve 13. století formovat a pronikat do společnosti, jsou již dříve zmíněné glosy, které se objevují v hojném počtu napříč různými obory – mnoho se jich nachází v latinském slovníku *Mater verborum* v kapitolách z medicíny, přírodovědy, z terminologie pohanské, zemědělské nebo řemeslné a z mnoha dalších –, důležité je však zmínit, že do zmiňované knihy připsal mnoho glos známý falzátor Václav Hanka až později v 19. století, přesto jsou však některé glosy pravé a dochované z poloviny 13. století, proto tento pramen zde zmiňuji. Původní glosy ze 13. století rozpoznal po paleografické analýze Adolf Patera, který objevil 13 nejstarších glos, dále 42člennou skupinu interlineárních glos a 285člennou mladší skupinu pravých glos, zbylé glosy, přibližně asi 1 200 přípisků, jsou obrozenecká falza.²⁵ Postupné poněmčování šlechty, dvora a země za Přemyslovců dopomohlo k ještě většímu rozvoji českého jazyka v průběhu 14. století, neboť nejenže byla v Čechách hojně rozvinutá síť církevního školství, ale samotné zavádění německého jazyka podněcovalo českou společnost ke kultivaci českého jazyka v nejrůznějších oborech, a hlavně i v literatuře. Stěžejním dílem se také samozřejmě později stala *Dalimilova kronika* ze 14. století, která je první dochovanou česky psanou veršovanou kronikou. V tomto období také vznikla díla, jako jsou *Život svatě Kateřiny*, *Hradecký rukopis* nebo *Životy svatých otců*.²⁶

Bylo zapotřebí, aby se čeština, která začala mít ambici plnit funkci spisovného jazyka, více zformovala i po stránce syntaktické umožňující využívat složitějších souvětí. Stará a nová čeština se samozřejmě v mnohém lišily.²⁷

²² D. Šlosar, *Spisovný jazyk*, s. 33.

²³ č = <c> (<Bocko>), <s> (<Bosko>), <z> (<Bozko>), <ch> (<Bochko>), <cz> (<Boczko>), <chch> (<Bochchek>), <ths> (<Bothsko>), tj. Boček, <cs> (<Bics>, tj. Bič), <sc> (<Scasta>), <tsch> (<Tschazta>), tj. Částa, <sh> (<Milish> – Milíč), <ts> (<Tsyrynin>), <sch> (<Schyrmin>), tj. Čírnín, srov. Černín, <zch> (<Zchyrminus> – Čírninus), <csh> (<Zlugacsh> – Slugáč, později Sluháč). Viz NES: Primitivní pravopis [cit. 2024-03-04].

²⁴ „Pavel dal jest Ploskovicích zem' u Vlach dal jest Dolás zem' u Bogu i sv'atému Ščepánu se dvěma dušnikoma, Bogučēja a Sedlatu.“ (Převzato z: D. Šlosar, *Spisovný jazyk*, s. 36.)

²⁵ D. Šlosar, *Spisovný jazyk*.

²⁶ O těchto dílech píše D. Šlosar v publikaci *Spisovný jazyk*.

²⁷ O rozdílu mezi užíváním minulého času, skladby vět a podobně se můžeme více dozvědět v knize *Dějiny českého jazyka* od Miroslava Komárka. (KOMÁREK, Miroslav, BLÁHA, Ondřej (ed.). *Dějiny českého jazyka*. Brno: Host, 2012. (Dále jen: M. Komárek, *Dějiny českého jazyka*.)

Soupis změn, ke kterým došlo během období primitivního pravopisu od 11. století až do pol. 13. století, můžeme najít níže v tabulce (Tabulka 1) převzaté z Nového encyklopedického slovníku.²⁸ Zkratky v horní části tabulky symbolizují, v jaké podobě, formě nebo rukopisu byly nalezeny – B = bohemika, G = glosy, P = přípisky, NeP = Nekrolog podlažický, ST = souvislé texty. Tučně jsou vyznačeny nejčastější zápisy dané hlásky v daném období, které jsou použity v aplikaci „Automatická transkripce“, která je výstupem této práce.

fonémy		1. pol.	1. pol.	2. pol.	1. pol.	1. pol.	1. pol.	2. pol.	2. pol.	2. pol.
	11. st.	12. st.	12. st.	12. st.	13. st.	13. st.	13. st.	13. st.	13. st.	13. st.
	B	B	G	B	B	NeP	P	B	GP	ST
c	cs	cs, c	c, cz	c , cs, z, ch, cz, tz, zc	c, tz, z, ch , ts, cz, zz, thz	c	c	cz, z, c, ch, tz, zc, zt, s	c , ch, s, cz, cc, cs	c, ch
č	c	c , cs, c h	c, f	c, ch, sch, szh, cs, cz, sc, tz, z, sh, csh, xh, th, cc	c, ch, s, sch, cz, tsch, ts, sc, z, tz, chch, zch	c	c	c, ch, z, sch, cs, s, sz, ts, tsch, cz, tsh, sh, zc, ths	c, ch, s, cs, sc, zc, est, cc	c, ch
r ²⁹	r	r	r	r	r, rs	r	-	r, rs, s, rz, rsh	r, rs, rz	r, rs, rz, zr
s ³⁰	s, z, s s	s, z	f, z	s, z, zz, c, ss	s, z, zz, ss, sz, zs, sc, cz	s, z, ss	s	s, z, zs, zz, ss, ch, sh, sc, sch, sz	s, z, zz, ssz, zsz, sz, szs	s, z
š	s, ss	s, ss	f, z	s, ss, sc, hs	s, ss, sc, sh, sch, si, z	s, ss, z	s	s, ss, z, sch, ts, ch	s, ss, z	s, ss
z ³¹	z	z	z	z, s	z, s	z	z	z, zz, s, sc, czs, zs, zh	z, zz, s	z, s
ž	-	s, z	z, f	s	s, z, sch	s, z	-	s, z	z, s	z, s

Tabulka 1 – Primitivní pravopis.

2.1.2 Spřežkový pravopis

Ve 14. století byl původní jednoduchý primitivní pravopis nahrazen pravopisem spřežkovým – první oficiálním pravopisným systémem. Důvodem jeho vzniku byl stále rostoucí význam psaného jazyka. Spřežkový pravopis dal možnost odlišit hlásky, které se v latině nevyskytovaly, neboť jak jsem již psala výše, v latinské abecedě se nacházelo méně

²⁸ NES: Primitivní pravopis [cit. 2024-03-02].

²⁹ Od roku 1237 se zapisovalo jako „ř“.

³⁰ Nebo také s', do začátku 13. století.

³¹ Nebo také z', do začátku 13. století.

grafémů, než bylo českých fonémů. Tento nedostatek pravopis řešil pomocí spřežek.³² Jsou zachované dva typy tohoto pravopisu – starší (složitější) a mladší (jednodušší) typ.

Pravopis starší, který vznikl na počátku 14. století, byl přesnější a více rozlišoval znělé sykavky od neznělých a tupé od ostrých. Používal se pouze krátkou dobu a je dochován jen v několika písemnostech. Tento pravopis využíval následujících spřežek: zz = s, ff = š, z = z, s = ž, cz = c, chz = č. Další hlásky se značily takto: ie = ě, c (před a, u, o, r) = k, ch = ch, rs = ř, g (před i, e) = j.³³ Blíže změny popisuje tabulka (Tabulka 2)³⁴ níže. Kvantita vokálů se označovala jen zřídka, buď vertikálním digrafem „ě“, nebo diakritickým znaménkem „á“. Čárka nad „í“, které se často psalo bez tečky, sloužila k formálnímu rozlišení litery „í“ od dvou nebo tří nožiček sousedních liter „u“, „n“ a „m“.³⁵

grafém	foném	příklady
cz	c	otecz
chz	č	chzazz
zz	s	zzam, zzie
ff, f	š	lepffim, wfem
z	z	wzhoru, mezi
s, f	ž	siwot
rs, rf, ojedinele rz	ř	morse, hrziech
ch	ch	ducha
ie, ye	ě, ie	wieczce, chtyel
c zvl. před a, o, u, r	k	cozztelu
g; v přejetých slovech	j; g	gedno, gmagie; z galiley
w, ojedinele u	v	wffelikā, sŷuotēm
u, v	u	uteczzi, vzrsie

Tabulka 2 – Starší spřežkový pravopis.

Oproti tomu mladší spřežkový pravopis byl komfortnější a jednodušší pro jeho uživatele, avšak rozlišoval pouze znělé hlásky od neznělých, nerozlišoval sykavkové dvojice (s-š, z-ž, c-č) – tento typ i přes svou nepřesnost posléze převládl. Spřežkový systém vypadal takto: z = z/ž, f/ff = s/š, cz = c/č, s = s/š, rz = ř. Kvantita vokálů se zapisovalo stále pouze zřídka pomocí zdvojení písmene „aa“, nebo čárky „á“. Jak je vidět na ukázce spřežek, pravopis nebyl jednotný, což dokazuje i tabulka (Tabulka 3)³⁶ níže. Mladším spřežkovým pravopisem byly také psány první pokusy o překlad textu evangelií z první poloviny 14. století.

grafém	foném	příklady
cz	c, č	czo, czas
s, f	s, š	fama, hruška
ss, ff	s, š	nebeffa, maff
rz, rf, rs	ř	przied, prfied

³² Spřežka je kombinace dvou a více písmen, která se v českém textu běžně nevyskytovala vedle sebe. Převzato z: Nový encyklopedický slovník: Spřežkový pravopis [online]. [cit. 2024-03-05]. Dostupné z: https://www.czechency.org/slovník/SPŘEŽKOVÝ_PRAVOPIS (Dále jen: NES: Spřežkový pravopis.)

³³ D. Šlosar, Spisovný jazyk.

³⁴ NES: Spřežkový pravopis [cit. 2024-03-05].

³⁵ Týž [cit. 2024-03-05].

³⁶ Týž [cit. 2024-03-05].

z	z, ž	iazik, když
w, méně v, u	v	welmy, velmi, ziuota
u, méně w, v	u	druhé, wyyty, vcho

Tabulka 3 – Mladší spřežkový pravopis.

Zároveň s výše zmíněným pravopisem se ve 14. století začala do soustavy hlásek dostávat různá diakritická znaménka především označující délku samohlásek. Jejich převzetí následovně dalo vzniknout diakritickému pravopisu. Velkému rozvoji češtiny se v tomto století dostalo díky Karlu IV., který podporoval českou kulturu, literaturu a jazyk. Taktéž ve 14. století vzniklo velkolepé slovníkářské dílo mistra Klareta, ve kterém shrnul starší českou terminologii. Důležité bylo i dílo Tomáše ze Štítného rozebírající křesťanský život. Samotný spřežkový pravopis se zachoval v jednotlivostech až do počátku 19. století.

Komparaci různých typů písemných památek ze 14. století postihující nejednoznačnost zápisu spřežek v tomto období a zároveň informace o výskytu a využívání interpunkce během období spřežkového pravopisu můžeme nalézt na stránkách Nového encyklopedického slovníku.³⁷

Spřežkovým pravopisem je také psaná Bible padeřovská, kterou podrobím zkoumání v praktické části této práce, kde na ni bude aplikován automatický transkripční model.

2.1.3 Diakritický pravopis

Mistr Jan Hus se v 1. polovině 15. století pravděpodobně zasloužil o tzv. diakritický pravopis. Jak jsem již psala výše, do dřívějšího spřežkového pravopisu začalo již ve 14. století přibývat označování délek, které jsou jedním ze symbolů diakritického pravopisu – právě tento pravopis řeší výše zmíněný nestejný počet latinských grafémů a českých fonémů pomocí diakritiky – čárek a teček nad písmeny. Stěžejním dílem tehdejší doby a také dílem pokládající základy současnému pravopisu se stal Husův traktát *Ortographia Bohemica* (nejspíše 1406).

Tečkou (tzv. nabodeničkem) nad písmeny se začaly označovat měkké souhlásky (hlásky <č>, <š>, <ž>, <ř>, <d>, <t>, <n>, <ň>, <í> označovaly dnešní <č>, <š>, <ž>, <ř>, <d>, <t>, <ň>)³⁸ a tvrdé l [l]. Později tečku nahradil háček. Délka samohlásek a slabikotvorných likvid (r, l) se označovala čárkou (gracilis virgula) nad písmenem (á, é, í, ...). Taktéž Hus vyžadoval rozlišování písmen „i“ a „y“, což se dochovalo až dodnes. Cílem Husovy pravopisné reformy bylo zjednodušení psaní pomocí jednoho písmene namísto předchozích spřežek – jediná spřežka, která se zachovala, je spřežka „ch“ používaná do dnešních dní. Hlavním účelem zavedení pravopisu bylo zajistit rychlejší, levnější a efektivnější opisování díky novému pravopisu. Pravopis však nebyl i přes nově vzniklý jednodušší systém ustálený, protože stále na území Čech docházelo k různým kombinacím již vzniklých pravopisných systémů, nebo k jejich nedodržování. Ve vlastních místních jménech³⁹ z 15. století můžeme najít zachované některé spřežky (<cz> = c, č; <tz> = c; <z> = z, ž; <rz> = ř; <s>, <f> = s; <ss>, <ff>, <sf>, <sch> = š).⁴⁰

³⁷ NES: Spřežkový pravopis [cit. 2024-03-05].

³⁸ Nový encyklopedický slovník: Diakritický pravopis [online]. [cit. 2024-03-05]. Dostupné z: https://www.czechency.org/slovník/DIAKRITICKÝ_PRAVOPIŠ. (Dále jen: NES: Diakritický pravopis.)

³⁹ O vlastních místních jménech se můžeme více dočíst na stránkách Nového encyklopedického slovníku, stejně tak jako o ponechání užívání některých cizojazyčných spřežek v určitých písemnostech. Odkaz: NES: Diakritický pravopis. [cit. 2024-03-05].

⁴⁰ Pokud se však hláska „š“ zapisovala pomocí spřežky „ff“, mnohdy takto zůstala a tečka značící měkkost se nad ni nezapisovala, neboť to působilo potíže. Viz NES: Diakritický pravopis. [cit. 2024-03-05].

Husitské války měly vliv na rozvoj českého jazyka – zrychlily jej. Bylo totiž zapotřebí s ním kázat, stal se tak bohoslužebným jazykem, zanikly uměle udržované zastaralé jevy, a také se spisovný jazyk přiblížil jazyku lidovému. Na druhou stranu kvůli období husitství upadl zájem o veršované legendy. Od 15. století získala čeština taktéž význam mezinárodní. Začala se užívat jako úřední jazyk ve Slezsku, dále jako diplomatický jazyk a v neposlední řadě jako kulturní jazyk v Polsku a Uhrách na dvoře Matyáše Korvína. Také zesílil vliv českého jazyka na Slovensku kvůli vojskům Jana Jiskry z Brandýsa za husitských válek a díky absolventům Karlovy univerzity.

2.1.4 Bratrský pravopis

Již na konci 15. století začal vcházet do českého prostředí pravopis bratrský, který byl využíván souběžně vedle pravopisu diakritického a spřežkového a který nakonec starší pravopisné systémy po vzájemném ovlivňování vytlačil.⁴¹ Tento pravopis byl inovativní, neboť ještě více zjednodušil a zmenšil počet používaných liter a tím šetřil více místa a času při psaní. Avšak opět docházelo k nejednoznačnosti zápisu, což dokazuje příklad z Nového encyklopedického slovníku, který zobrazuje některé z 21 možných hypotetických možností zápisu slova „odepříti“.⁴² Nejednoznačnost ukazuje i nepravidelné zaznamenání kvantity a diakritiky u hlásek a další jevy, které později ovlivnil rozvíjející se knihtisk (od poloviny 15. století) a který byl důležitým faktorem pro následnou ustálenost a jednotnost pravopisné normy.

Název pravopisu – bratrský – by mohl evokovat, že vznikl v tiscích Jednoty bratrské, avšak přesto dle Taťány Vykypělové⁴³ vzešel z kulturního středu českého utrakvismu – teoretické základy tohoto pravopisu byly představeny v české mluvnici *Grammatyka česká* (1533), jejíž autoři byli stoupenci českého utrakvismu. Samotný název pravopisu byl odvozen až později od bratrských tisků, které se staly vzorem jazykové správnosti – někteří odborníci dokonce nazývají bratrský pravopis jako tiskařský z důvodu hojného užívání po vzniku knihtisku. Toto pojmenování potvrzuje i fakt, že v době rozrůstajícího se knihtisku se knihy, například Bible, začaly tisknout ve velkém měřítku – dřívější dlouholeté ruční opisování se mnohonásobně zrychlilo a tiskárny byly schopny během stejné doby vytisknout několik set až tisíc exemplářů této knihy, čímž se kniha stala levnější a dostupnější.⁴⁴

Inventář bratrského pravopisu vypadal takto: <a>, <á>, , <c>, <č>, <d>, <d̄>, <e>, <é>, <f>, <g>, <h>, <i>, <ij>, <y>, <ý>, <k>, <l>, <m>, <n>, <ň>, <o>, <p>, <r>, <ř>, <ř̄>, <ř̄̄>, <s>, <ff>, <fs>, <š̄>, <t>, <t̄>, <v>, <u>, <ú>, <w>, <x>, <z>, <ž̄>.⁴⁵

⁴¹ V průběhu 15. století a 1. pol. 16. století se stále objevovaly hybridní zápisy kombinující spřežkový a diakritický pravopis, jako je <rž> pro „ř“ nebo <cž> pro „č“. V 15. století také docházelo k tomu, že v mluveném jazyce se začaly využívat určité pravopisné změny dříve, než se dostaly do jazyka psaného, jednalo se o změny: ie > í, uo > ú <ú>, ú > ou, ý > ej, é > í a další změny. Zdroj: Nový encyklopedický slovník: Bratrský pravopis [online]. [cit. 2024-03-06]. Dostupné z: https://www.czechency.org/slovník/BRATRSKÝ_PRAVOPIŠ. (Dále jen: NES: Bratrský pravopis.)

⁴² Např.: <odeprzieti>, <odeprzieti>, <odepřieti>, <odeprzijti>, <odeprzjti>, <odepřjti>, <odeprzijeti>, <odeprzijeti>, <odepřijeti>, <odeprzjti> a další. Zdroj: Týž.

⁴³ VYKYPĚLOVÁ Tat'ána. Wege zum Neutschechischen. Studien zur Geschichte der tschechischen Schriftsprache. Hamburg: Verlag Dr. Kovač. 2013.

⁴⁴ PORÁK, Jaroslav. Humanistická čeština: hláskosloví a pravopis. Acta Universitatis Carolinae. Praha: Univerzita Karlova, 1983. (Dále jen: J. Porák, Humanistická čeština.)

⁴⁵ J. Porák, Humanistická čeština.

Bratrský pravopis lze obecně rozdělit na dvě období, obdobně jako tomu bylo u primitivního pravopisu a předchozího spřežkového pravopisu. První etapa pravopisu se datuje od konce 15. století, kdy ještě nebyla plně ustálená forma a pravopis kolísal, a na druhé období od poloviny do konce 16. století, při němž došlo k ukotvení pravidel díky veleslavínskému a kralickému úzu.

V husitské době bylo zapotřebí, aby čeština začala plnit funkci agitační – v mluvené i psané formě –, v písemné podobě i manifestační a dále také funkci literární (vzdělávací). V 15. století totiž znalost písma a písemná kultura obecně již nebyla určena pouze pro duchovenstvo a šlechtu, ale začala pronikat i do řad měšťanstva a širších lidových vrstev. Aby se pravopis dobře šířil a byl lehce zapamatovatelný, Mistr Jan Hus dokonce sestavil abecedu, v níž každé písmeno pojmenoval slovem, která dohromady tvoří souvislý text jako mnemotechnická pomůcka (viz obrázek (Obrázek 29) v příloze této práce).⁴⁶

Od 16. a v 17. století začala být čeština otevřená k přejatým slovům různého původu. Přejala například z německého jazyka slova „handlovati“, „gruntovní“, „koštovati“, „šichtmistr“ a z latinského jazyka „exemplum“, „apelovati“ nebo „kalendář“. Taktéž se přejímalo i z románských jazyků. Slovní zásoba se tím doplňovala o slova z oblasti hornictví, odborné a administrativní, řemeslné výroby, nebo z oblasti vojenské či námořnické. Humanistická společnost na našem území si dala za úkol srovnat český jazyk s latinou a řečtinou. V 16. století bylo napsáno několik gramatik – psal je například Jan Blahoslav nebo Vavřinec Benedikt z Nudožer. Vedle mnohých stylistických a rétorických příruček také vyšly dva slovníky Daniela Adama z Veleslavína vydané roku 1598 shrnující dobovou slovní zásobu a terminologii. Všechny tyto zmíněné publikace vyzdvihující český jazyk vyvrcholily Bibli kralickou – tištěným českým překladem Bible – z let 1579-1593. Čeština v Bibli kralické je kompromisem mezi západní a středočeskou češtinou s nářečím na východní Moravě – v knize se tak užívá hlásek „ou“ namísto „au“ (mauka – mouka), „ey“ za „aj“ (původní dey – bratrské daj), ale zachovávají se hlásky „ý“ (strýc) a „é“ (dobré mléko). Také došlo v určitých slovech k odstranění protetického v- (voko – oko). Rozeznávání dvojího „l/ľ“ zůstalo obdobné původnímu pravopisu. Díky Bibli kralické byl přijat nový pravopis – již zmíněný bratrský pravopis –, který byl využíván až do počátku 19. století. Bratrský pravopis přejal mnohé z diakritického pravopisu, zachoval také spřežku „ľ“ – současné „š“ – to mohlo být psáno také jako „š“, nebo „š“. V inventáři pravopisu se zachovala až dodnes i spřežka „ch“ díky tomu, že přetrvávalo její používání i v latinských textech. Dalším rysem pravopisu byla hláska „w“, která nahradila „v“ (vy – wy), písmeno „u“ se na začátku slova psalo jako „v“ (ubohý – vbohý), stejně jako v iniciální slovní pozici. Znak „y“ nahradilo „j“, avšak pouze pokud se nacházelo ve dvojhlásce (kraj – kray) a hláska „j“ se před samohláskou zapisovala jako „g“ (jako – gako) a samotné „j“ zastupovalo grafém „í“ (dají – dagj), ale „i/í“ psané po „c, z, s“ bylo psáno pomocí „y/ý“ (cizí – cyzý). Taktéž, jelikož se jednalo o kompromisní řešení vzhledem k předešlým systémům, bratrský pravopis změnil tečku značící měkkost nad hláskami na háček, neboť byl praktičtější pro rozrůstající se tisk – lépe se tiskl a byl viditelný. Více můžeme vidět v tabulce (Tabulka 4) níže. Některé části změn v bratrském pravopisu, například psaní hlásky „š“, se řídily specifickými pravidly, to popisuje J. Porák ve své publikaci – například psaní skupiny „šľ“, která se využívala převážně na konci slova a podobně.⁴⁷ Více se však tato práce této specifčnosti dotýkat nebude, neboť stěžejními daty pro praktický výstup jsou samotné změny a varianty zápisu, nikoli jejich jednotlivá využívání.

⁴⁶ Týž.

⁴⁷ J. Porák, Humanistická čeština.

fonémy	majuskulní grafémy	minuskulní grafémy
s	S γ	s f
š	Sf γ f	ff š sf
l	L	l ł
j	G (J/I)	g y i
g	G Ġ (Ĝ)	g ġ (ğ)
ř	Ř Rž R	ř rž [rz]
ž	Ž Zi Z Zě	ž zi zě
č	Č Cž C	č cž cz [tc tč]
b	B	b b
ě	E JE	ě ie
u-	V- U-	v- u-
ú	V U	ũ ũ ú uo ü
ou	Au	au
í	J I	ij j ie
i	I J Y	i y
ď t ň	Ď Ď Ć	ď t ň
	Ti Di Ni	di ti ni
	T D N	d t n
h	H	h χ

Tabulka 4 – Přehled některých grafémů (alografů) bratrského (tiskařského) pravopisu.⁴⁸
Hlásky v hranatých závorkách byly převzaty z publikace od J. Poráka.⁴⁹

Ve 2. pol. 16. století došlo k revizi systému od J. Blahoslava, který doplnil do existujícího bratrského pravopisu některé alomorfy: „ů“ pro psaní dlouhého „ú“ vzniklého monoftongizací z diftongu „uo“; „j“ pro dlouhé „í“, které se vyvinulo monoftongizací z diftongu „ie“ – to se však neujalo a nahradilo jej pouhé „j“; grafém „ğ“ sloužil pro označení fonému „g“, přičemž tečka nad písmenem sloužila k jeho odlišení od grafému „g“ psané pomocí fonému „j“.

V průběhu 16. století postupně dochází k rozchodu dvou tendencí bratrského pravopisu – „písařské“, která je značně rozkolísaná a která se drží nejednotné tradice spřežkově-diakritické, a tendence „tiskařské“, klonící se, nebo naopak odkloňující se od bratrského pravopisu. Jiní odborníci tyto tendence pojmenovávají souslovími „texty spíše soukromého rázu“ („pro foro interno“) a „texty určené širšímu publiku“ („pro foro externo“), další odborníci dělí pravopis na „písemnosti psané novogotickou kurzívou nebo polokurzívou“ a na „písemnosti psané novogotickým knižním písmem“.⁵⁰ V Novém encyklopedickém slovníku⁵¹ také můžeme najít kapitoly o distribuci a užívání majuskul a interpunkčních znamének v tomto období – v této práci to však již nebude dále popisováno.

Po bitvě na Bílé hoře (1620) se vývoj češtiny zpomalil, přestože stále v některých okruzích vrcholil – například v díle Jana Amose Komenského. Jak píše Dušan Šlosar: „*Slibný rozvoj spisovné češtiny, jaký jsme konstatovali pro období 16. století, byl podloměn v důsledku*

⁴⁸ Přehled některých grafémů (alografů) bratrského (tiskařského) pravopisu 17. a 18. stol. Případy odlišené tučným písmem jsou ojedinělé nebo typické pro písařský úzus. Litery <đ> a <χ> představují alografy grafémů a <h>, které se různě objevují v tiskařském úzu tohoto období – jednalo se o nedotažené „horní bráško“. Zdroj: NES: Bratrský pravopis [cit. 2024-03-05].

⁴⁹ J. Porák, Humanistická čeština.

⁵⁰ NES: Bratrský pravopis [cit. 2024-03-05].

⁵¹ NES: Bratrský pravopis [cit. 2024-03-05].

událostí, jež následovaly po Bílé hoře. Toto období bývá obvykle charakterizováno jako doba úpadku českého spisovného jazyka. To je zjednodušující pohled, který nevystihuje podstatu tehdejšího vývoje.⁵² Úpadek češtiny, přestože to není přesné pojmenování její tehdejší podoby, byl zapříčiněn odchodem velké části české inteligence do exilu a úbytkem čtenářů, kteří četli českou literaturu, tj. drobné a střední šlechty a měšťanstva, neboť z dob předbělohorských stále rostla síla poněmčování šlechty a měst – nově příchozí šlechtické rody totiž neměly o domácí kulturní tradice velký zájem, avšak i přesto český jazyk stále plnil funkci jazyka úředního. Přestože dle M. Komárka český jazyk v 17. století upadá, v knize se dále autor zmiňuje, že v 2. polovině 17. století vyšla obsáhlá gramatika s neologismy od V. J. Rosy (*Čechořečnost seu Grammatica linguae Bohemicae*, tiskem 1672) jako snaha o očištění a povznesení českého jazyka – tento fakt dokládá citací výše.⁵³ Vedle tohoto činu vyšla i tiskem známá náměšťská mluvnice (*Grammatyka česká v dvojí stránce*, první vydání 1533) dokonce v pěti vydáních. Bratrský pravopis byl stále užíván v některých dílech exulantů, dále také i v českém překladu Bible – v Bibli svatováclavské, kterou využijí jako příklad v praktické části této práce. Dále se s pravopisem mohli lidé setkat v různých vydáních starších textů – například spisech J. A. Komenského od jezuitů. Tento pravopis byl také stále používán v kronikách – například katolické kroniky Pešiny a Beckovského –, v kázáních a světské literatuře. Samotná náboženská poezie vrcholila od 2. poloviny 17. století autory jako byl Bedřich Bridel (*Co Bůh? Člověk?*), Felix Kadlinský aj. a byla inspirací pro pozdější obrozeneckou češtinu. Nejen u nás se v některých úředních kruzích na panstvích cizích vrchností a na venkově psalo a mluvilo českým jazykem, ale dělo se tak i na Slovensku v protestantské části státu – vycházela zde například česká literatura od slovenského autora Daniela Sinapiuse-Horničky. V barokní době 2. poloviny 17. století také vyšlo mnoho gramatik a příruček, inteligence se totiž začala více zajímat o teoretickou část českého jazyka. Do této doby můžeme řadit například dílo jezuita Jana Drachovského (*Grammatica Boëmica*, 1660), Jiřího Konstance (*Brus jazyka českého*, 1667), nebo Matěje Václava Šteyera, překladatele Bible svatováclavské (*Žáček aneb Výborně dobrý způsob, jak se má dobře po česku psáti neb tisknouti*, 1668). Do čela těchto děl se dostal autor Pavel Doležal se svou publikací *Grammatica Slavico-Bohemica* (1746), která se stala jednou z nejvýznamnějších mluvnic pro předobrozenecké období.⁵⁴

Více podrobně se tomuto období věnuje publikace od Jaroslava Poráka – *Humanistická čeština*.⁵⁵ Porák zmiňuje, že dané českobratské (humanistické) období bylo pro vývoj spisovné češtiny klíčové. Snaha najít sounáležitost s požadavky a stylovými ideály probíhajícího humanismu vypomohla k dříve zmíněnému ustálení a vývoji jazyka. O gramatikách a českých mluvnicích píše také kniha *Česká bible v dějinách národního písemnictví* od Vladimíra Kyase, která zaznamenává jejich vývoj a vliv na českou společnost.⁵⁶

2.1.5 Obrozenecký (novočeský) pravopis

Díky zrušení nevolnictví v poslední čtvrtině 18. století (1781) se začala vrstva tehdejších poddaných a venkovského obyvatelstva stěhovat do měst, český jazyk byl však stále pod náporom ještě silnější germanizace a němčina se tak stala jediným úředním jazykem, čímž

⁵² D. Šlosar, *Spisovný jazyk*, s. 91.

⁵³ M. Komárek, *Dějiny českého jazyka*.

⁵⁴ Týž.

⁵⁵ J. Porák, *Humanistická čeština*.

⁵⁶ KYAS, Vladimír. *Česká bible v dějinách národního písemnictví*. Praha: Vyšehrad, 1997. (Dále jen: V. Kyas: *Česká bible*.)

docházelo k míšení těchto dvou jazyků. Rozvoj hospodářství a společenské vrstvy dal vzniknout na přelomu 18. a 19. století tzv. národnímu obrození, při němž došlo k velkému rozmachu i v rovině lingvistické a literární. Čeština se musela přizpůsobit novým potřebám české společnosti – muselo dojít k dotvoření gramatických norem spisovného jazyka, k ustálení slovní zásoby a k vytvoření a ustálení pevných jazykových prostředků. Z toho důvodu také začalo vznikat mnoho nových mluvnic. Důležitým milníkem pro vznik obrozeneckého pravopisu byla gramatika Josefa Dobrovského (vychází ve dvou vydáních r. 1809 a 1819) zpřisňující normy, které uváděl František Jan Tomsa ve své gramatice (*Böhmische Sprachlehre*, 1782)⁵⁷ a omezující „knižnost“ Františka Martina Pelcla využívajícího ji ve své gramatice (*Grundsätze der Böhmischen Grammatik*, 1795/1796). Dále také vyšla mluvnice Jana Nejedlého (*Böhmische Grammatik*, 1804), ovlivněná názory Dobrovského.⁵⁸

Pravopis Josefa Dobrovského (tzv. analogická reforma) obsahoval několik změn – jednalo se například o doporučení užívat pouze jedno „l“, rozlišovat po souhláskách „s/z“ psaní „i/y“ a „í<j>/ý“ (původní: psy štěkali – vidjm psy, změna na: psi štěkali – vidjm psy) – tedy zapisování podle morfologické analogie (např. vzorů jmen). Po souhlásce „c“ Dobrovský zavedl psaní pouze „i/í<j>“ (vojácy – vojáci), v ostatních záležitostech se autor přikláněl k staršímu bratrskému pravopisu, který je popsán v předchozí podkapitole. Dobrovský také dával přednost ohebným přechodníkům před neohebnými, dále nepřipouštěl u přivlastňovacích přídavných jmen v jednotném čísle složené tvary (kromě 7. pádu) (z matčiné / matčiný – z matčiny strany) a v 1. pádě množného čísla středního rodu tvrdých zájmen a přídavných jmen uznával pouze tvary „-a/-á“ (ty krásný / krásné – ta krásná města). Dalším významným dílem pro vývoj češtiny byl *Slovník česko-německý* (1835-1839) Josefa Jungmana, který do tehdejšího jazyka přinesl mnoho nových slov.

Samotný návrh Dobrovského analogické reformy vyvolal ve společnosti diskusi. Odborníci se rozdělili na dvě skupiny – tzv. „iotisté“ (V. Hanka, J. Jungmann) a „ypsilonisté“ (J. Nejedlý, J. L. Ziegler) –, kteří buď podporovali, nebo zavrhovali pravopis Dobrovského. O prosazení prvního systému se posléze zasloužil V. Hanka, který jej vydal v knize *Pravopis český podle základu grammatiky J. Dobrowského* (1817). V tomto spisu ještě zavedl psaní „u“ a „ú“ tam, kde se do té doby psalo „v“ (dřívější „vzřjti“ – nové „uzřjti“). Vzhledem k neustálému odporu „ypsilonistů“, kteří nepodporovali Dobrovského analogický pravopis, úřední kruhy v roce 1823 rozhodly, že se školní knihy budou tisknout bratrským pravopisem a knihy pro dospělé analogickým, avšak rozhodnutí nevedlo ke kýženému konci a boje mezi stranami ještě vzrostly. Spor byl ukončen až ve 40. letech 19. století, kdy analogický pravopis přijala Matice česká, přičemž se posléze roku 1842 stal i oficiálním školským pravopisem. V těchto letech český jazyk prošel dalšími změnami – jednalo se o tzv. reformu skladnou od V. Hanky. Písmeno „í“ se začalo psát namísto původního „j“ (dřívější wjra – reformovaná wíra). Hlávka „j“ nahradila „g/y“ (geden – jeden, heytman – hejtman), čímž uvolněná hlávka „g“ nahradila původní hlávku „ğ“ a sama o sobě tak nabyla dnešní známé formy.⁵⁹

Od poloviny 19. století spisovná čeština opět nabývala na své váze – zavedla se i do zbylých vyšších škol, na kterých se do té doby mluvilo německy a později nabyla i na významu úředním. Došlo k rozvoji české publicistiky a dalších odvětví. Také nastaly poslední závažné

⁵⁷ F. J. Tomsa ve své gramatice navrhnul například psaní <í> namísto dosavadního <j>, psaní <j> namísto <g>, <u> místo <v>, zrušení litery <ł> a psaní <š> místo <ll>. Zdroj: Nový encyklopedický slovník: Obrozenecký pravopis [online]. [cit. 2024-03-07]. Dostupné z: https://www.czechency.org/slovník/OBROZENSKÝ_PRAVOPIS. (Dále jen: NES: Obrozenecký pravopis.)

⁵⁸ M. Komárek, Dějiny českého jazyka.

⁵⁹ NES: Obrozenecký pravopis [cit. 2024-03-07].

pravopisné změny opět z iniciativy V. Hanky s pomocí od F. L. Čelakovského, J. Franty Šumavského a J. E. Purkyně. Došlo ke změně psaní „w“ na „v“ (wůbec – vůbec) a „au“ na „ou“ (maudrost – moudrost). Poslední zachovalá spřežka „ff“ se začala psát jako „š“, neboť došlo k náhradě dřívější fraktury a švabachu antikvou v tištěných textech. Také se upevnilo psaní „u“ namísto „v“ (wyvčiti – vyučiti).

Během 20. století došlo k mnoha změnám, které češtinu postihly – světové války, germanizace a silící dopad cizích jazyků (ruského a anglického jazyka), tyto změny se ale primárně dotkly slovní zásoby, proto je v této práci rozebírat nebude. Více o českém jazyce na konci 19. století a ve 20. století – kdy například vyšla důležitá mluvnice Jana Gebauera (*Mluvnice česká pro školy střední a ústavy učitelské I-II*, 1890), nebo první obecně závazná *Pravidla českého pravopisu* (1902) založená na předchozí zmíněné kodifikaci – se můžeme dočíst v publikaci Miroslava Komárka⁶⁰ nebo Dušana Šlosara.⁶¹ Psaní interpunkce a problematika dělení slov se podrobněji rozebírá na stránce Nového encyklopedického slovníku.⁶²

2.2 Komparace fonologických změn ve vývojových etapách českého jazyka

Fonologický vývoj českého jazyka je úzce spjat s vývojem spisovné češtiny, kterou jsem popisovala v předchozí kapitole. Lze jej rozdělit do několika vývojových etap: pračeština (cca. r. 1000-1150), stará čeština (cca. r. 1150-1500), střední čeština (cca. r. 1500-1775) a nová čeština (cca. r. 1775-dosud). Některé jednotlivé vývojové etapy lze ještě rozdělit do podskupin: stará čeština: raná stará čeština, čeština 14. stol. a čeština doby husitské; střední čeština: humanistická a barokní čeština; nová čeština: obrozenecká čeština, poobrozenecká čeština 19. stol., čeština 1. pol. 20. stol. a čeština 2. pol. 20. stol.⁶³

Během těchto období došlo v českém jazyce k několika fonologickým změnám, které nalezneme v tabulce (Tabulka 5) nacházející se na další straně. Jednotlivé změny však nebudou v této práci blíže rozebírány, neboť nejsou stěžejním tématem této práce – jednotlivé změny pouze zkomparuji do tabulky, která se stane základními vstupními daty pro vlastní vytvořenou online aplikaci „Automatická transkripce“ (<http://penkavova-jcu.wz.cz>). Více o samotných změnách a datech, které jsem vložila do tabulky, pojednávají publikace: *Dějiny českého jazyka* od Miroslava Komárka,⁶⁴ *Dějiny češtiny* od kolektivu autorů,⁶⁵ *Humanistická čeština* od Jaroslava Poráka,⁶⁶ dále okrajově i *Historická mluvnice češtiny I* od Pavla Koska,⁶⁷ dílo Dušana Šlosara – *Spisovný jazyk v dějinách české společnosti*,⁶⁸ podrobně informace nalezneme i ve všech edičních zásadách zmíněných dříve⁶⁹ a přehledně ve skriptech *Základy historické*

⁶⁰ M. Komárek, *Dějiny českého jazyka*.

⁶¹ D. Šlosar, *Spisovný jazyk*.

⁶² NES: Obrozenecký pravopis [cit. 2024-03-07].

⁶³ Nový encyklopedický slovník: Periodizace vývoje češtiny [online]. [cit. 2024-03-14]. Dostupné z: https://www.czechency.org/slovník/PERIODIZACE_VÝVOJE_ČEŠTINY.

⁶⁴ M. Komárek, *Dějiny českého jazyka*.

⁶⁵ FIDLEROVÁ, Alena A., Robert DITTMANN, František MARTÍNEK a Kateřina VOLEKOVÁ. *Dějiny češtiny*. 2013.

⁶⁶ J. Porák, *Humanistická čeština*.

⁶⁷ KOSEK, Pavel. *Historická mluvnice češtiny I*. Brno: Masarykova univerzita, 2014.

⁶⁸ D. Šlosar, *Spisovný jazyk*.

⁶⁹ J. Daňhelka: *Směrnice*.; J. Daňhelka: *Obecné zásady o starém jazyce českém*.; Daňhelka: *Obecné zásady o češtině 15. století*.; J. Vintr: *Zásady*.; I. Šťovíček: *Zásady*.

mluvnice češtiny s texty k rozboru od Hany Marešové.⁷⁰ Tabulka taktéž nepohlíží na změny týkající se nářečí, neboť cílem práce je zjistit, zda lze vytvořit univerzální aplikaci sloužící k přepisu staré češtiny, ve které si každý uživatel bude moci sám zvolit případné nářeční odchylky či jiné změny v sekci s vlastními pravidly. Bližší popis a charakteristika jednotlivých fonologických změn by si zasloužila vlastní publikaci, z toho důvodu odkazuji na dříve zmíněné již vydané práce a nevěnuji se každé změně zvlášť.

Definice zkoumaného tématu z oblasti diachronní češtiny, hlavně tedy posuzované fonologické změny, se mnohdy v publikacích liší. Při komparaci jsem našla několik různých zdrojů uvádějících různé letopočty a míry změn, rozhodla jsem se tedy zvolit kompromis mezi těmito pracemi a letopočty píši od nejstaršího nalezeného až po ten nejmladší. Fonologické změny zaznamenávám takové, které buď celouzemně ovlivnily vývoj českého jazyka, nebo byly jinak významné.

název změny	změna	období
Zánik souhlásek v koncových slabikách	is > ь; us > ь	4.-8. století
Změny z indoevropštiny do praslovanštiny	i > ь; u > ь; ī > i; ū > y; ē > ě (po ž, š, č, j se ě dále změnilo v 'a); s (po hláskách i, u, r, k) > ch; dt/tt > st; ǣ > o; ǣ > e	4.-8. století
Monoftongizace diftongů	ei (eî) > i; oi/ai (oî/aî) > ě; oī/aī > i; au/ou (aũ) > u; eu (eũ) > 'u	5. století
Vznik nosovek	em, en, im, in, ьm, ьn, ēm, ēn > e; am, an, om, on, ām, ān > o	5. století
1. palatalizace	k, g, ch (před i, ь, e, ě, e): k > č', g > ž', ch > š'	kolem r. 400
2. palatalizace	k, g, ch (před ě, i [diftongického původu]): k > c', g > dz', ch > š'	5.-6. století
3. palatalizace	k, g, ch (za i, ь, e, ě, e): k > c', g > dz', ch > š'	není známé
Palatalizace před jotací (formální palatalizace)	kj, gj, chj > č, ž, š; bj, pj, mj, vj > b', p', m', v'; rj, lj, nj > r', l', ň; sj, zj > š, ž; tj > c', dj > dz'; -kti, -gti > cí; -kť, -gť > c'ь; slj > šl'; skj, stj > šč; zgj > ždž; zdj > ždž aj.	7.-8. století
Palatalizační tendence u samohlásek	jo > je; jy > ji; jь > jь	5.-8. století
Metateze likvid (přesmyknutí)	orT-, olT- > raT-, laT-; TorT, TolT > TraT, TlaT; TerT, TelT > TrěT, TlěT (.,T“ zde označuje souhlásku)	8. století
Vznik protetických hlásek	j- (před e, e, ь), v- (před ь, y)	8. století

⁷⁰ MAREŠOVÁ, Hana. *Základy historické mluvnice češtiny: s texty k rozboru*. Olomouc: Univerzita Palackého v Olomouci, 2008.

Kontrakce (stahování)	aja > á; ěje > ie; ěja > 'á, iji > í; oja > á; yji > ý; yjъ > ý; ѣjъ > ý; ѣji > í; ѣjъ > í; ѣje > ie; ѣjo > ō; ѣje > e; eju > ú; ije > ě; oja > á; oje > é; oje > é; oju > ú; ojo > ō; ѣja > 'á; ѣju > 'ú apod.	počátek 10. století
Ztráta jerů	tzv. Havlíkovo pravidlo: počítáno od konce, silné (= sudé) jery se mění v „e“, slabé (= liché) jery zanikají	polovina 10. století
Zánik nosovek (denazalizace)	ę > ä; ě > ā; o > u; ō > ū	konec 10. století
Zjednodušení souhláskové skupiny (denazalizace)	dz' > z'	10.-11. století
Pračeská depalatalizace	t'et > tet; t'ät > tät > tat; t' ät (depalatalizace nastala před d, t, n, l, r, s, z, c, č, š)	konec 10. století až 12. století
Změna palatalizačních odstínů (plná palatalizace)	t', d', n' (před i, ě) > t', d', ñ	10.-14. století
Přehláska nízkých samohlásek (přehláska)	'a > 'ä > ě, 'á > ie; ä > a, 'ā > ie; (ä před "k" zůstalo a přehlasovalo v "ě") - uprostřed slova musí následovat měkká nebo změkčená hláska	12.-13. století
Souhláskové změny	g > γ > h	13. století
Změny slabikotvorných likvid	(ɔ)l, (ɔ)l (dříve zaniklé ztrátou jerů) > lu; čr, žr, ščr > čer, žer, ščer	13. století
Přehláska vysokých a středových samohlásek	'u > 'ü > 'i; 'ú > 'ű > 'í; 'o > 'ö > 'ě; 'ó > 'ő > 'ie	14. století
Asibilace (sykavkový šum)	t's, d'z > c', dz'; r's > ř	13.-14. století
Asibilace	r' > ř	14. století
Hlavní historická depalatalizace	p'e, b'e, m'e, v'e, s'e, z'e, c'e i d'e, t'e, n'e > pe, be, me ve, se, ze, ce; de, te, ne; c'ě > ce, s'ě > se, z'ě > ze, č'ě > če, š'ě > še, ž'ě > že; c'i, s'i, z'i; č'i, š'i, ž'i > cy, sy, zy, čy, šy, žy; b'i, p'i, m'i, v'i > bi, pi, mi, vi	konec 14. století
Vydělení jotace	p'ě, b'ě, m'ě, v'ě > bje (bě), pje (pě), mje (mě), vje (vě); 'ě > ie	konec 14. století
Změna slabikotvorné likvidy -l-	l'ě > le	13. století

Diftongizace a Ztráta jotace	ó > uó; ě > e; ý > ei > ej (hlavně v mluvené formě); ú > au > ou	14.-16. století
Vývoj vokálu -y-	y > i	2. pol. 14. st.-15. století
Disimilace	v > u (před labiálami p, b, m); šč > šť > št; ždž > žď; žč > žť; čč, čšč > čť; čř > tř	14.-15. století
Asimilace artikulační	mj > mň; sč > šč > šť	14.-15. století
Vznik a zánik konsonantů	sr' > stř	14.-15. století
Monoftongizace	ie, ié > í; uo, uó > ů (ú)	15.-16. století
Úžení	é > í	15.-16. století
Téžeslabičná změna diftongu	aj > ej	15.-16. století

Tabulka 5 – Komparace fonologických změn.

3 Automatizace přepisu starých i novodobých textů

3.1 Dostupné transliterační a transkripční aplikace

V dnešní době existují aplikace, které dokáží text ztransliterovat, avšak nikoli primárně ztranskribovat. Vedle aplikace Transkribus, které se budu věnovat v další kapitole, existují aplikace jako je například Projekt Pero,⁷¹ Diptychon,⁷² eScriptorium,⁷³ OCR4all⁷⁴ a Rescribe.xyz.⁷⁵ Každá aplikace má však své výhody a nevýhody. Taktéž využívají různého rozpoznávání textu – buďto v dnešní době velice známého OCR (Optical Character Recognition – optické rozpoznávání znaků), nebo HTR (Handwriting Text Recognition – rozpoznávání ručně psaných textů), o rozdílech mezi nimi píše více až v další části této práce.

Pro svou práci jsem si vybrala aplikaci Transkribus z důvodu existence českého transliteračního modelu, na jehož vzniku jsem se podílela, a taktéž z důvodu, že mi z výše uvedených aplikací připadá uživatelsky nejpřívětivější, přestože v dnešní době dochází k zavedení poplatků, bez kterých uživatel nemůže využívat určité užitečné a zajímavé funkcionality.

Projekt Pero si klade za cíl vytvořit technologii a nástroje, které zlepší a rozšíří přístupnost digitalizovaných historických dokumentů odborné i laické veřejnosti. Taktéž se projekt snaží dosáhnout vytvoření aplikace, která bude schopná automaticky zpracovat obsah i špatně čitelných a ručně psaných dokumentů na úrovni full-textového vyhledávání.⁷⁶

Diptychon je výzkumný projekt financovaný německou výzkumnou nadací DFG (Deutsche Forschungsgemeinschaft) a je společným projektem Univerzity v Brémách a Technické univerzity v Berlíně. Zkoumá metodu, která umožňuje znalcům paleografie přesně extrahovat jednotlivá písmena libovolných rukopisů z obrazů dokumentů. Uživatelé mají možnost stáhnout si soubor s instalačním balíčkem přímo na stránkách aplikace Diptychon.⁷⁷

Aplikace eScriptorium je součástí projektů Scripta, RESILIENCE a Biblissima+. Finanční prostředky získala díky Universitě PSL (Univerzita Paris sciences et lettres) a z programu Evropské unie Horizont 2020 Research and Innovation Program. Jeho cílem je poskytnout vědcům v humanitních vědách integrovaný soubor nástrojů pro přepis, anotaci, překlad a publikování historických dokumentů.⁷⁸

OCR4all jako všechny ostatní výše zmíněné aplikace slouží ke kvalitnímu zpracování náročných rukopisů až po hromadné full-textové rozpoznávání tisků. Jeho výhodou je, na rozdíl od aplikace Transkribus, že je zcela zdarma a webové stránky, na kterých se projekt prezentuje, jsou oproti ostatním třem výše zmíněným velmi přehledné.⁷⁹

⁷¹ Projekt PERO [online]. [cit. 2024-02-19]. Dostupné z: <https://pero.fit.vutbr.cz/>.

⁷² Diptychon [online]. [cit. 2024-02-19]. Dostupné z: <https://diptychon.informatik.uni-bremen.de/html/index.html>.

⁷³ eScriptorium [online]. [cit. 2024-02-19]. Dostupné z: <https://www.sofer.info/>.

⁷⁴ OCR4all [online]. [cit. 2024-02-19]. Dostupné z: <https://www.ocr4all.org/>.

⁷⁵ Rescribe [online]. [cit. 2024-02-19]. Dostupné z: <https://rescribe.xyz/>.

⁷⁶ Projekt PERO: O projektu [online]. [cit. 2024-02-19]. Dostupné z: <https://pero.fit.vutbr.cz/about>.

⁷⁷ Diptychon [online]. [cit. 2024-02-19]. Dostupné z: <https://diptychon.informatik.uni-bremen.de/html/index.html>.

⁷⁸ eScriptorium: Open source code [online]. [cit. 2024-02-19]. Dostupné z: <https://gitlab.com/scripta/escriptorium>.

⁷⁹ Diptychon [online]. [cit. 2024-02-19]. Dostupné z: <https://diptychon.informatik.uni-bremen.de/html/index.html>.

Poslední zmíněnou aplikací sloužící k automatizaci přepisu textů je Rescribe. Aplikaci spravuje výzkumný tým, který se zaměřuje na software umožňující optické rozpoznávání znaků (OCR) a na školení k historickým textům. Opět se jedná o projekt, který nabízí, stejně jako OCR4all a Diptychon, možnost si zcela zdarma stáhnout open source software s nástroji sloužícími k rozpoznávání textu.⁸⁰

Více o jednotlivých aplikacích se můžeme dozvědět přímo ze stránek projektů, popřípadě z literatury, která však prozatím existuje primárně v anglickém jazyce – Projekt Pero,⁸¹ Diptychon,⁸² eScriptorium,⁸³ OCR4all⁸⁴ a Rescribe.xyz.⁸⁵

3.2 Transkribus

Jak jsem již dříve zmínila, nejvíce se budu v této práci věnovat aplikaci Transkribus. Konkrétně popíši její fungování, výhody i nevýhody. Transkribus je platforma fungující na bázi umělé inteligence a slouží k rozpoznávání, přepisu a vyhledávání historických dokumentů – odkudkoli, kdykoli a v jakémkoli jazyce.⁸⁶ Aplikace je součástí projektu READ-COOP SCE,⁸⁷ který byl založen roku 2019. Samotný Transkribus vznikl v rámci projektu Horizon 2020 „READ“ EU v čele s Univerzitou v Innsbrucku. Počátky tohoto projektu se však datují již do roku 2016, kdy proběhla první konference. Součástí tohoto projektu ale není pouze platforma Transkribus, ale i projekt ScanTent⁸⁸ a Smart Search.⁸⁹ Tyto tři hlavní projekty slouží ke komplexnímu řešení digitalizace, automatické transliterace a full-textového vyhledávání v historických textech. První aplikace ScanTent (Obrázek 1) a její mobilní verze DocScan⁹⁰ totiž slouží k digitalizaci dokumentů. Pomocí druhé aplikace – Transkribus – se automatizuje transliterace historických textů – této aplikaci se budu věnovat dopodrobna níže. Poslední zmíněná aplikace, která rozšiřuje samotný Transkribus – Smart Search –, umožňuje fulltextové vyhledávání, což je způsob vyhledávání informací v celých elektronických textech podle

⁸⁰ Rescribe [online]. [cit. 2024-02-19]. Dostupné z: <https://rescribe.xyz/>.

⁸¹ ZAVŘELOVÁ, Alžbeta. Projekt PERO – OCR pro historické texty. Duha: Informace o knihách a knihovnách z Moravy. Brno, 2020, roč. 34, č. 4. ISSN 1804-4255.

⁸² GOTTFRIED, Björn; Wegner, Marius; Worch, Jan-Hendrik; Lawo, Mathias. (2015). Diptychon: A transcription assistant system for the separation of glyphs in medieval manuscript texts. In: K. Herbers und V. Trenkle (Eds.), *Automatische Handschriftenerkennung und historische Dokumentenanalyse*. Göttingen: res doctae. Dokumentenserver der Akademie der Wissenschaften zu Göttingen, 2015.

⁸³ KIESSLING, Benjamin; Tissot, Robin & Stokes, Peter & Stoekl Ben Ezra, Daniel. (2019). eScriptorium: An Open Source Platform for Historical Document Analysis. 19-19. 10.1109/ICDARW.2019.10032.

⁸⁴ REUL, Christian, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner, and Frank Puppe. 2019. "OCR4all—An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings" *Applied Sciences* 9, no. 22: 4853. <https://doi.org/10.3390/app9224853>.

⁸⁵ Rescribe: Download [online]. [cit. 2024-02-23]. Dostupné z: <https://rescribe.xyz/rescribe/>.

⁸⁶ Přeloženo z: Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-19]. Dostupné z: <https://www.transkribus.org/>.

⁸⁷ Český název po rozložení zkratk projektu READ a SCE je „rozpoznávání a obohacování archivních dokumentů“ a „evropská družstevní společnost“.

⁸⁸ Transkribus: ScanTent [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/scantent/>.

⁸⁹ Transkribus: Smart Search [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/smart-search>.

⁹⁰ Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/docscan>.

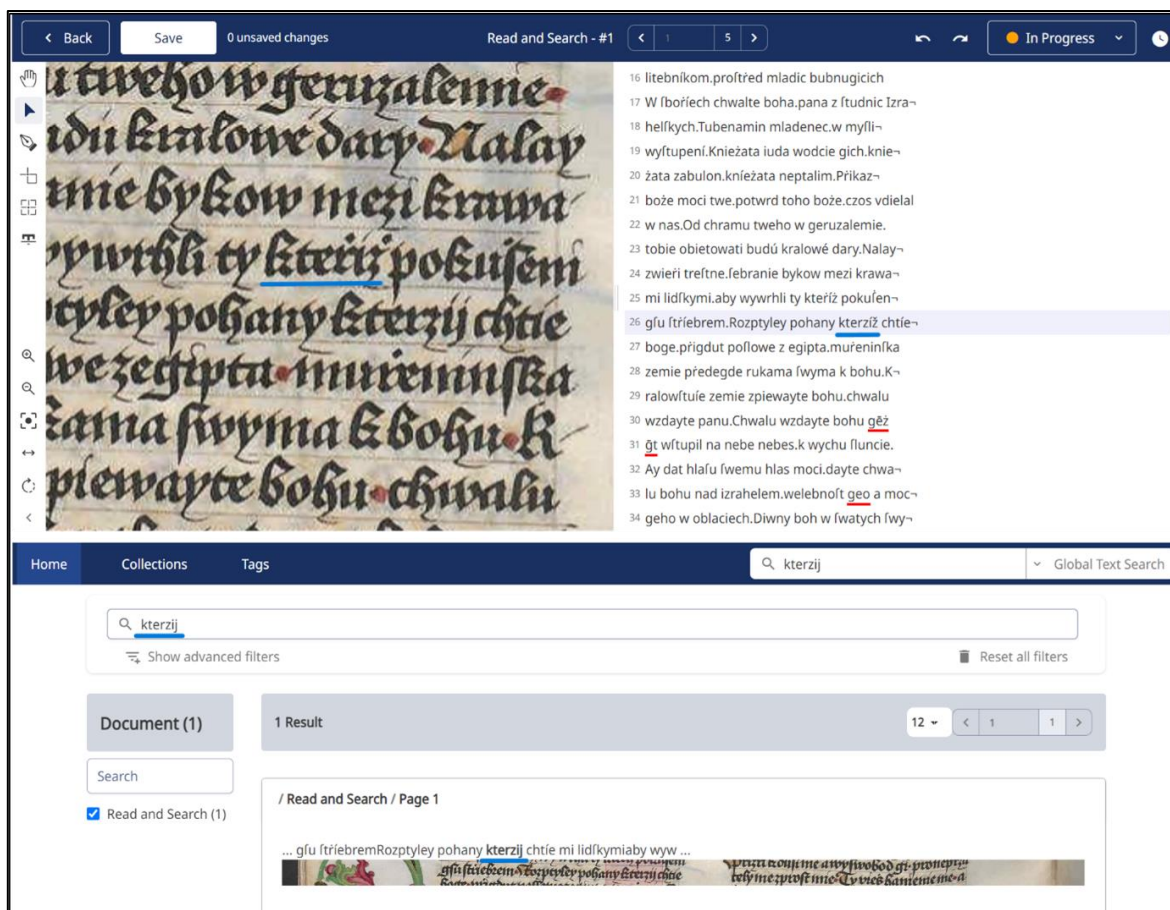
zadaného řetězce znaků (slov, jejich částí apod.)⁹¹ s přihlédnutím k chybovosti. Díky této poslední části projektu se tak může mnohonásobně zrychlit hledání a doplňování slovníků a korpusů pro starší češtinu, protože funkce umožňuje rychlé vyhledávání skrze automaticky transliterovaný text s přihlédnutím i k případným chybám, které by mohly vzniknout při automatickém rozpoznávání. Znamená to tedy, že pokud hledáme v celém automaticky transliterovaném textu slovo „kterzij“ („kteří“), pomocí Smart Search nám aplikace nabídne i podobně vypadající slova, která mohl Transkribus přechíst nesprávně, např. slovo „kterzíž“ (Obrázek 2). Tato funkce se musí zapnout u dokumentu ještě před samotným rozpoznáváním písma pomocí tlačítka „Smart Search“ v sekci „Start Recognition“, kterou vysvětluji níže. V nedávné době Transkribus zavedl placené předplatné plány, ve kterých je zahrnuta i funkce Smart Search, která sama o sobě z důvodu náročného zpracování výpočetních úkolů stojí o 50 % kreditů více, než je výsledný počet kreditů při tvorbě automatického přepisu. Dle obrázků níže je však patrné, že určitý nejspíše zjednodušený model této funkcionality funguje i bez placeného předplatného, neboť účet, na kterém je aplikace představována, nedisponuje placeným obsahem.⁹²



Obrázek 1 – ScanTent nacházející se v Akademické knihovně Jihočeské univerzity.

⁹¹ Akademický slovník cizích slov: A-Ž. Edited by Věra Petráčková – Jiří Kraus. Dotisk 1. vyd. Praha: Academia, 2000.

⁹² Ke dni 1. 5. 2024 existují čtyři plány. Prvním je „Individual“, s ním začíná každý zaregistrovaný uživatel a který obsahuje základní funkce. Dalším plánem je „Scholar“ a „Organisation“ – tyto plány již v sobě mají více funkcionalit. První placený plán vyjde na 14,9 € měsíčně, popřípadě 163,9 € za rok (cena obsahuje i 20% rakouskou daň). Druhý placený plán nemá přesně určenou cenu, jeho výše závisí na velikosti společnosti a odvíjí se od plánovaného projektu. Taktéž zpočátku každý uživatel dostane kredity na rozpoznávání textu zdarma, po jejichž vyčerpání si musí dokoupit další skrze výše zmíněné plány. Všechny uvedené ceny a funkce se vztahují na 100 možností zpracování stran za měsíc. Pokud bude chtít uživatel zpracovat více stránek, cena se mu navýší. Posledním nejnovějším plánem je „Team“. Nabízí podobné funkce jako dva předchozí zmíněné, ale navíc k tomu umožňuje sdílet kredity mezi až 5 uživateli, kteří získají i zmíněné výhody, ve vašem „týmu“. Tento plán uživatele vyjde na 59,9 € za měsíc, nebo 658,91 € za rok.



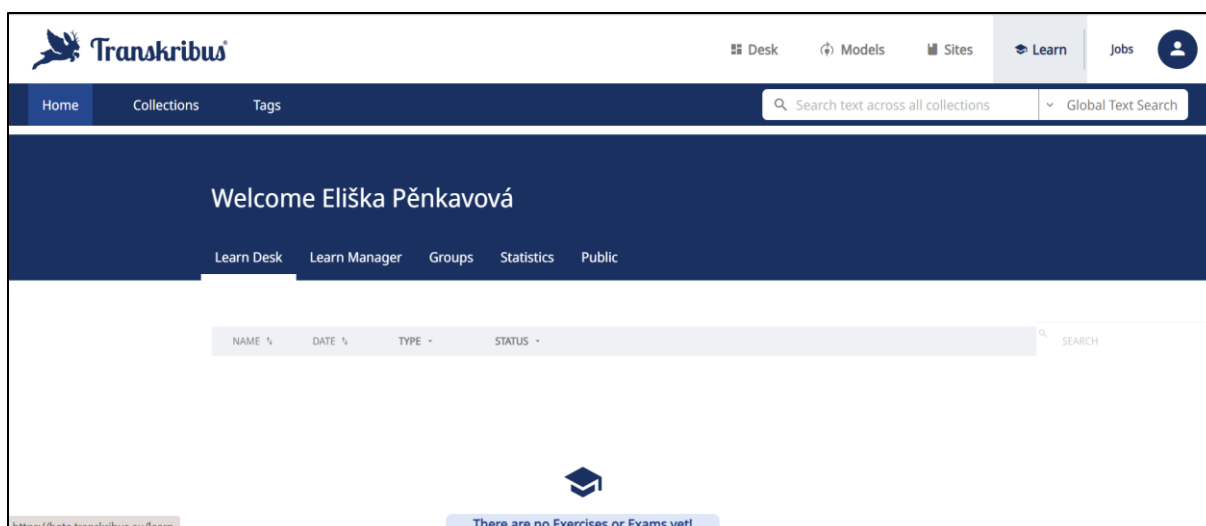
Obrázek 2 – Smart Search.

Na horním obrázku můžeme vidět samotné slovo automaticky přečtené v dokumentu (podtržené modrou linií), následně na obrázku pod ním vidíme funkci Smart Search, která již dané slovo našla pod námi hledaným výrazem ve správné formě (opět podtržené modře).

Zajímavým projektem je taktéž Transkribus LEARN.⁹³ Tento projekt byl vytvořen pro širokou veřejnost, která by se ráda naučila číst historické texty – je to tedy něco jako elektronická paleografická čítanka. Taktéž byl vytvořen i jako nástroj pro znalce, kteří si chtějí čtení procvičit nebo se v něm zdokonalit. Funguje na velmi jednoduché bázi, kdy aplikace označí slovo, které máme přečíst, poté ho napíšeme do pole pod dokumentem a následně Transkribus LEARN řekne, zda jsme slovo přečetli správně, či nikoli. Tuto aplikaci najdeme v online rozhraní aplikace Transkribus v pravém rohu horní lišty pod názvem „Learn“ (Obrázek 3) – prozatím se tato funkce nachází pouze v Transkribus Beta.⁹⁴ Následně si stačí vybrat z publikovaného dokumentu v záložce „Public“, nebo si zvolit vlastní soubor.

⁹³ Transkribus: Beta [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://beta.transkribus.eu/learn/learndesk?type=none&status=none>.

⁹⁴ Týž.



Obrázek 3 – Transkribus Learn.

Nyní již k vysvětlení samotné aplikace Transkribus. Aplikaci můžeme spustit buďto jako desktopovou aplikaci⁹⁵ nazývanou Transkribus eXpert (Obrázek 4), kterou lze stáhnout z internetových stránek projektu,⁹⁶ nebo pomocí druhé možnosti skrze internetový prohlížeč (Obrázek 6). Obě prostředí fungují díky online rozhraní – rozpracovaný dokument tedy uvidíme v obou verzích projektu. Základním vstupem platformy je naskenovaný dokument – může být psaný ručně nebo tištěný – Transkribus podporuje všechny varianty. Samotný sken se následně nahraje do prostředí aplikace Transkribus pomocí tlačítka „Import document(s)...“ (Obrázek 5) v desktopové aplikaci, nebo „Upload files“ v internetovém prostředí (Obrázek 6). Aplikace Transkribus podporuje několik formátů, jak jednotlivé stránky do prostředí nahrát – jedná se o formáty JPEG/JPG (10 MB), PNG (10 MB) a PDF (200 MB), přičemž maximální počet stran je 3 000. Na rozdíl od nyní zmíněných formátů dostupných v internetovém rozhraní desktopová aplikace podporuje jiné a pokročilejší formáty. Nabízí totiž možnost nahrát dokumenty skrze Private FTP,⁹⁷ URL z DFG Viewer METS,⁹⁸ URL z IIIF manifest,⁹⁹ dále extrahovat a nahrát obrázky z PDF formátu, nebo nahrát jednotlivé dokumenty. Následně ve vlastní kolekci, kam jsme dokument nahráli, jednotlivé stránky a dokumenty uvidíme – zatím bez jakéhokoli označení a rozpoznání řádků.¹⁰⁰

⁹⁵ Desktopová aplikace je softwarový program, který je určený k instalaci a použití na stolních nebo přenosných počítačích. Na rozdíl od webových aplikací, které fungují v prohlížeči, desktopové aplikace jsou instalovány přímo na pevný disk počítače a obecně poskytují rychlejší a plynulejší výkon a lepší integraci s operačním systémem a jinými aplikacemi. Definice viz: DAMI: Nás IT slovník [online]. [cit. 2024-02-20]. Dostupné z: <https://www.damidev.com/slovník/desktopova-aplikace>.

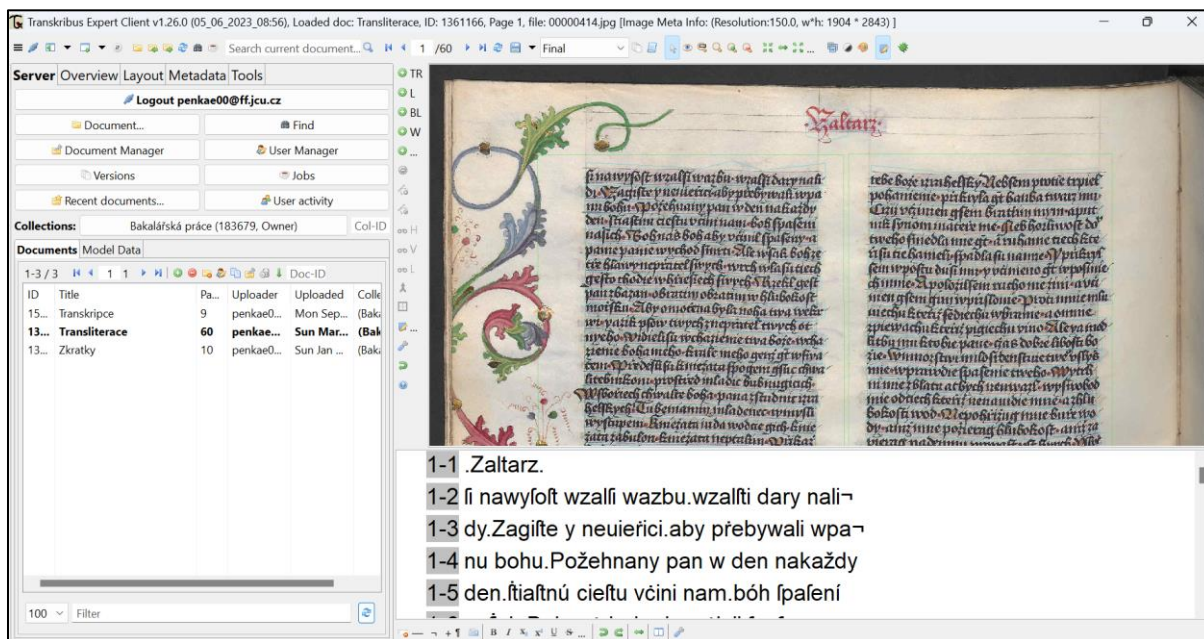
⁹⁶ Vývojáři se však rozhodli postupně od této verze odstoupit, a tak desktopová aplikace nadále nebude dostávat další aktualizace, taktéž mnoho nových funkcionalit je dostupných pouze v internetovém prohlížeči.

⁹⁷ Možnost FTP umožňuje nahrát více než jednu složku/soubor PDF najednou. Pracuje se soubory PDF, JPEG, PNG a TIFF, ale je zapotřebí mít nainstalovaného FTP klienta na počítači.

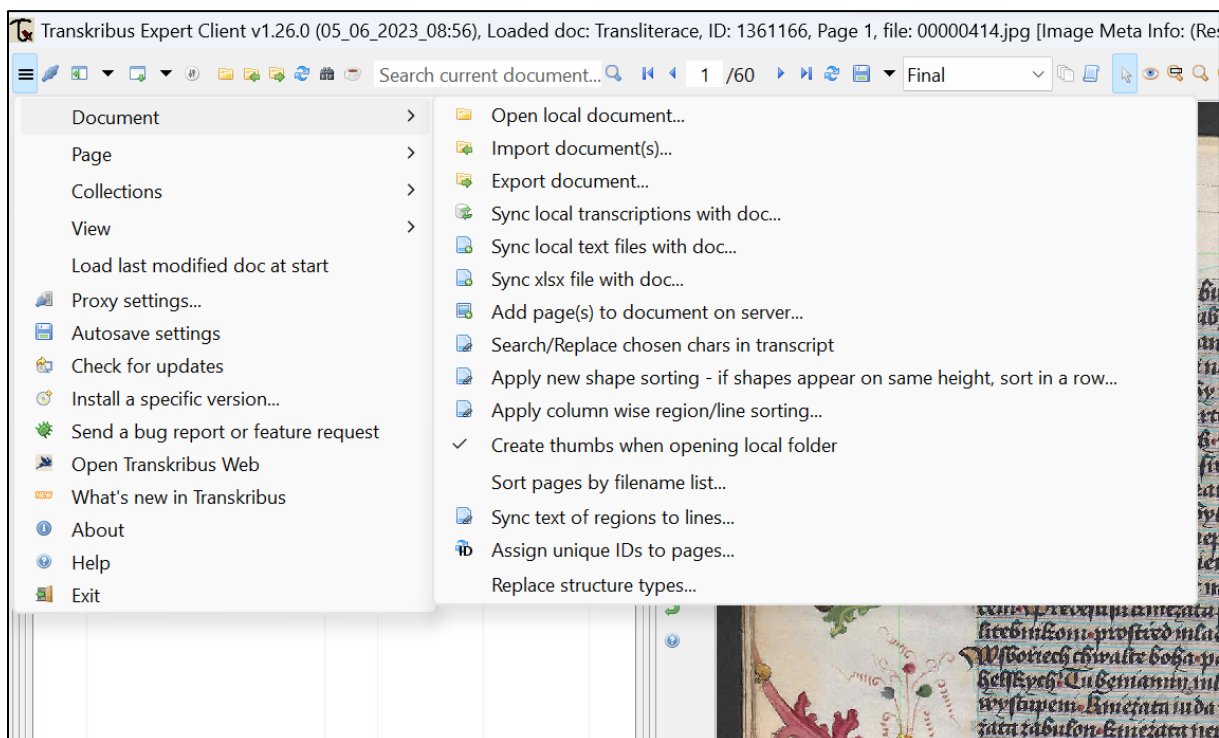
⁹⁸ DFG Viewer METS je rozhraní a formát, který se používá při digitalizaci v některých knihovnách nebo jiných institucích. Pokud tedy daná instituce zdigitalizovala dokument pomocí tohoto rozhraní, stačí pouze tuto adresu zkopírovat a nahrát do aplikace Transkribus.

⁹⁹ IIIF manifest formát je obdobný formátu DFG Viewer METS. Je to jeden z formátů, který mohou instituce využít při digitalizaci dat, takže opět stačí pouze zkopírovat odkaz na digitalizát a vložit jej do aplikace Transkribus.

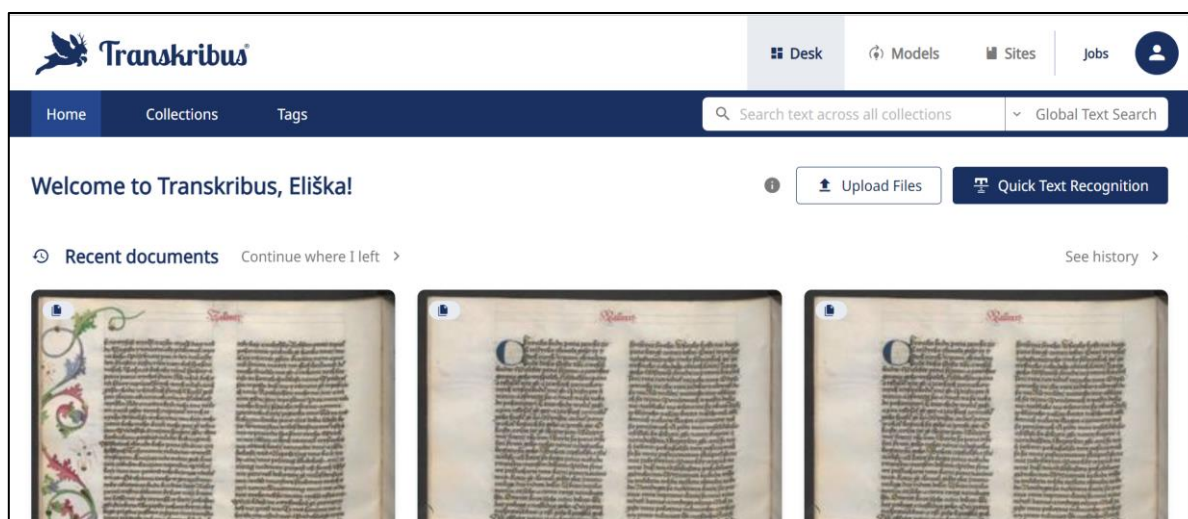
¹⁰⁰ Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/uploading-files-to-transkribus>.



Obrázek 4 – Transkribus eXpert (desktopová aplikace).

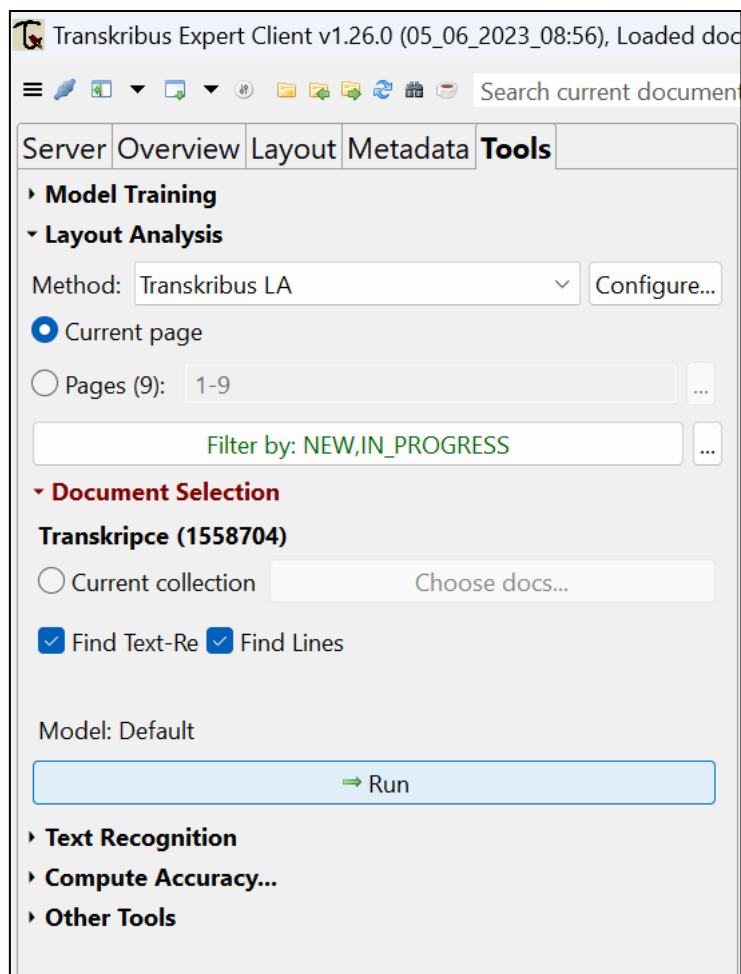


Obrázek 5 – Import document(s) – nahrání souboru.

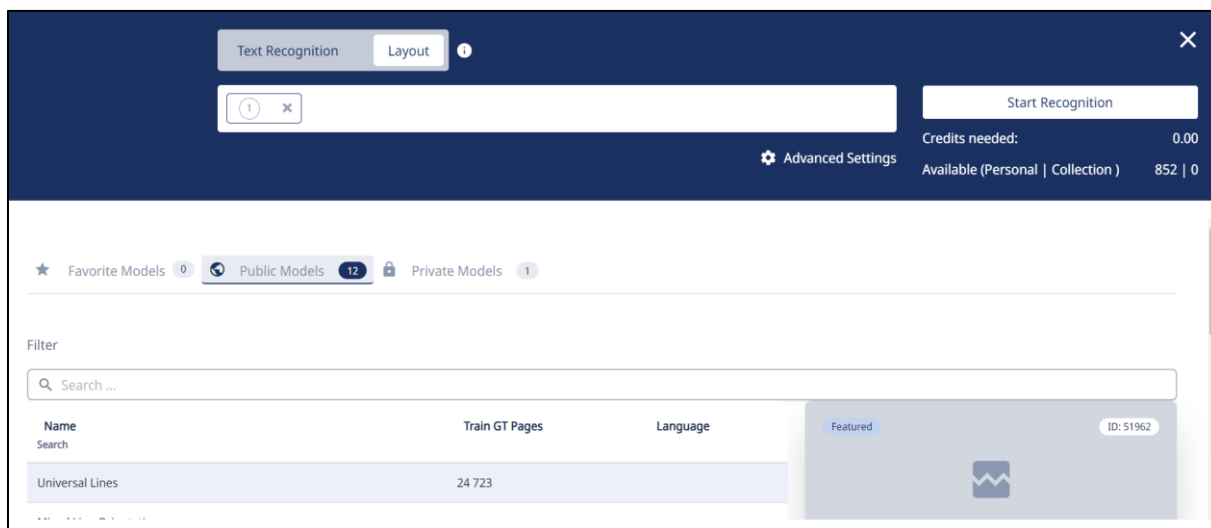


Obrázek 6 – Transkribus internetové prostředí.

Automatické rozpoznání řádků – což jednoduše znamená, že dáme aplikaci za úkol nalézt všechny řádky textu – funguje pomocí sekce „Layout Analysis“ (Obrázek 7). Abychom danou sekci našli, musíme mít otevřenou konkrétní stránku nahraného dokumentu, následně stačí pouze pokračovat přes „Tools“ → „Layout Analysis“ → „Run“. V internetovém prostředí najdeme funkci po kliknutí na tlačítko „Start automatic transcription“ → „Layout“ (Obrázek 8). V sekci si můžeme zvolit počet stran, na kterých se rozpoznávají řádky, stejně tak jako možnost zvolit automatické nalezení textového regionu a řádek. Pokud máme dokument, který je náročnější a složitější na rozložení sloupců a řádek (například urbáře), doporučuji tyto možnosti odznačit a manuálně na stránkách samotné textové regiony a řádky vyznačit pomocí tlačítek „Add a text region“ a „Add lines“ nacházející se v liště po rozkliknutí jednotlivých listů. V internetovém prostředí se navíc vybírají konkrétní modely rozpoznávající řádky a textové regiony, jejich rozdíly a specifika však zůstávají podobné jako v desktopové aplikaci. Po následné volbě možností stačí pouze kliknout na tlačítko „Start Recognition“, přičemž nám automaticky začne běžet proces hledání řádků a regionů textu. Je nutné si uvědomit, že aplikace potřebuje nějaký čas na zpracování jednotlivých úkonů v poměru k tomu, jak je dokument objemný, proto daný výsledek nevidíme hned, ale například až za pár minut. Samotný postup úkonů můžeme sledovat v „Show jobs“, v internetovém prohlížeči pouze „Jobs“ na hlavní straně. Následně, poté co rozpoznávání skončí, se nám po opětovném načtení dokumentu v kolekci zobrazí vygenerované a požadované řádky a textové regiony. Může se stát, že umělá inteligence prozatím nedokáže najít úplně přesně všechny řádky a textové regiony na stránce – v tom případě je stačí pouze ručně přidat pomocí výše zmíněných tlačítek („Add a text region“, „Add lines“), popřípadě manuálně již vytvořené řádky prodloužit potáhnutím konce, sloučit dvě a více řádek a podobné –, to vše nalezneme v levém panelu v desktopové aplikaci, nebo v „Guide“ v internetovém prohlížeči.



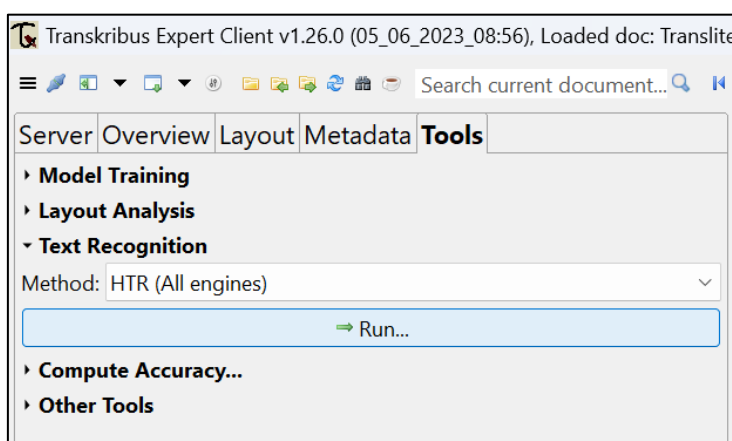
Obrázek 7 – Layout Analysis – Automatické rozpoznání řádků v aplikaci Transkribus eXpert.



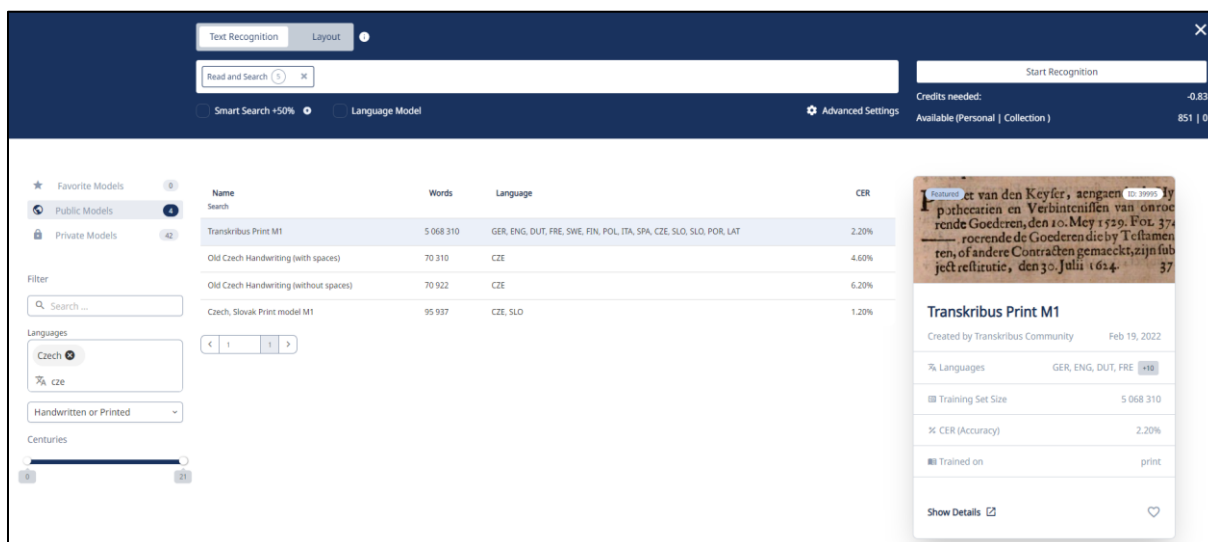
Obrázek 8 – Layout v internetovém rozhraní.

Nyní jsme již úspěšně zvládli najít všechny textové regiony a řádky našeho dokumentu a můžeme přejít k samotnému rozpoznávání písma. Aby se tak stalo, musíme nejprve otevřít daný dokument a následně kliknout na ikonu „Tools“ → „Text recognition“ → „Run“ (Obrázek 9), v internetovém prohlížeči v sekci, která se nám objeví po označení stran v dokumentu, pod tlačítkem „Recognize“ → „Text recognition“ (Obrázek 10). Poté se objeví tabulka

s doplňujícími informacemi (Obrázek 11) – výběr stran dokumentu, modelu a nastavení řádek. Samotný výběr modelu může být zpočátku náročný, protože jich v tuto dobu existuje přibližně kolem 140 a stále jich přibývá. K výběru modelu však velmi dobře poslouží samotné stránky aplikace Transkribus, kde je soupis všech dostupných modelů¹⁰¹ – můžeme si zde zvolit, jaké konkrétní období hledáme, jaký jazyk či styl písma potřebujeme, popřípadě zvolit procentuální hodnotu CER (Character Error Rate), blíže pojem specifikuji níže. Výhodou je, že si můžeme nechat na této stránce jeden list svého dokumentu zcela zdarma automaticky ztransliterovat a zjistit tak, zda je daný model vhodný pro náš dokument. Ten nejhodnější posléze stačí najít v soupisu při vybírání modelu v sekci „Text recognition“. Následně potvrdíme výběr v tabulce a během chvíle proces započne.

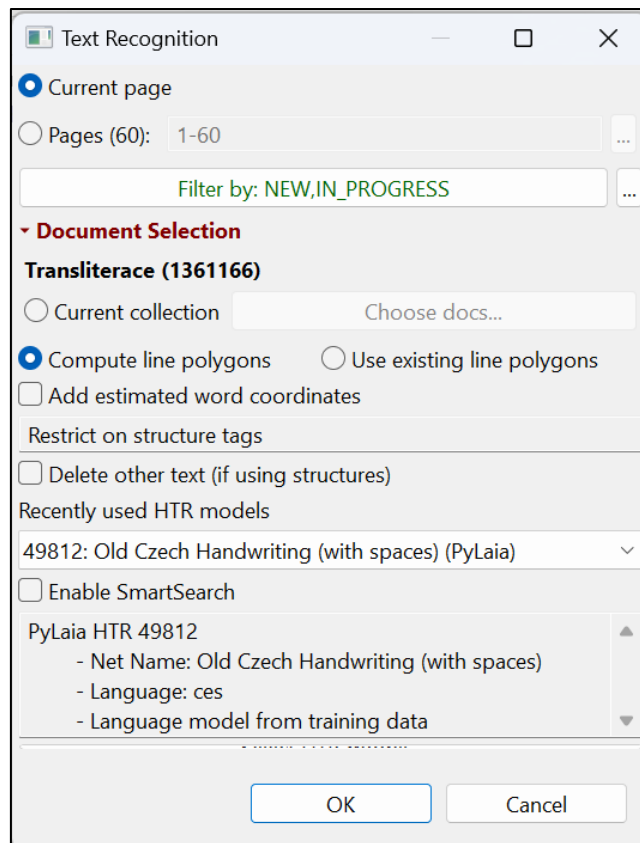


Obrázek 9 – Text Recognition – automatické rozpoznávání textu, úvodní tabulka ještě před samotným výběrem modelu a dalších podrobností.



Obrázek 10 – Text Recognition – internetové rozhraní.

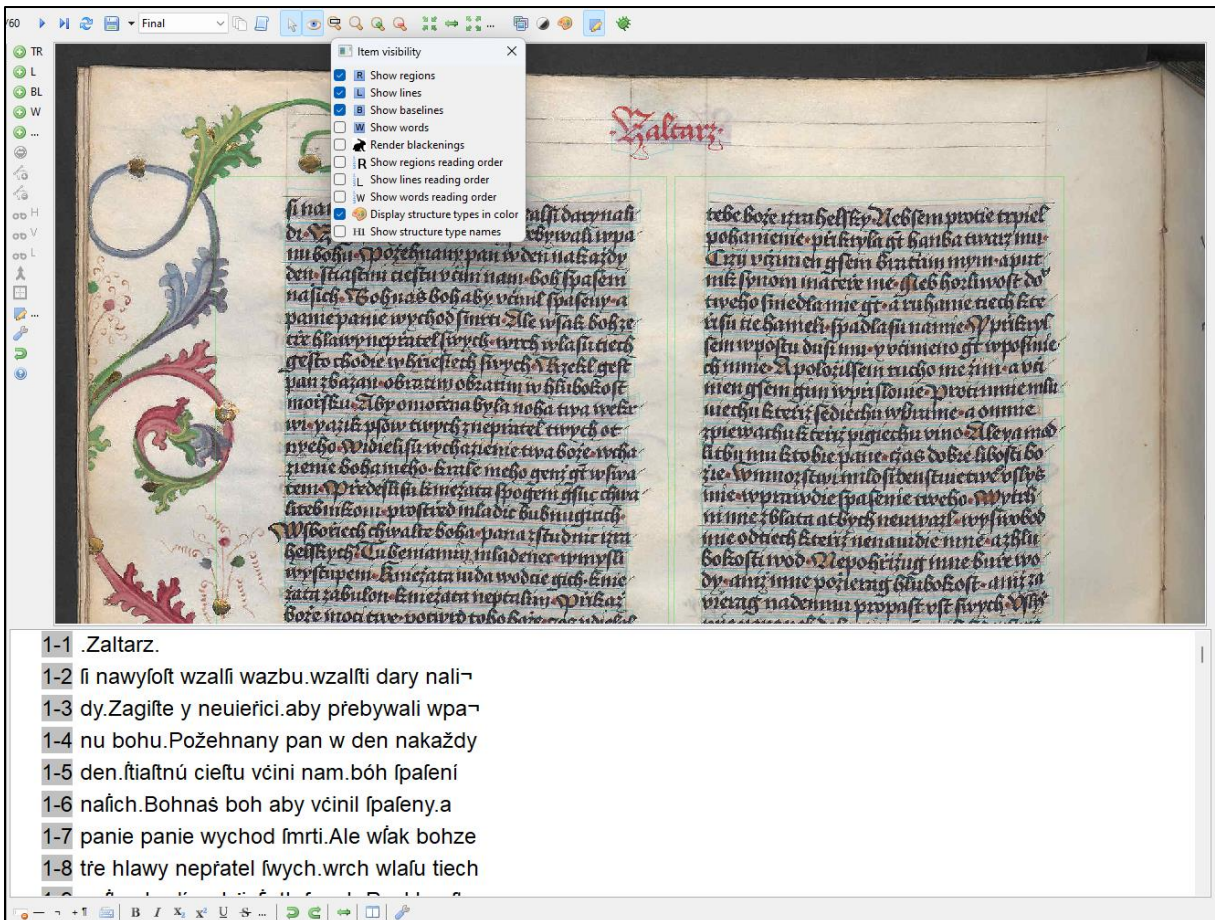
¹⁰¹ Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/transkribus/public-models/>.



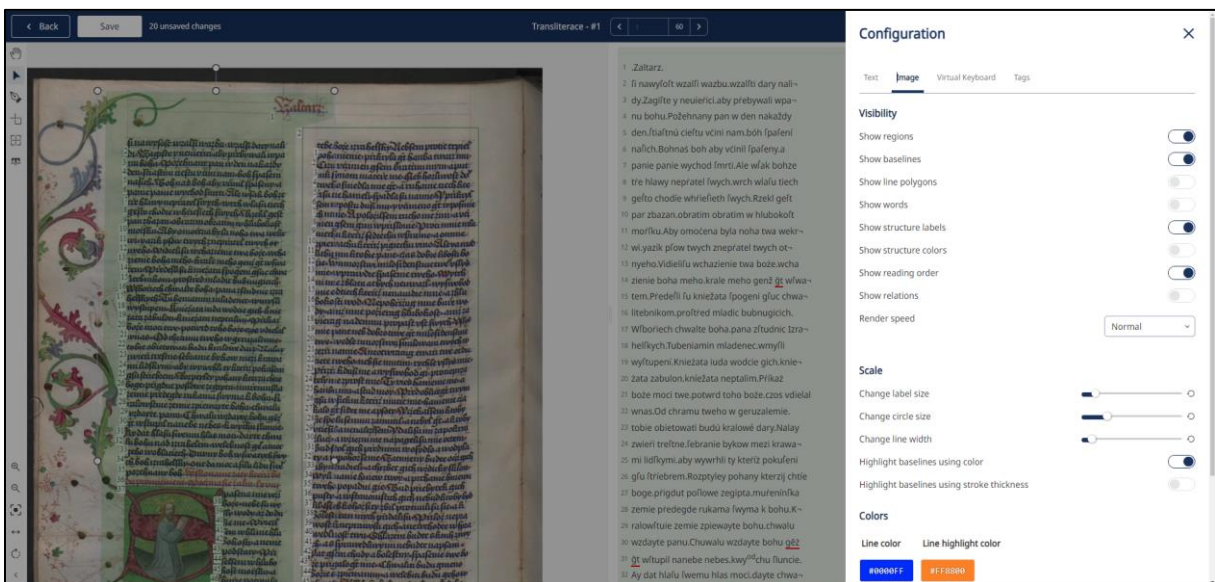
Obrázek 11 – Text Recognition – tabulka s podrobnostmi a nastavením.

Po dokončení procesu automatického čtení se nám v nalezených řádkách pomocí Layout zobrazí text. Tento text je seřazen dle jednotlivých řádků a rozdělen podle textových regionů. V systému si taktéž můžeme nastavit skrytí, nebo zobrazení jednotlivých funkcionalit – zda chceme vidět textové regiony, základní řádky, číslování a podobné –, v desktopové aplikaci se toto nachází pod ikonou oka (Obrázek 12), v internetovém prohlížeči po rozkliknutí nastavení – „Settings“ / „Configuration“ – v pravém dolním rohu (Obrázek 13). Tyto možnosti mohou pomoci k snadnějšímu a přehlednějšímu procházení přepisu. Automaticky přečtený text již můžeme libovolně měnit a označovat různé věci v něm – například zvýrazňovat jména, místa nebo označit zkratky (Abbrevs) (Obrázek 14). Samotné zkratky pak můžeme rozepsat v „expansion“ – v desktopové aplikaci v sekci „Metadata“ → „Textual“ (Obrázek 15), v internetovém prostředí stačí kliknout. Samotné označení se provede pouhým označením slova či sousloví a následným vybráním, v Transkribus eXpert kliknutím pravým tlačítkem (Obrázek 15). Tato označení se nazývají tagy (Tags) a jak napovídá i samotný název sekce v aplikaci Transkribus eXpert, využívají se jako metadata¹⁰² k výstupnímu textu.

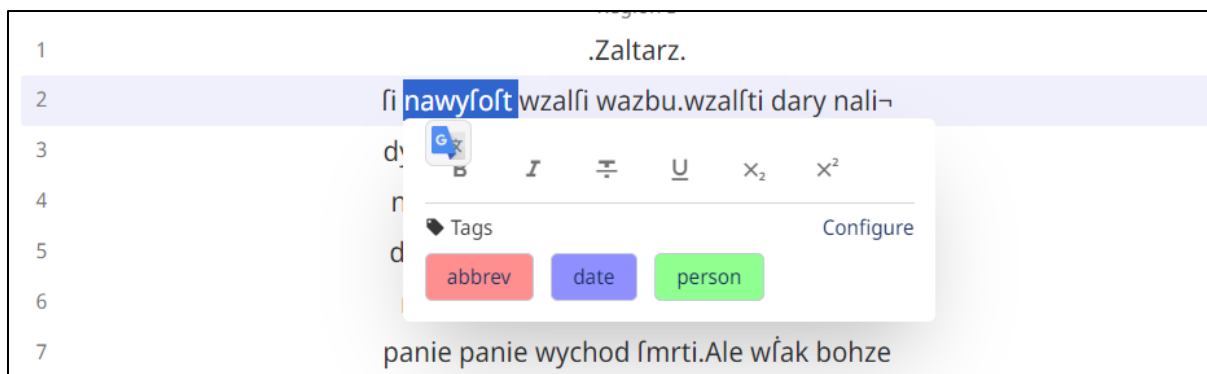
¹⁰² Metadata jsou jednoduše řečeno data o datech – popisují vlastnosti a charakter samotného dokumentu. Uživatelům poskytují kontext, aby se dokázali lépe orientovat v souboru.



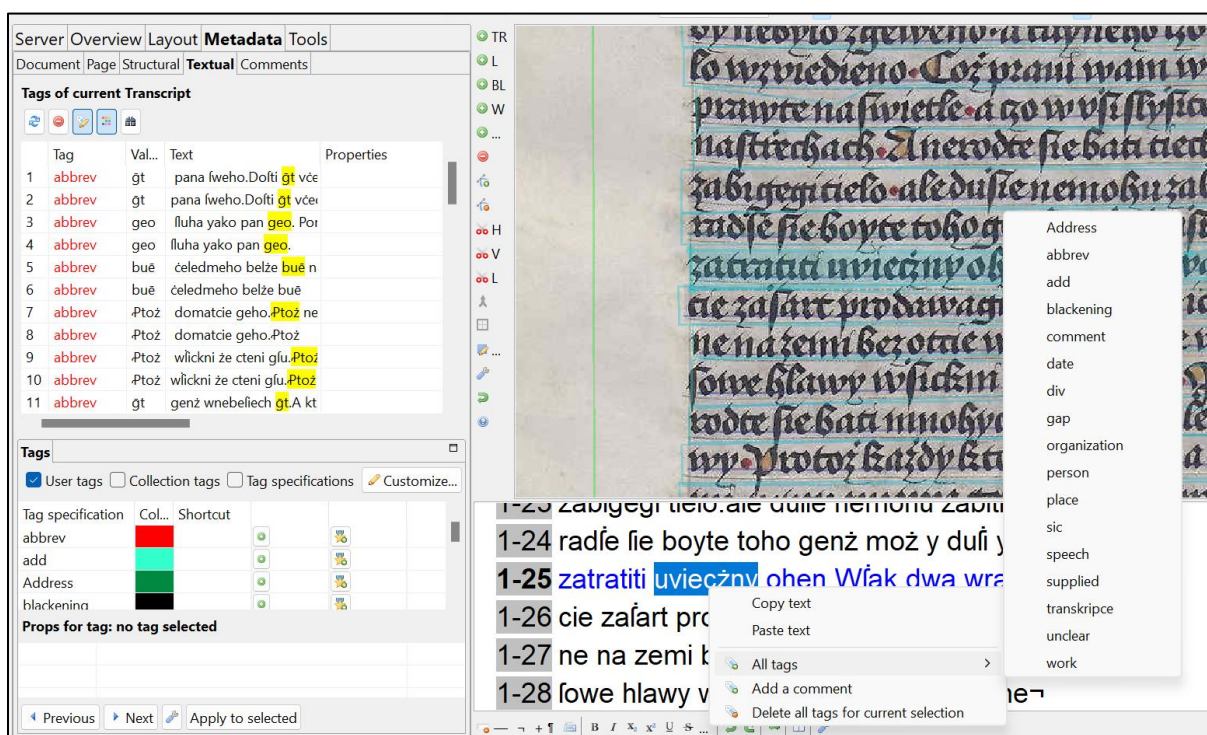
Obrázek 12 – Pokročilé nastavení zobrazování regionů, řádků apod.



Obrázek 13 – Nastavení v internetovém prostředí.



Obrázek 14 – Přidání speciálních označení – Tags – do textu.



Obrázek 15 – Přidání speciálních označení – Tags – do textu v Transkribus eXpert.

Transkribus funguje primárně na bázi HTR (handwriting text recognition¹⁰³), což je druh rozpoznávání textu. Když budeme procházet aplikací hlouběji, zjistíme, že můžeme z primárního HTR přejít na OCR (optical character recognition¹⁰⁴),¹⁰⁵ tuto funkci jsem však našla pouze v desktopové aplikaci v sekci „Text recognition“ při volbě metody. OCR tedy slouží k převodu obrazové předlohy do textu, oproti tomu HTR umožňuje převod i ručně psaných digitalizovaných dokumentů do textu. Důležité si je uvědomit, že pro počítač je strojově psaný text poskytnutý uživatelem v podobě obrázku zcela nečitelný – důvodem je, že

¹⁰³ Rozpoznávání ručně psaného textu.

¹⁰⁴ Optické rozpoznávání znaků.

¹⁰⁵ Technologiím HTR a OCR a problematice automatizace dokumentů se více věnuje diplomová práce Ondřeje Tomišky, který však na tyto technologie pohlíží více z odborného a vědeckého hlediska s důrazem na využití v archivnictví a dalším zpracování. Taktéž práce není určena pro širší veřejnost, která není tolik seznámena s fungováním umělé inteligence, technologií s ní spojené a s anglickým jazykem obecně. Práce: TOMIŠKA, Ondřej. Automatizovaný přepis rukopisných historických dokumentů a jejich využití pomocí moderních IT [online]. Hradec Králové, 2021 [cit. 2024-02-17]. Dostupné z: <https://theses.cz/id/9lbwy8/>. Diplomová práce. Univerzita Hradec Králové, Filozofická fakulta. Vedoucí práce doc. RNDr. Štěpán Hubálovský, Ph.D.

se obrázků neskládá ze znaků, ale z pixelů. Představme si, že bychom počítači poskytli knihu – pro něj by to byla pouhá směs pixelů bez možnosti přečtení. Z toho důvodu také vznikla technologie OCR. Ta dokáže pomocí vložených algoritmů porovnat vnitřní známá data s poskytnutou knihou a následně přečíst, ke kterým znakům se dané písmeno v knize nejvíce podobá. Problém ale nastává při ručně psaných textech, kterých je nespočet různých druhů v závislosti na počtu písařů. Taktéž se kvalita OCR horší při nekvalitním naskenování tisků, šmouhách nebo škrtech – proto také vznikla technologie HTR. Ta se dokáže hluboce učit různým druhům písem a naučit se z vložených dat vytvořit nový originální model.¹⁰⁶ Jednoduše řečeno – OCR pracuje tak, že má uloženou pevnou kolekci vzorů pro jednotlivá písmena a porovnává je s tvary, které nalezne na obrázku, aby určil, co je to za písmeno. Oproti tomu HTR místo toho, aby porovnával tvary s pevnými vzory, jako to dělá OCR, HTR používá pokročilejší výpočty a učení, aby se naučil rozpoznávat různé styly písma s tím, že se jeho šablona písmene postupně modeluje a tvaruje dle vložených dat.

S technologií OCR se denně setkáváme – aplikace, které dokáží přečíst slova a věty v PDF, rozpoznávání podobnosti obrázků, full-textové vyhledání v digitálních knihovnách jako je například digitální knihovna Kramerius a mnoho dalších. Funguje obecná rovnice při rozpoznávání textu pomocí OCR a tou je – čím novější je text, tím úspěšnější bude OCR.¹⁰⁷ Technologie totiž funguje na principu algoritmů, jak již bylo zmíněno výše, které porovnávají známé vzorce – nedokáže se tedy sama učit, pouze přebírá to, co již bylo do algoritmu vloženo. Tato metoda se používá hlavně u tištěných textů, ale například Projekt Pero ji používá i mimo ně. Oproti tomu HTR je technologie, která funguje díky modelu umělé neuronové sítě vytvářející automatický přepis. Tato síť využívá manuálního přepisu o rozsahu minimálně kolem 15 000 až 20 000 slov jako vstupní komoditu.¹⁰⁸ Největší rozdíl mezi těmito technologiemi spočívá v schopnosti učení se, neboť HTR již není programován ručně, ale automaticky se učí z vložených dat. Nejčastěji používaným typem sítě pro potřebu rozpoznávání obrazu je takzvaná konvoluční neuronová síť (convnet) (Obrázek 16).¹⁰⁹ Na obrázku si můžeme všimnout struktury konvoluční sítě. Úplně vlevo se nachází náš vstup (například písmeno „a“), který je do sítě vložen, přesněji se jedná o pixely obrázku. Následně dochází k tzv. konvoluci (convolution), což je proces, kdy filtry procházejí obrázkem a zdůrazňují určité charakteristické rysy (u písmene „a“ by tomu bylo například břicho, pata písmene a podobné), ze kterých vytvoří Feature Maps (prvkové mapy / mapy znaků / mapa rysů). Další fází je Max-pooling (maximální slučování a redukce), která zmenší rozměry, ale zachová důležité informace. Poté se tyto poslední dvě fáze opakují – dochází tedy k extrakci a zjemnění rysů. Dále přichází fáze klasifikace (classification), kdy dojde k porovnání Feature maps s databází, přičemž se k nám dostane samotný výsledek – poznání, či nerozpoznání písmene „a“ v závislosti na vložených trénovacích datech. Výstupem, tedy tím, co „vidí“

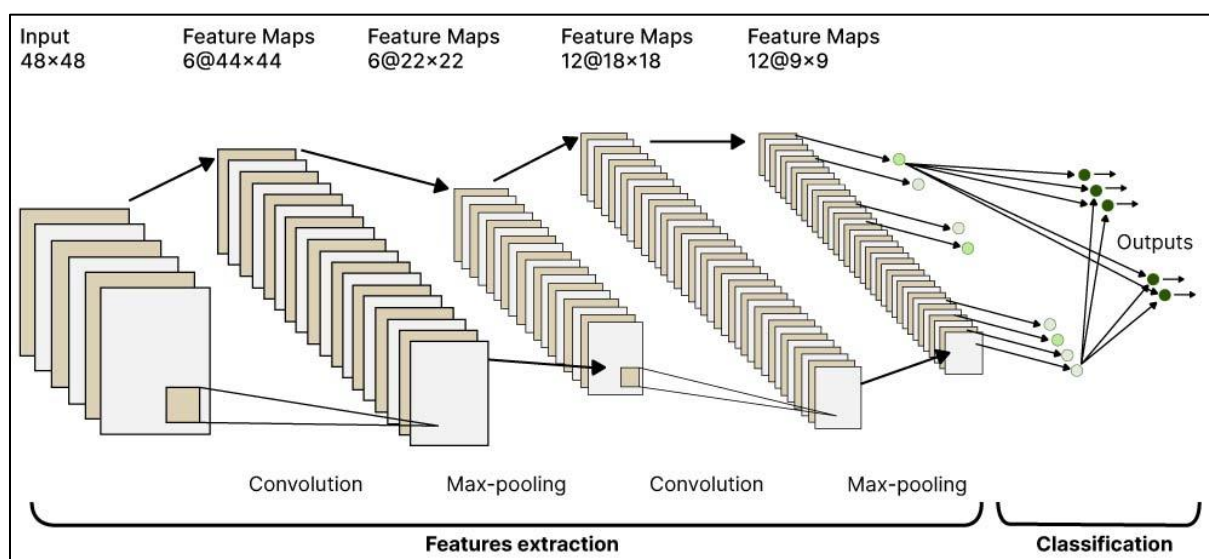
¹⁰⁶ Transkribus: OCR vs HTR [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/insights/ocr-vs-htr/>. (Dále jen: Transkribus: OCR vs HTR)

¹⁰⁷ KBLAB [online], 2014 [cit. 2024-02-20]. Dostupné z: <https://lab.kb.nl/about-us/blog/entangled-histories-ocr-htr-atr-automatic-text-recognition>.

¹⁰⁸ Jak si však ukážeme později, tento počet slouží pouze doporučení, neboť model může fungovat dobře na rukopisu od stejného písaře i z menšího počtu vložených dat.

¹⁰⁹ Transkribus: OCR vs HTR [cit. 2024-02-20].

konvoluční neuronová síť, je vektor čísel, který se používá ke klasifikaci znaku.¹¹⁰ Tento proces se také nazývá trénování. Model se učí na sadě vzorových obrázků se známým obsahem, například písmeno „e“ a „o“. Výstup modelu se poté porovnává se správnými výsledky a na základě nich se model upravuje – pokud tedy například model přečte písmeno „e“ jako „o“, v dalším trénovacím kole se zdokonalí. Trénink modelu pokračuje, dokud se nestane sám pro sebe dokonalým pomocí dosažení požadované úrovně, tedy dokud nedokáže s jistotou poznat rozdíl mezi písmeny „e“ a „o“ z vložených dat. Na základě rozdílu mezi předpovědí modelu a základní pravdou (Ground Truth)¹¹¹ se aktualizují parametry v samotné síti – neuronová síť se učí stejně jako dítě, postupným opakováním a následným opravováním chyb – pomocí námi vložených správných a bezchybných dat se zlepšuje a zdokonaluje.



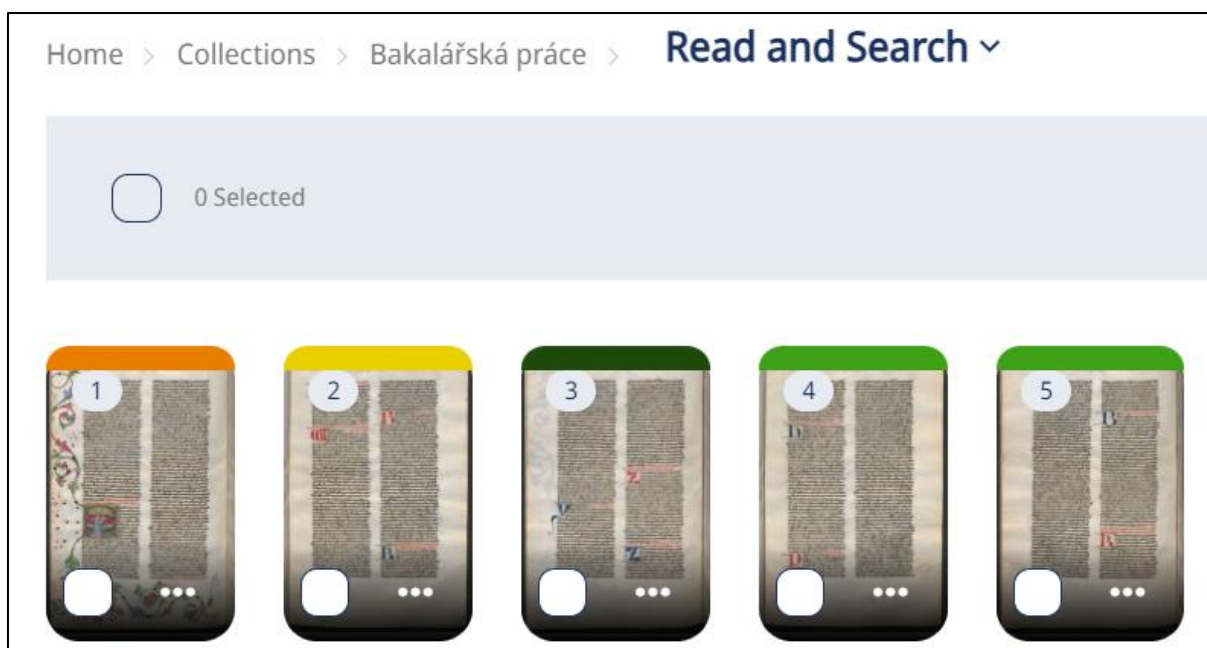
Obrázek 16 – Konvoluční neuronová síť.

Výše jsme si zjednodušeně shrnuli, na jaké bázi funguje tvoření vlastních modelů, nyní však vysvětlím, jak konkrétně si je na platformě Transkribus můžeme vytvořit. Potřebným materiálem pro tvorbu modelu je Ground Truth.

Transkribus umožňuje jednotlivé stránky označit různými značkami (dle toho, jak je dokument rozpracován). Jedná se o značky: „In progress“, „Done“, „Ground Truth“ a „Final“. Pomáhá to při lepší orientaci v jednotlivých dokumentech, neboť můžeme hned ještě před otevřením listu vidět barevně označený horní proužek dokumentu signalizující míru dokončenosti dokumentu (Obrázek 17). Barevné označení se taktéž může využívat při sdílených projektech, protože samotné označení ostatním uživatelům signalizuje, jaké strany jsou již bezchybné, či na jakých se ještě pracuje.

¹¹⁰ Microsoft: Studentské Trenérské Centrum [online]. [cit. 2024-02-20]. Dostupné z: <https://studuj.digital/2020/12/09/umela-inteligence-6-dil/>.

¹¹¹ Ground Truth je strojopisná kopie dokumentu, která musí být 100 % správná. Využívá se totiž k učení modelů, přičemž slouží jako kontrolní síla – při chybně poznaném znaku při trénování modelu z ní neuronová síť čerpá a učí se z ní jako ze vzoru, proto i nese název „základní pravda“. Převzato ze stránky: Transkribus: Ground Truth [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/glossary/ground-truth/>.



Obrázek 17 – Barevné označení jednotlivých stran dokumentu.

První list z dokumentu označuje „In progress“, druhý „Done“, třetí „Ground Truth“ a čtvrtý a pátý „Final“.

Nyní jsme si již vysvětlili, jak nahrát, upravit a automaticky přizpůsobit zdigitalizovaný dokument v aplikaci Transkribus tak, aby byl připravený na prvotní manuální přepis. Ten je totiž důležitým krokem k tomu, abychom mohli dále tvořit vlastní modely na míru. Samozřejmě můžeme k manuálnímu přepisu využít již existující model, avšak musíme vzít v úvahu existenci a kompatibilitnost těchto již vytvořených modelů k našemu dokumentu. V dnešní době totiž existují pouze tři modely podporující češtinu – dva čistě české vytvořené při Winter School 2022 ve Vídni, které jsem se účastnila, a jeden univerzální Print Multi-Language. Během následujících měsíců by měl být taktéž zveřejněn další český model trénovaný opět na Winter School, která proběhla na přelomu let 2023-2024 rovněž ve Vídni a které jsem se také zúčastnila. Z důvodu prozatím malého počtu českých modelů a rozmanitosti druhu písem si nejspíše budeme muset u česky psaných dokumentů vytvořit model vlastní – tedy z manuálního přepisu několika stránek. V případě jiných jazyků již existuje mnoho modelů a je pravděpodobné, že nebude potřeba vytvořit vlastní model. Pokud bude námi nově vytvořený model větších rozměrů, můžeme jej také nechat zveřejnit i pro ostatní badatele pomocí domluvy se samotnými vývojáři aplikace Transkribus, čímž pomůžeme i ostatním uživatelům využívat odvedené práce. Model je totiž po vytvoření vždy soukromý a vidíme ho pouze my, může jej vidět maximálně uživatel, kterého jsme přidali do stejné kolekce.

Nyní jsme se dostali do fáze, kdy si chceme jako uživatel vytvořit vlastní model, neboť neexistuje žádný vhodný k našemu dokumentu. Tuto funkci vyvoláme tlačítkem, které nalezneme v sekci „Tools“ → „Train a new model...“ v Transkribus eXpert (Obrázek 18). V internetovém prohlížeči to nalezneme v oddělení „Models“ → „Train new model“ – v tomto prostředí ještě dostaneme možnost výběru mezi trénováním „Text Recognition Model“ a „Baselines Model“ (Obrázek 19). Druhá varianta („Baselines Model“) slouží primárně pro dokumenty se složitým rozvržením textových regionů jako jsou pohlednice, rejstříky nebo dokumenty s poznámkami. Model by se tedy trénoval konkrétně podle našich zadaných

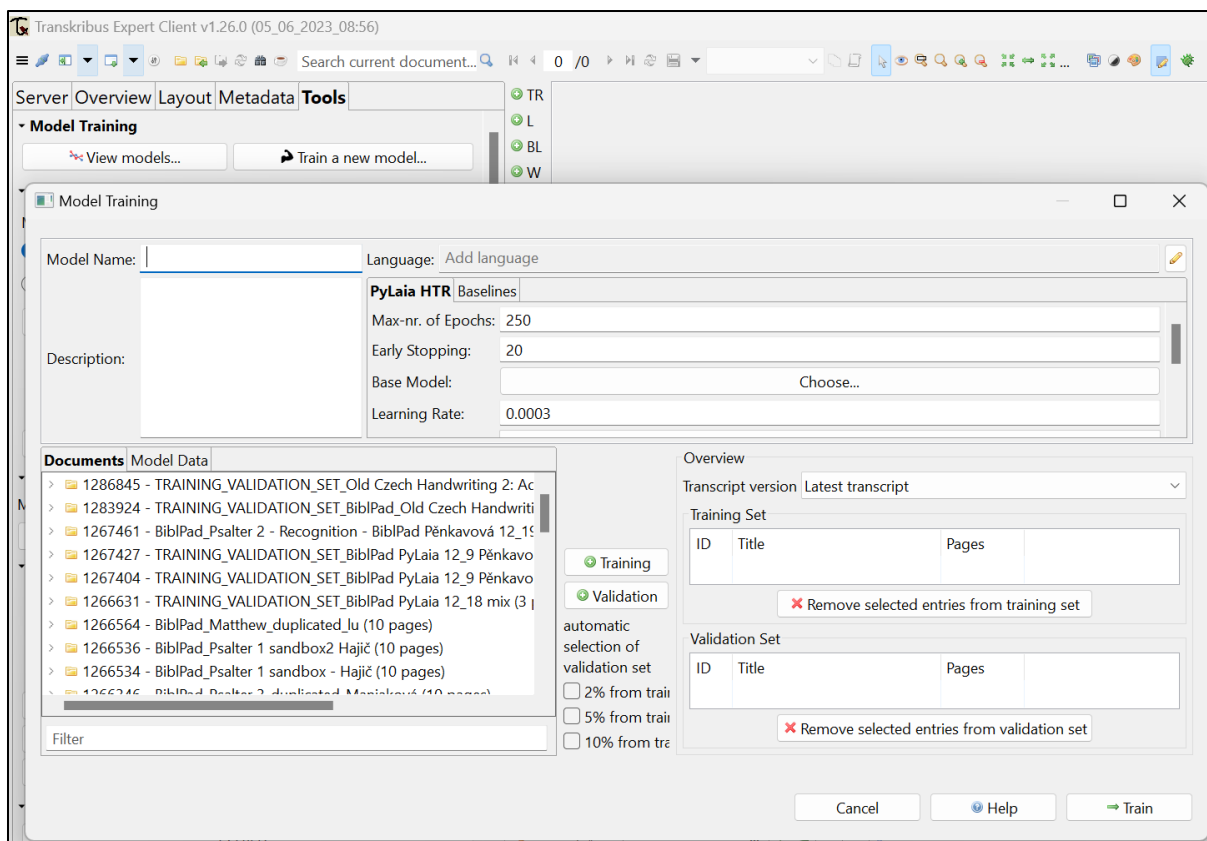
základních linií (baselines). Rozdíl mezi touto variantou a „Text Region Model“ je hlavně v tom, že model „Baselines“ automaticky nerozpoznává textové oblasti v dokumentu, ale pouze řádky (čáry). Výhoda je v tom, že je vytvořen přímo na konkrétní dokument, takže by se neměly vyskytnout problémy s chybným označením textových regionů. Transkribus v tomto případě doporučuje trénovat na 50 stranách našeho bezchybného textu (Ground Truth).¹¹²

Nyní se však vrátím zpět k modelu trénování „Text Region Model“. V obou rozhraních po nás aplikace bude chtít nejdříve „Training data“ – to je ta výše zmíněná Ground Truth, tedy námi ručně bezchybně přepsané listy dokumentu – z ní se bude učit model tvořit tvary jednotlivých znaků. Následně si aplikace vyžádá „Validation Data“ – to by mělo být přibližně 10 % z námi vložené Ground Truth – tento datový set slouží k validaci (kontrola, ověření) dat, které jsme my napsali, přičemž je bude porovnávat s modelem, který postupně umělá inteligence vylepšuje ze zmiňovaných „Training data“. Znamená to tedy, že některé námi napsané listy musíme vložit do sekce „Validation data“. Uvědomme si, že data obou sad nemůžou být stejná, nedávalo by to smysl. „Validation data“ slouží jako ověřovací sada, kterou po dokončení tvorby modelu uvidíme v naší kolekci již automaticky přečtenou modelem. Dále musíme u modelu nastavit jeho název, popis, jazyk, kterým je dokument napsán, popřípadě můžeme nahrát obrázek, který se bude zobrazovat při opětovném načtení modelu z kolekce modelů. Taktéž si můžeme zvolit, zda chceme model trénovat na datech již jiného existujícího modelu – ikona „Base Model“ – to může být výhodné v momentě, kdy postupně rozšiřujeme svůj menší model do většího a obecnějšího modelu například na základě jednoho druhu písma.

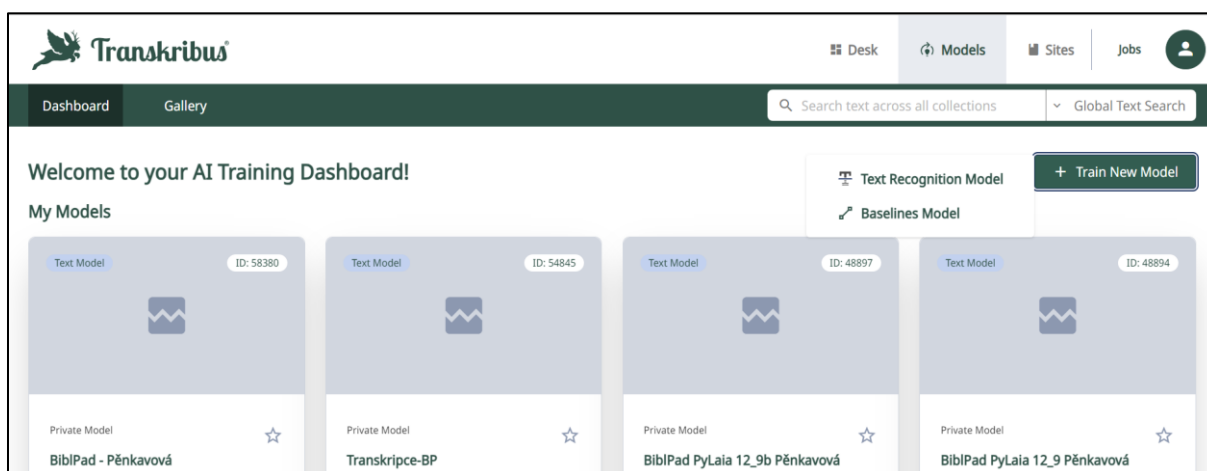
Dále zde můžeme nastavit „Max-nr. of Epochs“ / „Training Cycles“, což je počet opakování, kolikrát síť projde tréninkovými daty. Čím větší je číslo u tohoto nastavení, tím bude modelu delší doba trvat proces učení. Také je důležité zmínit, že se proces učení může sám ukončit dříve, pokud už nebude docházet k jeho zlepšení. Tuto funkci dřívějšího ukončení můžeme nastavit v sekci „Early Stopping“. Ta signalizuje minimální počet cyklů, který musí síť vykonat před dřívějším zastavením – je to tedy povinný počet opakování, který musí model provést.¹¹³ Poté ještě můžeme zvolit, zda chceme naučit model rozepisovat zkratky – „Train Abbrevs with expansion“ –, což je velmi nápomocné při exportu hotového dokumentu. Při učení také máme na výběr zvolit si, jak bude umělá inteligence pracovat s tagy – to je ale již pokročilejší funkce, která připomíná princip metadat, jejichž funkci jsme si vysvětlovali výše. Následně už stačí pouze pokračovat a kliknout na tlačítko „Start“. Postup učení modelu uvidíme opět pod ikonou „Jobs“.

¹¹² Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/baselines-models>.

¹¹³ Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/model-setup-and-training>.



Obrázek 18 – Vytváření vlastního modelu v Transkribus eXpert.



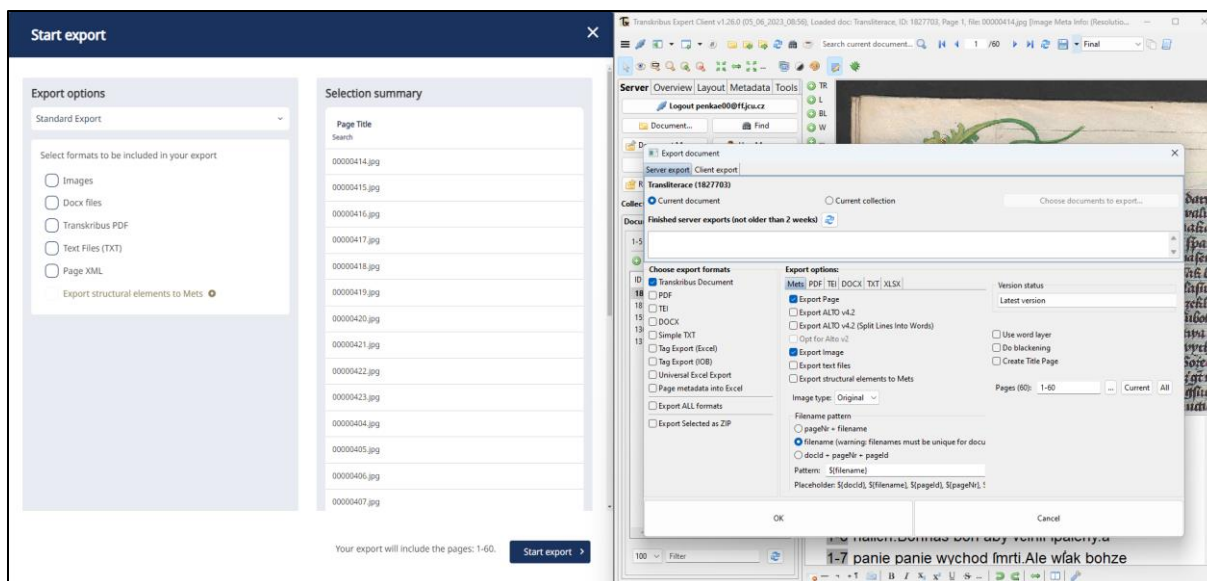
Obrázek 19 – Zobrazení možnosti výběru při tvoření nového modelu v internetovém prostředí.

Již výše jsem zmínila termín Character Error Rate (CER), podle kterého se můžeme orientovat ve zveřejněných modelech a který slouží k porovnání výše chybovosti v modelech. Tato míra porovnává Ground Truth, kterou jsme vložili při tvorbě modelu, s automaticky přečtenými „Validation data“. Čím je hodnota CER nižší, tím je rozpoznávání přesnější – výsledkem této hodnoty jsou procenta. Výpočet funguje tak, že se celkový počet znaků včetně mezer (n) vydělí součtem tří hodnot – počtem nově vložených znaků (i), počtem nahrazených znaků (s) a počtem odstraněných znaků (d) v porovnání s Ground Truth –, to vše se vynásobí 100, abychom získali výsledek v procentech. Vzorek vypadá takto: $CER = [(i + s + d) / n] * 100$. Nejdříve je tedy spočítán celkový počet chyb, které se při kontrole vyskytly, poté jsou chyby vyděleny celkovým počtem znaků Ground Truth a nakonec se celé toto číslo převede na

procenta. Pokud by tedy počítač našel v automaticky přečteném textu 8 znaků napsaných chybně a 2 musel odstranit při celkovém počtu 100 znaků v textu, vzorec by vypadal následovně: $CER = [(0 + 8 + 2) / 100] * 100 = 10 \%$. Znamená to tedy, že v dokumentu o 100 znacích model chybně přečetl 10 znaků, v dokumentu o 1 000 znacích by při 10 % chybovosti přečetl 100 znaků chybně a podobně. Chybně se počítá i to, pokud při automatickém čtení model neoznačí čárku nad písmenem, pokud nerozpozná majuskuli od minuskuli nebo například přečte písmeno „u“ namísto „v“.

Nyní jsme si již vysvětlili téměř celý postup průchodu zdigitalizovaného dokumentu skrze aplikaci Transkribus a poslední, co nám chybí, je export – stáhnutí hotového a automaticky ztransliterovaného dokumentu zpět do počítače. Tento proces se nachází pod tlačítkem „Export document“, který najdeme buďto v Transkribus eXpert vpravo nahoře v sekci „Document“, nebo v internetovém prohlížeči stačí pouze označit listy, které chceme stáhnout a kliknout na tlačítko „Export“ skrývající se za třemi tečkami v liště vedle tlačítka „Train model“. Následně se objeví tabulka, kterou můžeme vidět níže na obrázku – vlevo je internetové prostředí, vpravo desktopová aplikace (Obrázek 20). Je mnoho možností (různých formátů), jak daný dokument stáhnout – PDF, TEI, DOCX, TXT, Images a XML. Zajímavým výstupem je například formát PDF, neboť aplikace vytvoří obrazový PDF soubor s naskenovanými stranami našeho dokumentu obsahující v sobě další vrstvu, kterou můžeme zaktivovat a samotný strojově rozpoznávaný text se propíše na digitalizát. Důležité je také zmínit, že při přepisování či upravování souboru je nutné používat speciální znak sloužící k rozdělování slov na konci řádku a tím je „-“, neboť je jediný, který výstupní formáty chápou jako rozdělovač slov, čímž formát dokáže v následném exportovaném dokumentu napsat daná slova dohromady bez spojníku uprostřed slova.¹¹⁴ Taktéž si můžeme nastavit, zda chceme mít ve výstupu označené tagy, popřípadě rozepsané zkratky, které jsme v souboru označili, či nikoli. Poté, co potvrdíme výběr exportu, nám zanedlouho do e-mailu, přes který jsme přihlášení v aplikaci Transkribus, přijde zpráva s možností stáhnout námi zvolené formáty. Výstupní formát DOCX (Obrázek 21) je vhodný ke kombinaci s využitím aplikace Automatická transkripce, kterou jsem pro tuto práci vytvořila a o které budu více psát v dalších kapitolách.

¹¹⁴ Příkladem je spojení slova „w v-ftech“. Kdybychom jej do aplikace Transkribus napsali ve tvaru „w v-ftech“, v exportu bychom dostali stejné hodnoty, ale při první variantě se speciálním znakem se nám do výstupu přepíše dané spojení jako „w vftch“ (v ústech).



Obrázek 20 – Export – stáhnutí dokumentu.

.Zaltarz. fi nawyfoft wzalfi wazbu.wzalfti dary nalidy.Zagifte y neuierici.abym přebywali wpanu bohu.Požehnany pan w den nakaždy den.řtiaftnu ciefu wćini nam.bóh fpaení nařich.Bohnaš boh aby wćinil fpaeny.a panie panie wychod fmrti.Ale włak bohze tre hlavy nepratel fwych.wrch wřafu tiech geřto chodie whrieřieřth fwych.Rzekł geft par zbazan.obratim obratim w hlubokoft mořřku.Aby omoćena byla noha twa wekřwi.yazik přow twych znepratel twych otnyeho.Vidielifu wchazienie twa bože.wcha zienie boha meho.krale meho genž ģt wfwatem.Předeřli fu knieřzata fpogeni ģfuc chwalitebnikom.prořřred mladice bubnugicich. Wřbořiech chwalte boha.pana zřtdudnic lzrahelfkych.Tubeniamin mladeneć.wmyřli wyřftupeni.Knieřzata iuda wodcie ģich.knieřzata zabulon.knieřzata neptalim.Přikřař bože moci twe.potwřd toho bože.czos wdialat wnas.Od chramu tweho w ģeruzalemie. tobie obietowati budř kralowě dary.Nalay zwieri

Obrázek 21 – Zobrazení výstupních dat ve formátu DOCX.

Jedním z posledních rozšíření platformy Transkribus je funkce „Transkribus Sides“. Jedná se o online databázi, která umožňuje všem uživatelům přístup k námi zveřejněným dokumentům skrze stránku „Sides“. Rozšíření je k nalezení pod tlačítkem „Sides“ nacházející se na domovské uživatelské straně v online rozhraní Transkribus. Zde stačí pouze vytvořit stránky pomocí tlačítka „Create my first site“, následně vyplnit potřebné informace a připojit kolekci, kterou chcete sdílet.

Více informací o výše zmíněných funkcích a postupech můžeme najít přímo na stránkách podpory aplikace Transkribus, kde je ale vše popsáno v anglickém jazyce.¹¹⁵

3.3 Výhody a nevýhody aplikací sloužících k automatickému přepisu

Hlavní nevýhodou aplikací automaticky přepisující text je jistě jejich špatné učení se tomu, jak nalézt a rozpoznat iniciály. Vzhledem k jejich velikosti oproti normálnímu písmu a také různorodosti, jakými se zapisují, bychom totiž museli model učit jednotlivé iniciály zvlášť a několikrát. Samozřejmě, že to není nemožné, Transkribus i jiné aplikace by byly schopny se naučit číst iniciály, avšak potřebovali bychom k tomu dokumenty obsahující je ve

¹¹⁵ Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://help.transkribus.org/>.

velké míře, a přesto by se mohlo stát, že přečtení nebude dokonalé, vzhledem k výše zmíněným faktorům.

Dalším záparem by pro některé uživatele mohlo být anglické rozhraní většiny aplikací zaměřující se na automatizaci přepisu. Přestože Projekt Pero je českým projektem, stále se jedná o menší aplikaci, která nemá takové možnosti a funkcionality jako například Transkribus, jehož rozhraní je celé v anglickém jazyce.

Nevýhodou také může být přesnost modelů, přestože jsou mnohdy účinnější než manuální přepis. Technologie dokáže eliminovat lidské chyby, které se mohou vyskytnout při opisování. Na druhou stranu automatický přepis stále není dokonalý – špatná kvalita digitalizátu, škrtnání či pozdější připsané poznámky působí automatickému přepisu problém.

Nevýhodou se může stát i formátování, které při automatickém přepisu zůstává stejné jako v originále. Pokud tedy plánujeme dokument publikovat, zachování původního formátu textu může vyvolat komplikace s následnými úpravami – pokud však bude používat výstupní formát Word a podobné, tato komplikace se zde nenachází.

Možná, že někteří uživatelé nebudou brát tuto informaci za negativum, avšak jelikož se tato práce zaměřuje na širokou veřejnost, musím to zde zmínit. V dnešní době existuje sice mnoho publikací a článků, které se věnují umělé inteligenci a automatickému přepisu obecně, avšak nejsou psány v českém jazyce. Jak je totiž patrné z doplňující literatury k jednotlivým aplikacím sloužícím k automatickému přepisu výše, literatura momentálně existuje primárně v anglickém, nebo německém jazyce. Znamená to tedy, že většina návodů a manuálů k jednotlivým aplikacím by mohla být pro některé uživatele nedostupná – i z toho důvodu jsem věnovala jednu podkapitolu popisu fungování aplikace Transkribus s úmyslem vytvořit univerzální návod k této aplikaci v českém jazyce.

Pokud přejdu již konkrétně k aplikaci Transkribus, ještě větší nevýhodou, než je pro některé uživatele anglické rozhraní, je online prostředí, ve kterém není možnost pracovat offline – bez přístupu k internetu. I když si můžeme daný dokument stáhnout do počítače, jak jsme si vysvětlovali ve třetí kapitole, stále potřebujeme k přepisu dokumentu přístup k internetu.

Jak jsem již zmínila výše, záparem aplikace Transkribus jsou také v nedávné době zavedené placené plány, bez nichž je možno v aplikaci pracovat pouze s omezenými funkcemi, a dále čas potřebný k vykonání úkonu ze strany aplikace. Zároveň je ale důležité zmínit, že je aplikace Transkribus k dispozici zcela zdarma do doby, dokud si uživatel tvoří vlastní modely, až následné automatické rozpoznávání textu stojí kredity. Pokud by tedy uživatel chtěl pouze aplikaci, do které si může ručně přepisovat texty s kvalitním a přehledným uspořádáním řádků, Transkribus je při této variantě bezplatný. Samozřejmě si ale také uvědomuji, že takto velký projekt by se bez finanční opory neobešel.

Abych však pouze nepoukazovala na nedostatky aplikací sloužících k automatizaci přepisu, jejich velikou výhodou je samozřejmě jejich primární funkce – umožnit rychle a efektivně zpracovávat velký objem textu. Existují modely, které dokáží s téměř nulovou chybovostí přečíst tiskové texty, včetně tiskového švabachu nebo fraktury, a nebojím se říci, že se k podobnému číslu dostaneme i u ručně psaných textů, pokud se začne do vytváření modelů aktivně zapojovat stále více jedinců a institucí.¹¹⁶

¹¹⁶ Velmi dobře je například zpracovaná německá bastarda, jejíž model se s hodnotou CER dostal k 2,3 %, nebo německý kurent s CER 4,0 %. K nalezení zde: Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-21].

Výhodou samotné aplikace Transkribus je tzv. Virtual Keyboard (virtuální klávesnice), kterou nalezneme po otevření strany dokumentu v dolní liště. Tato funkce umožňuje přidat jakýkoli libovolný znak do její paměti, a tak při vytváření přepisu můžeme do ní složitější a neobvyklé znaky nahrát a nemusíme je stále pouze kopírovat – tzn. pokud se v přepisovaném textu nachází speciální znak, např. ē, nemusíme jej pokaždé kopírovat, ale stačí jej nahrát do virtuální klávesnice a z ní poté daný znak vyvolat kliknutím na něj, což velmi zrychluje a zjednodušuje přepis starších nebo složitějších textů, ve kterých se nachází dnes již nepoužívané znaky. Soupis některých těchto speciálních znaků se nachází v aplikaci „Automatická transkripce“, která je výstupem této práce, přesněji se znaky nacházejí na stránce „Vlastní transkripční pravidla“ v oddílu „Příručka pro zadávání vlastních transkripčních pravidel“.

Automatizace přepisu také snižuje náklady související s manuálním přepisem textů. Může tak dojít k přesunu potřebných prostředků do jiných sfér, například k hlubšímu zkoumání samotných textů a podobně. Samozřejmě je stále zapotřebí znalců paleografie, kteří by museli připravit modely a rozšiřovat stávající, ale ti, kteří chtějí primárně studovat obsah textu, by tímto dostali možnost získat kvantitativní přísun dat připravených ke zkoumání.

Nelze také opomenout následující výhodu automatického přepisu, a tou je umožnění přístupu k pramenům i široké veřejnosti. Neboť dojde-li společnost do fáze, kdy budou v aplikacích umožňujících přepis existovat kvalitně zpracované modely k různým jazykům a z různého období, může se tak každý jedinec dostat k pramenům, které by jinak bez paleografického vzdělání nedokázal přečíst.

Posledním faktorem, který bych ráda zmínila, je možnost trénování vlastních modelů, neboť z dříve uvedených aplikací sloužících k automatizaci přepisu umožňují tuto funkci pouze aplikace Transkribus, eScriptorium a Diptychon. Vidím to jako obrovskou výhodu pro tyto tři aplikace, neboť možnost vytvořit si model na konkrétní dokument, druh písma a styl písaře je neopomenutelným kladem – zpřístupňuje se nám tím možnost kvantitativně a kvalitativně zpracovávat dlouhé a paleograficky náročné texty, které jsou například poznamenané osobitým písařským stylem.

4 Možnost využití transliterace a transkripce

Vzhledem k biblickému zaměření této práce – využití Bible padeřovské a svatováclavské –, bych v této kapitole ráda zmínila některá možná online rozhraní, která využívají přepis Bible, neboť se jedná o jeden z možných výstupů, jak téma této práce – automatizace přepisu – využít v praxi a zužitkovat jej.

Mezi digitální stránky zaměřující se na přepis Bible patří například aplikace Diabible,¹¹⁷ Ebible,¹¹⁸ VulSearch,¹¹⁹ nebo Vokabulář webový.¹²⁰ To vše jsou webové aplikace zaměřující se, nebo alespoň částečně, na shromáždění textů z překladů Bible. Více se o těchto aplikacích můžeme dozvědět na jednotlivých stránkách aplikací.

Automatizace přepisu s sebou přináší možnost kvantitativního zpracování textů. Rozšíření aplikací, které automaticky přepisují texty, by přineslo velký obrat při kvantitativních výzkumech. Rychlé shromáždění velkého objemu dat posune výzkumy dopředu. Samozřejmě je zde ale úskalí, že automaticky transliterované texty obsahují chyby, přestože již lze v dnešní době dosáhnout téměř zanedbatelné chybovosti. Přesto však s sebou tato varianta přináší možnost, aby se odborníci mohli více soustředit na badatelskou činnost namísto čtení a přepisování. Pokud se podíváme na samotné využití přepisu textů, můžeme je vidět například ve formě edic – kterým bude věnována následující podkapitola –, slovníků, korpusů a mnoho dalšího. Obecně také větší počet zpřístupněných historických památek může dopomoci hlubšímu výzkumu jazyka a vývoje obecně, badatel by tak dostal šanci věnovat se zmíněnému bádání. Proto si myslím, že zpřístupňování stále většího počtu dokumentů, které jsou pro mnohé odborníky i širokou veřejnost paleograficky nepřístupné nebo časově náročné, je důležitým a neopomenutelným posunem vpřed, který čeká pouze na své přijetí a rozšíření v okruhu akademické obce a na seznámení široké veřejnosti s těmito možnostmi.

4.1 Tradice edičního přepisu starých českých textů

Tato podkapitola přiblíží ediční zásady a tvorbu edic. Budu se opírat o pět stěžejních děl týkající se popisu edičních zásad o českém jazyce a zároveň blíže nastíním jednu z variant možností využití automatizace přepisu. Zmiňovanými publikacemi jsou díla Jiřího Daňhelky *Směrnice pro vydávání starších českých textů* (1985),¹²¹ *Obecné zásady ediční a poučení o starém jazyce českém* (1957)¹²² a *Obecné zásady ediční a poučení o češtině 15. století*

¹¹⁷ Diabible [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka 2023. Aktualizace dat 19. 12. 2022 [cit. 2024-04-01]. Dostupné z: <http://diabible.com>.

¹¹⁸ The Bible: e-Bible [online]. [cit. 2024-04-01]. Dostupné z: <https://ebible.com/>.

¹¹⁹ Clementine Vulgate Project: The full text of the Clementine Vulgate, freely available online [online]. 2002-2005. Aktualizace dat 22. 10. 2022 [cit. 8. 3. 2022]. Dostupné z: <https://vulsearch.sourceforge.net/>.

¹²⁰ Vokabulář webový: Webové hnízdo pramenů k poznání historické češtiny [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka. © 2006–2020. Verze dat 1.1.15 [cit. 2024-04-01]. Dostupné z: <https://vokabular.ujc.cas.cz>.

¹²¹ J. Daňhelka: *Směrnice*.

¹²² DĀNHELKA, Jiří, *Obecné zásady ediční a poučení o starém jazyce českém*. In: Výbor z české literatury od počátků po dobu Husovu. Praha, Nakladatelství Československé akademie věd 1957, s. 25–35. (Dále jen: J. Daňhelka: *Obecné zásady o starém jazyce českém*.)

(1963),¹²³ dále text Josefa Vintra *Zásady transkripce českých textů z barokní doby* (1998)¹²⁴ a *Zásady vydávání novověkých historických pramenů z období od počátku 16. století do současnosti* vydané pod vedením Ivana Šťovíčka (2002).¹²⁵ Zajímavou polemikou je také článek Lenky Jirouškové *Edice versus transkripce aneb zapomenutá ediční praxe?* (2007), ve kterém autorka naráží na problematiku malého počtu vydávání nových edic a nedostatečných prostředků k jejich publikaci.¹²⁶ Autorka popisuje, jak se vydané prameny mohou zužitkovat, jak je lze interpretovat a zkoumat je. Nakonec Jiroušková píše, že je nutné znovuobnovit diskuse o formě vydávání edic a edičních pravidel. S tímto názorem musím nadměru souhlasit. Vezměme si, jak vypadají ediční vydání různých pramenů. Každý editor tvoří edici dle vlastního nejlepšího svědomí a uvážení. Mnohdy dochází k tomu, že editor provede jen některé fonologické změny, přestože se rozhodl vytvořit edici s transkribovaným pramenem. Proč tedy existují edice pouze částečně transkribované? Je nutné si vždy uvědomit, kdo je koncovým uživatelem, pro koho edici nebo přepis tvořím. Pokud se jedná o širokou veřejnost, proč bychom měli pouze některé fonologické změny provést? Pokud chci seznámit publikum, které je historickým vývojem „nepolíbené“, s danou památkou, proč nepoužíváme dnešní transkripci pouze například se zachováním původních tvarů sloves a podobných jevů, když chci zachovat autentičnost? Pokud však tvořím edici pro znalce, pro vědecké pracovníky, proč jim nepředám jednotnou edici, která je buď také celá upravená dle platných fonologických zásad, nebo naopak vytvořená tak, aby změny odpovídaly originálu? Protože pokud chce vědec zkoumat danou památku, její písmo, formu a tvar, není spíše vhodné mu předložit transliterovaný text pouze změněný dle pravopisných norem, aby obsahoval dnešní tvar daných znaků? Tím je myšleno s použitím dnešních platných grafémů. Z toho důvodu si myslím, že je nutné diskutovat o tvoření edic a edičních zásadách, neboť nejednotnost a individuálnost často vedou ke zmatku.

Nyní ale zpět k publikacím zmíněným dříve. Jak je patrné z dat vydání jednotlivých edičních zásad, poslední publikace týkající se tohoto tématu vyšla už v roce 2002, tedy již před téměř čtvrt stoletím, i to podporuje názor Lenky Jirouškové o potřebě znovuobnovit diskusi o edičních zásadách¹²⁷ – výjimku tvoří *Metodika přípravy a zpracování elektronických edic starších českých textů* (2015) od Aleny Černé a Borise Lehečky, která je zaměřena na elektronické edice, jak je patrné ze samotného názvu.¹²⁸ Více ji však rozebírat nebudu, neboť tato podkapitola primárně komparuje problematiku kolem edic vydaných tiskem, neboť tyto publikace mají jiná pravidla, než je tomu u elektronických edic.

Při tvoření edic dochází k několika problémům a jedinečným, kterým musí editor věnovat svou pozornost. Jedná se například o nesrozumitelnost dřívějšího jazyka pro dnešní

¹²³ DAÑHELKA, Jiří, Obecné zásady ediční a poučení o češtině 15. století. In: Výbor z české literatury doby husitské. Svazek první. Praha, Nakladatelství Československé akademie věd 1963, s. 31–41. J. (Dále jen: Daňhelka: Obecné zásady o češtině 15. století.)

¹²⁴ VINTR, Josef: Zásady transkripce českých textů z barokní doby. Listy filologické 121, 1998, s. 341–346. (Dále jen: J. Vintř: Zásady.)

¹²⁵ I. Šťovíček: Zásady.

¹²⁶ JIROUŠKOVÁ, Lenka. Edice versus transkripce aneb zapomenutá ediční praxe? Listy filologické, 2007, roč. 130, č. 3-4, s. 373-380. ISSN 0024-4457. (Dále jen: L. Jiroušková, Edice.)

¹²⁷ L. Jiroušková, Edice.

¹²⁸ ČERNÁ, Alena M., Boris LEHEČKA. Metodika přípravy a zpracování elektronických edic starších českých textů [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka 2015, aktualizace 11. 2. 2016 [cit. 2024-04-01]. Dostupné z:

https://vokabular.ujc.cas.cz/soubory/nastroje/Methodics/Methodika_pripravy_a_zpracovani_elektronickyh_edic_DF12P01OVV028.pdf. (Dále jen: A. Černá, B. Lehečka, Metodika.)

čtenáře – to samozřejmě obnáší fakt, že editor musí být seznámen s vývojovými stádii českého jazyka, aby dokázal určit, rozpoznat a správně pracovat s vývojovou tendencí dané památky. Dále se problémem může stát nezachovalost původního rukopisu – tedy zachování pouze jeho opisů, které podlely mnohokrát opisovačským nedbalostním chybám –, z toho důvodu může dojít k nepřesné interpretaci vývoje a podoby češtiny té doby. Mezi další problematiku můžeme řadit otázku autenticity textu – zda se jedná o naprosto prvotní vyjádření obsahu sdělení, nebo o opisy někdy i několikáté generace, které mohly za ten čas podlehnout mnoha změnám a posunům. Taktéž editoři musí operovat s problematikou spojenou s písaři či opisovači, například písařovo dodržování dobových norem, které bylo mnohdy nekonzistentní z důvodu neznalosti nebo nedbalosti, dále nepevný písařský úzus, jeho technické vybavení a podobně. V neposlední řadě se editor setkává s velkým problémem rozporu mezi obsahem a formou – Daňhelka v článku zdůrazňuje, že editorova práce není mechanické přepisování textu. Editor dle autora musí pochopit hloubku obsahu a formy a nalézt nejlepší způsob, jak je zkombinovat.¹²⁹ Nástrahou pro editory v přepisu památky z doby barokní dle Josefa Vintera může být i zachovávání protetických a hiátových hlásek („v“ a „h“), dále psaní souhláskových skupin podle originálu (např. ouředník, vzáctný), psaní předložek a předpon dle originálu („s“ a „z“), nebo zachování rozdílu mezi „ň-“ a „n-“ a „sem/sou/sme“ oproti „jsem/jsou/jsme“. V neposlední řadě píše o důrazu na zachování nářečních variant hlásek v textu. Také zmiňuje, že v kritické edici se musí uvádět každá odchylka od uvedených zásad v kritickém aparátu, ve čtenářské edici stačí změny obecně formulovat v ediční poznámce, přičemž doporučuje, aby se edice vydávaly pro obě tyto skupiny čtenářů dohromady – pro širokou veřejnost i odborníky.¹³⁰ Oproti tomu Daňhelka spíše doporučuje oddělování těchto dvou skupin a psaní edic zvlášť s přesným zacílením na danou skupinu, o čemž jsem psala výše.¹³¹

Daňhelka rozlišuje dva způsoby, jak je možné přistoupit k vydávání textu. Zaprvé buď jako k textu, tedy s respektem k jeho stavu jakožto výrazu doby, kdy vznikla jeho realizace, kterou má editor před očima – tj. kdy byl text opsán nebo napsán –, nebo zadruhé jako k památce, tj. ke konkrétnímu ztvárnění památky promítané do doby, kdy vznikla – kdy byl archetyp (předloha) vzniklý následným opakovaným opisováním vyvinut do podoby, kterou máme k dispozici. Taktéž autor píše: „*Ediční práci se nelze naučit pouhým osvojením edičních zásad a s nimi i transkripčních pravidel.*“¹³² Přestože se má předkládaná práce pokouší o přesný opak Daňhelkova tvrzení, jejím cílem nebylo a není jakkoli snižovat práci editorů, pouze se pokouší o poukázání na možné urychlení a zjednodušení jejich práce, neboť i přes případné osvojení si automatické transliterace a transkripce budou muset editoři dané texty pročíst a případné chyby vytvořené umělou inteligencí opravit, společně s případným opravením obsahu odpovídající formě, bude-li to nutné.

Edice se vydávají buď z důvodu, aby sloužily k vědeckému bádání, nebo jako vydání seznamující širší veřejnost s památkou. Existují různé druhy edic – podle Daňhelky se jedná o edice vědecké (vědecko-informativní) a čtenářské (kulturně-vzdělávací).¹³³ Pokud však edice není vydána podle vědeckých zásad, jedná se o parafrázi (překlad). Starší památky se mohou vydávat dvojím způsobem – buďto psány moderním (diakritickým) pravopisem, nebo paleograficky, což je cenné hlavně pro filologické bádání. Opět zde však narážíme na výše zmíněný problém toho, že mnozí editoři vydávají své edice na pomezí těchto dvou rovin –

¹²⁹ J. Daňhelka, Směrnice.

¹³⁰ J. Vintř: Zásady.

¹³¹ J. Daňhelka, Směrnice.

¹³² Týž.

¹³³ Týž.

někde mezi paleografickým a diakritickým čtením a přepisováním. Textové poznámky, které se přikládají k edici, založené na textologické analýze a transkripci, jsou klíčové pro porozumění edice. Také slouží jako zdroj pro možné přidání kritiky editorovy práce. Tyto ediční poznámky jsou zpravidla dvoudílné, přičemž jsou tyto dvě části od sebe graficky oddělené. První část těchto poznámek tvoří forma daného čtení, jak je upravená editorem, a vysvětlení, o jaké texty se daná úprava opírá. Druhá část – původní – obsahuje originální podobu pramene, například rukopisu. Při editorově úpravě se doporučuje, aby změny vycházející z jeho úvahy byly označeny zkratkou jeho jména objasněnou mezi zkratkami a značkami, stejně tak jako jménem označit úpravy, které vycházejí z odborných studií jiných autorů. Samotné poznámky by měly být nedílnou součástí každé edice a zpracované tak, aby objasňovaly metodu editorovy práce a umožňovaly případnou kritiku edice.¹³⁴

S publikací Jiřího Daňhelkovi lze porovnat publikaci Ivana Šťovíčka, který dělí edice dle typu, druhu a formy.¹³⁵ Dle typu na vědecké, vědeckopopulární, úřední, osvětové a didaktické. Dále edice hlavně vědecké a vědeckopopulární rozřazuje podle druhu, jehož určení závisí na pramenné kvalitě, provenienci a obsahu – vyplývá to z vnitřních a vnějších znaků samotných dokumentů. Mezi druhy edic patří edice jednoho druhu dokumentů, jednofundová, vícefundová, tematická, dokumentů jednoho původce či jedné osobnosti, periodického a neperiodického tisku, dokumentů nepísemné povahy a edice speciální vzniklé z potřeby studia pomocných věd historických. Poslední dělení je podle formy – kritériem je zde způsob zpřístupnění dokumentů a vnější podoba edice. První kritérium autor pojmenovává jako formu edic dle způsobu zpřístupnění dokumentů a řadí sem edice reprodukcující celé dokumenty, přepisující texty dokumentů, publikující zkrácené texty dokumentů, publikující edičně zpracovaný obsah dokumentů a edice kombinované využívající více forem reprodukcí. Druhá část rozdělení dle formy je roztríděna dle techniky a způsobu vydání, přičemž tato skupina obsahuje rozdělení na formu edice dle techniky reprodukce dokumentu a publikování jejich textů (edice vydané tiskem, technikou fotoreprodukce apod.) a edice dokumentů dle rozsahu a způsobu vydání (edice jednoho samostatně zveřejněného dokumentu, několika dokumentů aj.). Šťovíček dále v této publikaci velmi podrobně popisuje práce spojené se vznikem edice – heuristiku, výběr dokumentů, uspořádání a zpracování jich v edici, také se věnuje možnosti reprodukce dokumentů a samozřejmě i edičním značkám.¹³⁶ Zajímavými kapitolami jsou i samotné návody na přepis českých textů vzniklých od 16. století do roku 1848 a pro dokumenty vzniklé po roce 1848. Také se hojně věnuje přepisu cizojazyčných textů, zvláštěm vzniklých při přepisování, zpřístupňování obsahu a vědeckoinformačnímu aparátu edice.¹³⁷

Při přepisu pramene do diakritického pravopisu je nutné dbát na to, aby se odstranila stará grafika rukopisu či tisku, společně se vši primitivností a neustáleností. Je ale důležité zvážit, zda se jedná o záležitost grafickou, či jazykovou. Taktéž se texty jiného druhu písma přepisují do latinky a celý dokument se opatří interpunkcí obvyklou dnešní češtině, ale nikoli na úkor specifčnosti památky – „*nemodernizujeme však za každou cenu, nýbrž snažíme se uchovat členění originálu*“.¹³⁸ Pro filologické edice platí, že tištěné texty musí být na okraji očíslovány

¹³⁴ J. Daňhelka, Směrnice.

¹³⁵ I. Šťovíček: Zásady.

¹³⁶ [?]: Označení nejistého čtení textu. [!], [sic]: Upozornění na chybný text, věcnou chybu, chybu v dataci apod. []: Editorové zásahy do textu. [...]: Označení editorem vypuštěného textu. [.....]: Označení chybějícího nebo nečitelného textu (počet teček odpovídá počtu chybějících znaků). //, <>: Zásahy autora dokumentu do textu. (Viz I. Šťovíček: Zásady.)

¹³⁷ Týž.

¹³⁸ J. Vintř: Zásady.

po pěti řádcích nebo verších, přičemž následná čísla slouží k odkazování v poznámkovém aparátu, stejně tak jako paginování a foliování na okraji. V transliterovaných textech se rozepisují zkratky, číslice se nechávají v původním tvaru, slova se píší odděleně ve smyslu základních tendencí jazyka v díle a stejně jako v transkribovaných textech se do přepisu vkládá interpunkce dle dnešní normy, výjimkou mohou být opět rukopisy s vlastní interpunkční normou.¹³⁹

Směrnice Jiřího Daňhelky obsahuje soupis s přehledem problematických znaků v jednotlivých systémech pravopisu – viz tabulka (Tabulka 6) níže.¹⁴⁰ Můžeme si všimnout, že ne všechny změny v tabulce od Jiřího Daňhelky se shodují s aplikací „Automatická transkripce“, kterou jsem vytvořila z dat zaznamenaných v tabulkách ve druhé kapitole této práce (soupis těchto dat lze také vidět v příloze viz Obrázek 27). Ty byly vytvořené ze zdrojů: Nový encyklopedický slovník,¹⁴¹ Dějiny českého jazyka od Miroslava Komárka,¹⁴² Humanistická čeština od Jaroslava Poráka¹⁴³ a Spisovný jazyk v dějinách české společnosti od Dušana Šlosara.¹⁴⁴ Možnou příčinou může být datum vydání zmiňované Směrnice (1985), neboť zbylé zmíněné publikace byly vydány až ve 21. století – tedy kromě díla Humanistická čeština. Dále také může být příčinou různorodost pohledů filologů a historiků na tvorbu edic – v tomto ohledu poměrně rozdílné. V neposlední řadě to také mohla zapříčinit samotná nesourodost a neustálenost edičních zásad i vzhledem k posunu vědy v oblasti historického vývoje českého jazyka.

Hlásk	Jednoduchý p.	Spřežkový p.	Husův p.	Bratrský pravopis
á	a	a, aa, a	á	á
é	e	e, ee	é	é
ě	e, i	ie, ye	ie, ie, e, ě	ě
ie	e, i	ie, ye; iee, yee; yie, ije, íje	ié, íé; ýe, ýé; iě, ye	–
i/y	i, y	i, y, (často) í (s vlasovou čárkou)	i, y, í	i/y (rozlišuje se ale „y“ jako spojka a po „c“, „z“, „s“)
í/ý	i	i, y, ě, y, í, ý; y, ii, íi, ij, yy, yí, ýj, yg	í, ý	j; ij ý
ó	o	o, oo, o ^o , ó	ó	–
u	u	u, v, w	u, v, w	u, v-
ú	u	u, ú, uu, uv, vv	ú	v-
ou	–	–	au; ou	au
r	r	r ev. se zvl. fon. realizací: er, re, ri, ur, ru, ro	r	r
l	l	l ev. se zvl. fon. realizací: el, li, il, lu, ul, lo	l, l	l, l (rozlišeno)

¹³⁹ J. Daňhelka, Směrnice.

¹⁴⁰ Týž.

¹⁴¹ Nový encyklopedický slovník [online]. [cit. 2024-03-29]. Dostupné z: <https://www.czechency.org/>.

¹⁴² M. Komárek, Dějiny českého jazyka.

¹⁴³ J. Porák, Humanistická čeština.

¹⁴⁴ D. Šlosar, Spisovný jazyk.

ř	r, rs	rz, rs, rzs, rzss; zr, sr, zrs; rzz, rž, rz, rsz; rzy (po zániku jotace)	ř, ř	ř, ř
ň	n	ny; ni (po zániku jotace) n, nh; yn; ng	ň	(ně, né, ni)
t	t	t, tt, th	t	t
ť	t	t, th, tt; ty, ti (po zániku jotace); tg	ť, ť; t	(tě, te, ti)
ď	d	d, dh, dd; dy, di (po zániku jotace); dg	ď, đ, d	(dě, de, di)
v	v, u	u, uu, v, vv, w, wu	w, v, u	w
f	f	f, ff, ph, pf	f, ff, ph	f
k	c, ch	k, c, ch, ck	k	k
ch	h	h, ch	ch	ch
s	s, z	z, zz, f; s, sz, szz, ss	s, f	s, f
z	z, s	z, s, zs, ss	z	z
c	c	ch, cc, cz, cs; ts	c	c
š	s, z, f	s, z; ss, f, ff	ś, š, f	-š, ff
ž	z, f	z, s, zz, f	ž, ž	ž, ž
č	c, s, f	cs, sc, zs, sz; ch, chz, sch; cz	ć, č	ć, č
j	i, y	i, y, g; gh, gi, gy (po zániku jotace)	g, y	g, ale ya
g	g	g	g	ğ

Tabulka 6 – J. Daňhelka: Přehled problematických znaků v jednotlivých systémech pravopisu.

Jelikož výstupem této práce je aplikace, která se pokouší být univerzálním nástrojem pro automatickou transkripci jednotlivých vývojových etap českého jazyka, změny navržené Daňhelkou zohledním a vytvořím pro ně samostatnou transkripční stránku. Přesto si ale ve zmiňované aplikaci můžeme všimnout, že finální výstup je poměrně chaotický z důvodu uvádění mnoha variant pro jeden znak. Aplikace tak zpracovává jeden grafém jako několik možných variant jednotlivých písmen.

Téma vydávání edic, vytváření edičních zásad, řízení se jednotnými grafickými značeními a podobná témata s tím spojená je téma velmi rozsáhlé a pro zpracování náročné. Tato práce pouze okrajově popisuje danou problematiku – nejedná se tedy o návod k vytvoření edice –, neboť samotná edice je pouze jednou z možných variant, jak by mohl být využit výstup této práce – automaticky transliterovaný nebo transkribovaný text. Více se proto můžeme dočíst ve výše zmíněných zásadách, které jednotlivé jevy popisují podrobně, hlavně tedy nejnovější publikace od Ivana Šťovička¹⁴⁵. Proces vytváření elektronické edice poté může nalézt v publikaci od Aleny Černé a Borise Lehečky.¹⁴⁶

¹⁴⁵ I. Šťoviček, Zásady.

¹⁴⁶ A. Černá, B. Lehečka, Metodika.

5 Praktická část

5.1 Model Bible padeřovské a svatováclavské

Bibli padeřovskou si nechal vyhotovit Filip z Padeřova skrze písaře Jana z Prahy ve 30. letech 15. století. Jedná se o iluminovaný pečlivě psaný rukopis (dle Vladimíra Kyase psán pomocí diakritického pravopisu, přesto ale při zkoumání originálu můžeme vidět, že se jedná o pravopis spřežkový). Jedná se o bibli III. redakce.¹⁴⁷

Oproti tomu Bible svatováclavská je tisk z období baroka, přesněji Nový zákon pochází z roku 1677, Starý zákon z let 1712 až 1715. Bible vyšla pod nakladatelstvím Svatováclavského dědictví, které bylo založeno Matějem Václavem Šteyerem a jeho matkou Marií Šteyerovou. Samotná rodina vložila kapitál na vydání této Bible, dále přispěli ještě vlastenečtí jezuité Jan a Matěj Tannerovi, literát a historik Bohuslav Balbín, Jiří Cruger a František Retz. Výchozím textem se stal starý tisk Bible benátské – IV. redakce – z něhož bylo odstraněno vše, co se neshodovalo s latinskou Sixto-Klementinou, a všechna „jest“ a „jsú“ ve třetích osobách préterita. Konstanc a Šteyer společně pracovali na překladu Nového zákona, jeho vydání se však Konstanc nedožil. Obdobně tomu bylo i s vydáním Starého zákona, jehož vydání se nedožil Šteyer – jeho práci převzal Jan Barner. Roku 1715 tak konečně vyšla první úplná tištěná česká katolická bible vytisknutá pod názvem: „*Bibli česká, tj. Svaté Písmo podle starožitného a obecného latinského od všeobecné Církve svaté římské v Starém Městě Pražském v Koleji Sv. Klimenta Tovarýšstva P. Ježíše... léta MDCCXV*“.¹⁴⁸

Model vytvořený z dat přepisu Bible padeřovské (Model: Old Czech Handwriting (with/without spaces))¹⁴⁹ v aplikaci Transkribus vznikl v roce 2022 během HTR (Handwriting text recognition) Winter school¹⁵⁰ díky kolektivu jedinců pod vedením Mgr. Anny Michalcové.¹⁵¹ Tento model byl vytvořen jak na transliterovaných stranách Bible padeřovské, tak i Bible drážďanské, čímž se stal více univerzálním modelem pro česky psaná písma středověkou bastardou.¹⁵² I z toho důvodu jsem si zvolila Bibli padeřovskou jako jeden ze dvou příkladů textů, které podrobím testu automatické transkripce. Dalším důvodem je fakt, že Bible jakožto „kniha knih“ se stala základním pramenem zkoumání, a právě její velký rozsah a málo písařů je ideálním zdrojem k využití testování a následnému aplikování automatického přepisu. V 16. a 17. století také Bible patřila k nejvydávanějším knihám vůbec. Do roku 1620 bylo

¹⁴⁷ V. Kyas: Česká bible.

¹⁴⁸ V. Kyas: Česká bible.

¹⁴⁹ Transkribus: Old Czech Handwriting [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/model/old-czech-handwriting-with-spaces/>, <https://readcoop.eu/model/old-czech-handwriting-without-spaces/>.

¹⁵⁰ Článek o HTR Winter School viz Österreichische Akademie der Wissenschaften: Discovering the power of AI for reading medieval manuscripts: HTR Winter school 2022 [online]. [cit. 2024-04-01]. Dostupné z: <https://www.oew.ac.at/imafo/read/discovering-the-power-of-ai-for-reading-medieval-manuscripts-htr-winter-school-2022>.

¹⁵¹ Michalcová, A. (2022) „Padeřov-Bible-handwriting-ground-truth: Initial release“. Zenodo. doi: 10.5281/zenodo.7467034. [cit. 2024-03-31]. Dostupné z: <https://zenodo.org/records/7467034>.

¹⁵² Bastarda je písmo vzniklé ve 14. století na přechodu mezi pravidelnými a strnulými formami gotickými a volnějším, umírněně lomeným písmem novogotickým. Písmový obraz bastardy působí dekorativně, zdobně a vzdušně. K některým z jeho rysů patří jednobříškové minuskulní ‚a‘, dvoubříšková konstrukce ‚b‘ a ostře sbíhající dřívky ‚f‘ a ostrého ‚s‘. (Viz VOIT, Petr. Encyklopedie knihy: starší knihtisk a příbuzné obory mezi polovinou 15. a počátkem 19. století. Bibliotheca Strahoviensis. Praha: Libri ve spolupráci s Královskou kanonií premonstrátů na Strahově, 2006.)

vydáno celkem 15 českých překladů Bible, 27 vydání Nového zákona, 1 vydání Starého zákona a další desítky jednotlivých výtisků vydávající jednotlivé knihy Bible.¹⁵³ Z důvodu uchování celistvosti a jednotnosti této práce a následného testu jsem si za druhý text vybrala mladší tisk – Bibli svatováclavskou. U Bible padeřovské jsem zvolila jeden list ze Žaltáře, který jsem dříve jako jeden z několika přepisovala při zmiňované HTR Winter School. Samotná Bible padeřovská se nachází v Rakouské národní knihovně, která umožňuje volně nahlédnout do digitalizátu.¹⁵⁴ U Bible svatováclavské jsem přepsala první dva úvodní listy z dostupného elektronického digitalizátu exempláře z aplikace Manuscriptorium. Digitalizát patří Muzeu Polička.¹⁵⁵

5.2 Transkripční model aneb univerzální model pro starou češtinu

Jak již bylo psáno dříve, v dnešní době existují aplikace, které dokáží automaticky transliterovat text s minimální chybovostí. Aplikaci Transkribus byla věnována celá kapitola popisující její fungování. Cílem této práce ale bylo taktéž zjistit, zda by mohl být vytvořen model, který by automaticky transkriboval určený text – zda by mohl vzniknout univerzální model pro transkripci pramene z určitého období podle známých transkripčních pravidel. Pokusila jsem se proto vytvořit různé modely a aplikace, které se o to pokouší.

Je nutno zmínit, že téměř všechny navrhované varianty, které uvádím v příkladu, potřebují jako vstupní komoditu transliterovaný text, který by nám mohl v ideálním případě automaticky dodávat například Transkribus. Proto je důležité si uvědomit, že nastínění možných transkripčních automatických aplikací ve většině případech není možné bez automatické transliterace, která by umožnila kvantitativní přínos starých textů. Také je důležité zmínit, že se jedná pouze o nahlédnutí do možností automatické transkripce, z toho důvodu je nutné pochopit, že pokud by se měly navrhované možnosti přenést do každodenního používání, musí se na jejich rozvíjení podílet více odborníků.

A nyní již k samotným návrhům řešení automatické transkripce. První variantou, která je schopna automatické transkripce, je možnost vytvořit transkribovaný model přímo v aplikaci Transkribus, přičemž by se mohl model buď přímo učit psát skupinu znaků jako znak transkribovaný (Obrázek 31 v příloze), nebo transliterovaný model naučit pomocí zkratk rozepisovat určitou změnu (Obrázek 30 v příloze). Pokud to uvedu v praxi – víme, že se skupina hlásek ‚cz‘ ve spřežkovém pravopisu přepisovala ve většině případech jako ‚c‘, dále ‚ŕ‘ označovalo ‚š‘. V první variantě bychom tedy nový model v aplikaci Transkribus přímo naučili číst transliterované slovo ‚czaff‘ jako slovo ‚čas‘. Ve druhé verzi – se zkratkami – bychom označili slovo ‚czaff‘ jako abbrev. (zkratku) a do ‚properties‘ (vlastností) napsali již samotné transkribované slovo ‚čas‘. Následně po vhodném minimu ‚ozkratkovaných‘ listů bychom naučili nový model rozepisování zkratk. Tato druhá možnost je však velmi nepraktická

¹⁵³ J. Porák, Humanistická čeština.

¹⁵⁴ Vienna, Österreichische Nationalbibliothek, Cod. 1175, fol. 199v-200. Biblia bohemica, saec. XV. Dostupné z: <http://www.onb.ac.at/sammlungen/hschrift/bibliographie.htm>.

¹⁵⁵ Polička, Městské muzeum a galerie Polička, Sign. K 378, fol. V-VI. Svatováclavská bible, Starý zákon, 1. část (Genesis – Sírachovec). Dostupné z: https://www.manuscriptorium.com/apps/index.php?direct=record&pid=AIPDIG-MP18_K_378_____2Y5WNU6-cs.

a zdlouhavá, proto ji nedoporučuji, přesto ji ale uvádím, neboť by mohla být užitečná v případě částečné transliterace, která se soustředí jen na určité jevy. První možnost se však osvědčila poměrně kvalitně, přičemž i na menším počtu stran, ze kterých jsem vytvořila model, jsem získala velmi dobrý výsledek v závislosti na kvantitě vložených dat (Obrázek 32 v příloze). A jak je patrné i na druhém obrázku, model se dokázal vypořádat – vzhledem k velikosti modelu –, velmi kvalitně s pro něj neznámou stranou rukopisu (Obrázek 33 v příloze).

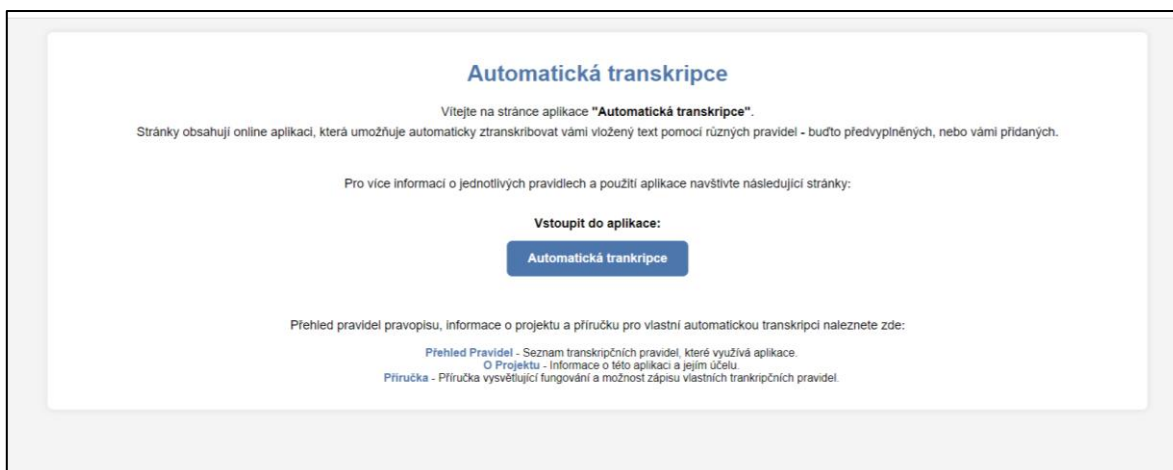
Další možností, která se nabízí, je vytvoření šablony na míru v aplikaci Word. Tato aplikace je totiž schopná prostého nahrazování určitých znaků za jiné, což je stejné jako návrh se zkratkami v Transkribus, Výhodou je, že je aplikace Word nejspíše pro většinu uživatelů cenově dostupnější, jelikož nabízí rodinné a firemní balíčky s předplatným. Musíme však vzít na vědomí, že zmíněné modely, hlavně tedy „Model – zkratky“ a „Word“, jsou velice primitivní a nedostatečné. Mohly by sloužit například k osobnímu účelu, aby uživatel mohl zkoumat v dokumentech odchylky od edičních a transkripčních zásad. Aplikace totiž prozatím nejsou schopny složitějšího uživatelského rozhraní, které by jim dovoľovalo zadávat složitější pravidla – tyto složitější příkazy můžeme v dnešní době vidět například v aplikaci Kontext od Českého národního korpusu.¹⁵⁶ Jelikož je tato navrhovaná varianta v aplikaci Word pouze jednodušší koncepcí mnou vytvořené aplikace „Automatická transkripce“, více se jí zabývat nebudu a vysvětlím pouze podrobně fungování již zmíněné webové aplikace v následující kapitole.

5.2.1 Aplikace „Automatická transkripce“

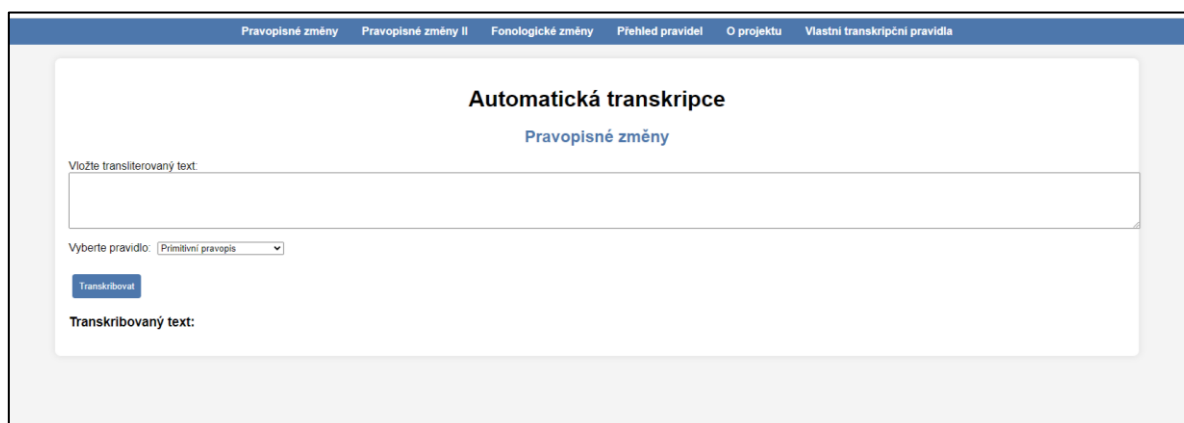
Jak již bylo několikrát zmíněno, výstupem této práce je webové rozhraní „Automatická transkripce“. Jedná se o jednoduchou online aplikaci ve formátu webových stránek, které obsahují základní transkripční pravidla. Lze ji nalézt po zadání adresy: <http://penkavova-jcu.wz.cz/>. Úvodní stránka slouží jako rozcestník a krátké vysvětlení projektu (Obrázek 22). Aplikace přesněji obsahuje tři základní stránky umožňující automatickou transkripci. První („Pravopisné změny“,¹⁵⁷ Obrázek 23) a druhá stránka („Fonologické změny“, Obrázek 25) nabízí předvyplněná transkripční pravidla, která jsem zkomparovala z nejčastějších transkripčních, fonologických a pravopisných změn jednotlivých období a která je možno vidět v dřívějších kapitolách této práce, přičemž všechny změny jsou vypsány i v samotné aplikaci na stránce „Přehled pravidel“ (Obrázek 27). Třetí stránka („Vlastní transkripční pravidla“, Obrázek 26) obsahuje volně nastavitelné prostředí vývojových změn, ve kterém si uživatel může sám zadat vlastní pravidla – taktéž zde uživatel může nalézt příručku se základními funkcemi, jak například zadávat složitější změny a seznam složitých a neobvyklých znaků ke zkopírování. Přestože by se mohlo zdát, že třetí část webové stránky by se zde nemusela vyskytovat a je přebytečnou, není tomu tak. Každý rukopis, někdy i tisk, je individuální, proto jsem vytvořila i tuto stránku, která tuto problematiku redukuje a umožňuje každému uživateli si navolit vlastní změny dle dokumentu a úzu písaře, či tiskaře. Stránka taktéž obsahuje dříve zmíněnou část („Přehled pravidel“), kde jsou sepsána všechna předvyplněná pravidla používaná na první a druhé straně.

¹⁵⁶ Český národní korpus: Kontext [online]. [cit. 2024-03-31]. Dostupné z: <https://www.korpus.cz/kontext/query?corpname=syn2020>.

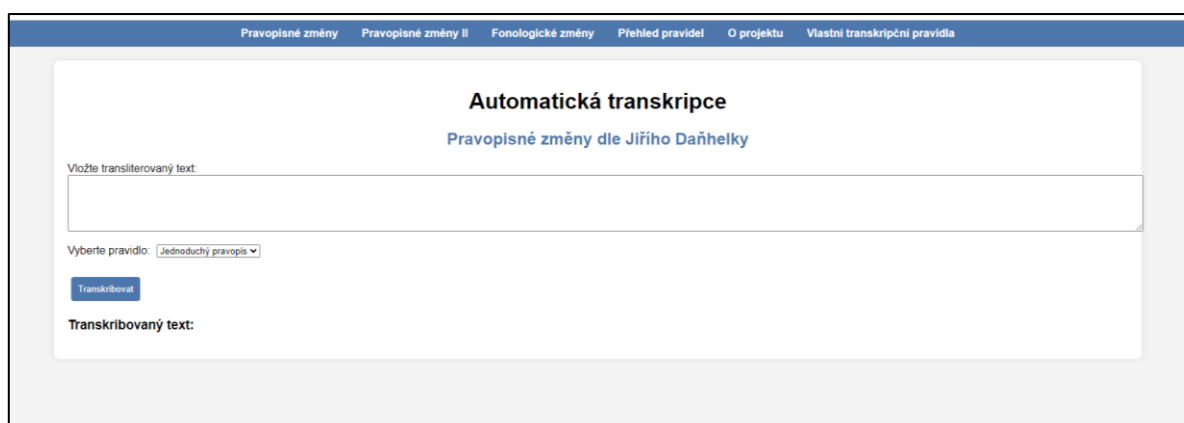
¹⁵⁷ Tato stránka obsahuje proklik na další pravidla „Pravopisné změny dle Jiřího Daňhelky“ (Obrázek 24).



Obrázek 22 – Úvodní rozcestník při zadání adresy: <http://penkavova-jcu.wz.cz/>.



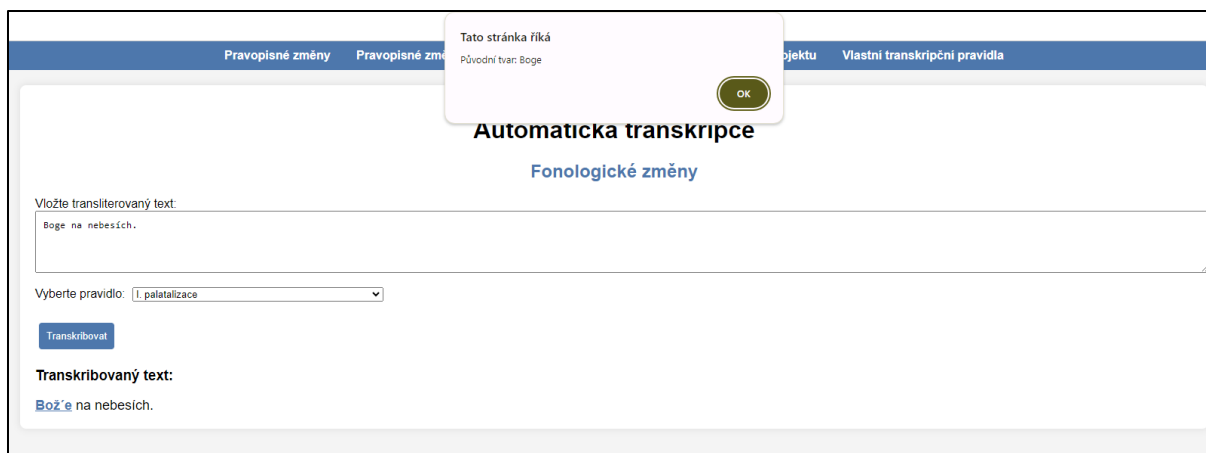
Obrázek 23 – Stránka s předvyplněnými pravidly dle druhů pravopisů.



Obrázek 24 – Stránka s předvyplněnými pravidly dle Směrnice od Jiřího Daňhelky.

Nyní již k tomu, jak aplikace „Automatická transkripce“ funguje. Do prvního pole na webových stránkách „Pravopisné změny“, „Pravopisné změny dle Jiřího Daňhelky“, a „Fonologické změny“ se zadává transliterovaný vlastní text. Následně si stačí vybrat model (pravidlo), kterým bude text transkribován – výběr usnadní výše zmíněná stránka „Přehled pravidel“, která vysvětluje, jaká pravidla model obsahuje. Následně už stačí pouze kliknout na tlačítko „Transkribovat“. Výsledný text se objeví jako text pod vloženou transliterací. Ta slova, která byla změněna, jsou označena pomocí podtržení a ztučnění. Samotné slovo lze i rozkliknout, přičemž nám následně aplikace ukáže okno s původním tvarem slova (Obrázek 28). Tato funkce byla přidána z důvodu, aby si uživatel mohl rychle a snadno prohlédnout, jak vypadal původní tvar slova, neboť se může stát, že písař daného dokumentu se nedržel dobovými pravopisnými nebo fonologickými zvyklostmi a aplikace jej tak transkribovala chybně.

Uvědomuji si, že dříve zmíněné fonologické a pravopisné změny nenastaly vždy, také z toho důvodu se jedná o univerzální nástroj, který předpokládá, že autor památky dodržoval pravidla dané doby. Jak ale víme z praxe, ne vždy k tomuto docházelo. Z toho důvodu se proto ve webové aplikaci nachází sekce sloužící k zadání vlastních pravidel („Vlastní transkripční pravidla“). Taktéž, jak již bylo zmíněno, při použití již existujících pravidel na stránkách pravopisných nebo fonologických změn lze u výstupu na dané transkribované slovo kliknout, přičemž se zobrazí původní tvar slova, abychom ihned viděli, jak vypadal původní nepřepsaný tvar a mohli jej snadno upravit.



Obrázek 28 – Rozkliknutí a odkaz na původní tvar slova u ztučněného a podtrženého slova.

Úskalím této aplikace je má nedostatečná znalost programování, neboť jsem například nebyla schopna vytvořit změny „Ztráta jerů“ a „Změna téžeslabičného diftongu“ na stránce s fonologickými změnami – prozatím jsou tedy funkce prázdné nebo nepřesné. Tento fakt však není nijak nepřekonatelný, neboť existuje mnoho zkušených odborníků, kteří se s tím dokáží snadno vypořádat – z toho důvodu by byla vhodná podpora od znalých programátorů, kteří by dokázali nedostatky výstupu eliminovat.

A jak vůbec může vypadat praktické využití jednotlivých stránek? Prvního příkladu si můžeme všimnout hned v přílohách. Jak je totiž patrné, pro přepis obou Biblií jsem využila vlastních transkripčních pravidel (Obrázek 34, Obrázek 35 v příloze). Je to z toho důvodu, že přestože se jedná o Bibli – „knihu knih“, kde bychom předpokládali držení se normy –, rukopis

i tisk jsou psané s různou měrou proběhlých a neproběhlých změn. Nedá se tedy na ně použít univerzální nástroj předpokládající přesné dodržování změn dle dané doby. Taktéž jsem využila vlastních pravidel z důvodu, abych ilustrovala, jak lze zmiňovanou stránku využít v praxi. Můžeme si všimnout, že automaticky transkribovaný model Bible svatováclavské má lepší výsledky než Bible padeřovská. Myslím si, že je to z důvodu stáří jednotlivých biblí, a hlavně z toho důvodu, že první zmiňovaná Bible – svatováclavská – je tisk, ve kterém se více dbalo na dodržování určité normy a pravidelnosti. Zde je tedy patrné, že aplikace může sloužit primárně prozatím – jelikož jsou nyní stránky v měřítku profesionálního programování velmi primitivní –, jako ulehčující nástroj při dlouhém přepisování určitých opakujících se jevů. Uvědomuji si také, že míra přepisu není dokonalá a stále by mohla být ještě propracovanější. Mým cílem však v této práci nebylo vytvořit dokonalý přepis jedné strany Bible, ale demonstrovat, že aplikace je schopna automatického přepisu na větším počtu dat. Dále se dá jistě použít webová stránka jako učební materiál, na kterém je možné snadno ukázat, jak dané změny v praxi proběhly. Taktéž lze stránku využít jako nástroj na změnu jen určitých znaků, které například do transliterované edice nechceme vkládat, neboť se nevyužívají. A v neposlední řadě lze také využít aplikaci jako nástroj pro přepis textu dle jeho fonologických změn – tím například myslím, že pokud bychom chtěli na daném textu provést ve velkém měřítku metatezi likvid, nástroj je toho schopný.

Závěr

Tato práce měla za cíl zjistit, zda lze vytvořit aplikaci nebo model, který by byl schopen automaticky transkribovat dokumenty. Jak je patrné v samotné práci a v přílohách, existuje několik možností, jak toho docílit. Nejvíce se osvědčila varianta zadávání vlastních transkripčních pravidel do vytvořené aplikace podle jednotlivého dokumentu. Bylo zjištěno, že dokument mladší, který již nemá tolik rozkolísaný pravopis a proběhlé fonologické změny, přinesl velmi perspektivní výsledek. S trochou úsilí a při zapojení většího počtu odborníků by jistě aplikace dokázala kvalitně a téměř bezchybně automaticky transkribovat texty. Pro novější texty je to určitě perspektivní cesta, jak je rychleji, kvantitativněji, a hlavně kvalitněji zpracovávat.

Druhou nejúspěšnější metodou pro automatickou transkripci se stal model v aplikaci Transkribus, které byla věnována celá jedna kapitola, v níž bylo vysvětleno, jak v dnešní době tuto aplikaci použít k získání kvalitní automatické transliterace. Ve stejné kapitole se taktéž nachází návod k aplikaci s cílem přiblížit odborníkům a široké veřejnosti toto rozhraní tak, aby jej byli schopni všichni využívat.

Výzkum, který byl součástí této práce, však zjistil, že je možné aplikaci Transkribus použít i jako nástroj pro automatickou transkripci, neboť byl vytvořen zjednodušený model schopný velmi dobrých výsledků. Také bylo zjištěno, že poměrně velkým problémem se stala různorodost pohledu dosavadních publikací na vývojové etapy pravopisného systému a edičních zásad. Tento aspekt byl však vyřešen přidáním stránky umožňující zadat vlastní pravidla v nově vytvořené aplikaci „Automatická transkripce“ (<http://penkavova-jcu.wz.cz/>). I samotné zvýšení počtu pramenů, které by se staly při využívání automatizace přístupnými, by napomohlo k jednoznačnému ustálení zásad a k jednotnému pochopení vývojových změn.

Tato práce otevírá dveře k revolučnímu pokroku v oblasti automatické transliterace a transkripce historických dokumentů. Výsledky viditelně dokazují, že s trochou dalšího výzkumu a investic lze dosáhnout i téměř bezchybné a rychlé automatizace přepisů, která zrychlí a zefektivní práci historiků, filologů a dalších odborníků a tím umožní zpřístupnit data i široké veřejnosti, pro níž jsou historické prameny z důvodu vědomostí nepřístupné.

Zdroje

Prameny

Polička, Městské muzeum a galerie Polička, Sign. K 378, fol. V-VI. Svatováclavská bible, Starý zákon, 1. část (Genesis – Sírachovec). Dostupné z:

https://www.manuscriptorium.com/apps/index.php?direct=record&pid=AIPDIG-MP18__K_378_____2Y5WNU6-cs.

Vienna, Österreichische Nationalbibliothek, Cod. 1175, fol. 199v-200. Biblia bohemica, saec. XV. Dostupné z: <http://www.onb.ac.at/sammlungen/hschrift/bibliographie.htm>.

Literatura

Akademický slovník cizích slov: A-Ž. Edited by Věra Petráčková – Jiří Kraus. Dotisk 1. vyd. Praha: Academia, 2000. s. 834.

DAŇHELKA, Jiří. Obecné zásady ediční a poučení o češtině 15. století. In: Výbor z české literatury doby husitské. Svazek první. Praha, Nakladatelství Československé akademie věd 1963, s. 31–41.

DAŇHELKA, Jiří. Směrnice pro vydávání starších českých textů. In: Husitský Tábor 8, 1985, s. 285–301.

FIDLEROVÁ, Alena A., Robert DITTMANN, František MARTÍNEK a Kateřina VOLEKOVÁ. Dějiny češtiny. 2013.

GEBAUER, Jan. Příruční mluvnice jazyka českého pro učitele a studium soukromé, 3. vyd., upravil F. Trávníček, 1925, s. 444–446.

HLAVSA, Zdeněk. Pravidla českého pravopisu: s dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky. Praha: Academia, 1993.

JIROUŠKOVÁ, Lenka. Edice versus transkripce aneb zapomenutá ediční praxe? Listy filologické, 2007, roč. 130, č. 3-4, s. 373-380. ISSN 0024-4457.

KOMÁREK, Miroslav, BLÁHA Ondřej (ed.). Dějiny českého jazyka. Brno: Host, 2012.

KOSEK, Pavel. Historická mluvnice češtiny I. Brno: Masarykova univerzita, 2014.

KYAS, Vladimír. Česká bible v dějinách národního písemnictví. Praha: Vyšehrad, 1997.

MAREŠOVÁ, Hana. Základy historické mluvnice češtiny: s texty k rozboru. Olomouc: Univerzita Palackého v Olomouci, 2008.

PORÁK, Jaroslav. Humanistická čeština: hláskosloví a pravopis. Acta Universitatis Carolinae. Praha: Univerzita Karlova, 1983.

ŠLOSAR, Dušan. Spisovný jazyk v dějinách české společnosti. Vyd. 3., opr. a dopl. Brno: Host, 2009.

ŠŤOVÍČEK, Ivan. Zásady vydávání novověkých historických pramenů z období od počátku 16. století do současnosti: příprava vědeckých edic dokumentů ze 16.-20. století pro potřeby historiografie. Praha: Archivní správa Ministerstva vnitra ČR, 2002.

TOMIŠKA, Ondřej. Automatizovaný přepis rukopisných historických dokumentů a jejich využití pomocí moderních IT [online]. Hradec Králové, 2021 [cit. 2024-02-17]. Dostupné z: <https://theses.cz/id/9lbwy8/>. Diplomová práce. Univerzita Hradec Králové, Filozofická fakulta. Vedoucí práce doc. RNDr. Štěpán Hubálovský, Ph.D.

VINTR, Josef. Zásady transkripce českých textů z barokní doby. Listy filologické 121, 1998, s. 341–346.

VOIT, Petr. Encyklopedie knihy: starší knihtisk a příbuzné obory mezi polovinou 15. a počátkem 19. století. Bibliotheca Strahoviensis. Praha: Libri ve spolupráci s Královskou kanonií premonstrátů na Strahově, 2006.

VYKYPĚLOVÁ, Tat'ána. Wege zum Neutschechischen. Studien zur Geschichte der tschechischen Schriftsprache. Hamburg: Verlag Dr. Kovač. 2013.

ZAVŘELOVÁ, Alžbeta. Projekt PERO – OCR pro historické texty. Duha: Informace o knihách a knihovnách z Moravy. Brno, 2020, roč. 34, č. 4.

Elektronické zdroje a aplikace:

Clementine Vulgate Project: The full text of the Clementine Vulgate, freely available online [online]. 2002-2005. Aktualizace dat 22. 10. 2022 [cit. 2024-03-08]. Dostupné z: <https://vulsearch.sourceforge.net/>.

ČERNÁ, Alena M., Boris LEHEČKA. Metodika přípravy a zpracování elektronických edic starších českých textů [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka 2015, aktualizace 11. 2. 2016 [cit. 2024-04-01]. Dostupné z: https://vokabular.ujc.cas.cz/soubory/nastroje/Methodics/Methodika_pripravy_a_zpracovani_elektronickych_edic_DF12P01OVV028.pdf.

DAMI: Nás IT slovník [online]. [cit. 2024-02-20]. Dostupné z: <https://www.damidev.com/slovník/desktopova-aplikace>.

Diabile [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka 2023. Aktualizace dat 19. 12. 2022 [cit. 2024-04-01]. Dostupné z: <http://diabile.com>.

Diptychon [online]. [cit. 2024-02-19]. Dostupné z: <https://diptychon.informatik.uni-bremen.de/html/index.html>.

eScriptorium [online]. [cit. 2024-02-19]. Dostupné z: <https://www.sofer.info/>.

Internetová jazyková příručka: Transkripce a transliterace z cyrilských abeced – obecné poučení [online]. [cit. 2024-2-23]. Dostupné z: <https://prirucka.ujc.cas.cz/?id=920>.

KBLAB [online]. 2014 [cit. 2024-02-20]. Dostupné z: <https://lab.kb.nl/about-us/blog/entangled-histories-ocr-htr-atr-automatic-text-recognition>.

Microsoft: Studentské Trenéřské Centrum [online]. [cit. 2024-02-20]. Dostupné z: <https://studuj.digital/2020/12/09/umela-inteligence-6-dil/>.

Michalcová, A. (2022) „Padeřov-Bible-handwriting-ground-truth: Initial release“. Zenodo. doi: 10.5281/zenodo.7467034. [cit. 2024-03-31]. Dostupné z: <https://zenodo.org/records/7467034>.

Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/>.

Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/CYRILICE>.

Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/LATINKA>.

Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/TRANSKRIPCE>.

Nový encyklopedický slovník [online]. [cit. 2024-02-23]. Dostupné z: <https://www.czechency.org/slovník/TRANSLITERACE>.

Nový encyklopedický slovník: Bratrský pravopis [online]. [cit. 2024-03-06]. Dostupné z: https://www.czechency.org/slovník/BRATRSKÝ_PRAVOPIS.

- Nový encyklopedický slovník: Diakritický pravopis [online]. [cit. 2024-03-05]. Dostupné z: https://www.czechency.org/slovník/DIAKRITICKÝ_PRAVOPIS.
- Nový encyklopedický slovník: Obrozenecký pravopis [online]. [cit. 2024-03-07]. Dostupné z: https://www.czechency.org/slovník/OBROZENSKÝ_PRAVOPIS.
- https://www.czechency.org/slovník/PERIODIZACE_VÝVOJE_ČEŠTINY.
- Nový encyklopedický slovník: Pračeština [online]. [cit. 2024-03-02]. Dostupné z: <https://www.czechency.org/slovník/PRAČEŠTINA>.
- Nový encyklopedický slovník: Primitivní pravopis [online]. [cit. 2024-03-04]. Dostupné z: https://www.czechency.org/slovník/PRIMITIVNÍ_PRAVOPIS.
- Nový encyklopedický slovník: Spřežkový pravopis [online]. [cit. 2024-03-05]. Dostupné z: https://www.czechency.org/slovník/SPŘEŽKOVÝ_PRAVOPIS.
- OCR4all [online]. [cit. 2024-02-19]. Dostupné z: <https://www.ocr4all.org/>.
- Projekt PERO [online]. [cit. 2024-02-19]. Dostupné z: <https://pero.fit.vutbr.cz/>.
- Projekt PERO: O projektu [online]. [cit. 2024-02-19]. Dostupné z: <https://pero.fit.vutbr.cz/about>.
- Österreichische Akademie der Wissenschaften: Discovering the power of AI for reading medieval manuscripts: HTR Winter school 2022 [online]. [cit. 2024-04-01]. Dostupné z: <https://www.oeaw.ac.at/imafo/read/discovering-the-power-of-ai-for-reading-medieval-manuscripts-htr-winter-school-2022>.
- Rescribe [online]. [cit. 2024-02-19]. Dostupné z: <https://rescribe.xyz/>.
- Rescribe: Download [online]. [cit. 2024-02-23]. Dostupné z: <https://rescribe.xyz/rescribe/>.
- RICHTEROVÁ, Alena. Alfabetička. In: KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online]. Praha: Národní knihovna ČR, 2003- [cit. 2024-02-23]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000002275&local_base=KTD.
- The Bible: e-Bible [online]. [cit. 2024-04-01]. Dostupné z: <https://ebible.com/>.
- Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-19]. Dostupné z: <https://www.transkribus.org/>.
- Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://readcoop.eu/transkribus/public-models/>.
- Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-21]. Dostupné z: https://readcoop.eu/model/bastarda_1460-1463_enhg_v2/.
- Transkribus [online]. READ-COOP SCE, 2019 [cit. 2024-02-21]. Dostupné z: https://readcoop.eu/model/kurrent_1515_enhg_v3/.
- Transkribus: Beta [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné z: <https://beta.transkribus.eu/learn/learndesk?type=none&status=none>.

Transkribus: Ground Truth [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://readcoop.eu/glossary/ground-truth/>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/docscan>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/uploading-files-to-transkribus>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/baselines-models>.

Transkribus: Help Center [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/model-setup-and-training>.

Transkribus: OCR vs HTR [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://readcoop.eu/insights/ocr-vs-htr/>.

Transkribus: ScanTent [online]. READ-COOP SCE, 2019 [cit. 2024-02-20]. Dostupné
z: <https://readcoop.eu/scantent/>.

Transkribus: Smart Search [online]. READ-COOP SCE, 2019 [cit. 2024-02-20].
Dostupné z: <https://help.transkribus.org/smart-search>.

Vokabulář webový: Webové hnízdo pramenů k poznání historické češtiny [online]. Praha:
Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka. © 2006–2020. Verze dat
1.1.15 [cit. 2024-04-01]. Dostupné z: <https://vokabular.ujc.cas.cz>.

Přílohy

Seznam obrázků:

- Obrázek 1 – ScanTent nacházející se v Akademické knihovně Jihočeské univerzity.
- Obrázek 2 – Smart Search.
- Obrázek 3 – Transkribus Learn.
- Obrázek 4 – Transkribus eXpert (desktopová aplikace).
- Obrázek 6 – Transkribus internetové prostředí.
- Obrázek 5 – Import document(s) – nahrání souboru.
- Obrázek 7 – Layout Analysis – Automatické rozpoznání řádků v aplikaci Transkribus eXpert.
- Obrázek 8 – Layout v internetovém rozhraní.
- Obrázek 9 – Text Recognition – automatické rozpoznávání textu, úvodní tabulka ještě před samotným výběrem modelu a dalších podrobností.
- Obrázek 11 – Text Recognition – tabulka s podrobnostmi a nastavením.
- Obrázek 10 – Text Recognition – internetové rozhraní.
- Obrázek 12 – Pokročilé nastavení zobrazování regionů, řádků apod.
- Obrázek 13 – Nastavení v internetovém prostředí.
- Obrázek 14 – Přidání speciálních označení – Tags – do textu.
- Obrázek 15 – Přidání speciálních označení – Tags – do textu v Transkribus eXpert.
- Obrázek 16 – Konvoluční neuronová síť.
- Obrázek 17 – Barevné označení jednotlivých stran dokumentu.
- Obrázek 18 – Vytváření vlastního modelu v Transkribus eXpert.
- Obrázek 19 – Zobrazení možnosti výběru při tvoření nového modelu v internetovém prostředí.
- Obrázek 20 – Export – stáhnutí dokumentu.
- Obrázek 21 – Zobrazení výstupních dat ve formátu DOCX.
- Obrázek 29 – Abeceda bratrského pravopisu vytvořený Mistrem Janem Husem.
- Obrázek 23 – Stránka s předvyplněnými pravidly dle druhů pravopisů.
- Obrázek 24 – Stránka s předvyplněnými pravidly dle Směrnice od Jiřího Daňhelky.
- Obrázek 25 – Stránka s předvyplněnými fonologickými pravidly.
- Obrázek 26 – Stránka sloužící k zadávání vlastních transkripčních pravidel.
- Obrázek 27 – Vypsání pravidla jednotlivých stránek.
- Obrázek 28 – Rozkliknutí a odkaz na původní tvar slova u ztučněného a podtrženého slova.
- Obrázek 29 – Abeceda bratrského pravopisu vytvořený Mistrem Janem Husem.
- Obrázek 30 – Transkribovaný model pomocí zkratk v aplikaci Transkribus.
- Obrázek 31 – Transkribovaný model v aplikaci Transkribus.
- Obrázek 32 – Model v aplikaci Transkribus (automatická transkripce).
- Obrázek 33 – Výsledek automatické transkripce vytvořený pomocí modelu v aplikaci Transkribus.
- Obrázek 34 – Vlastní transkripční pravidla použita na transliterovaný text Bible padeřovské.
- Obrázek 35 – Vlastní transkripční pravidla použita na transliterovaný text Bible svatováclavské.

Seznam tabulek:

Tabulka 1 – Primitivní pravopis.

Tabulka 2 – Starší spřežkový pravopis.

Tabulka 3 – Mladší spřežkový pravopis.

Tabulka 4 – Přehled některých grafémů (alografů) bratrského (tiskařského) pravopisu.

Tabulka 5 – Komparace fonologických změn.

Tabulka 6 – J. Daňhelka: Přehled problematických znaků v jednotlivých systémech pravopisu.

A	b	c	č ^{čé}	d	d ^{dje}	e	f	g	h	i	j	y	k	l	í	m	n	ň
o	p	q	r ^{er}	ř ^{er}	f	í ^{eš}	s	s ^{eš}	t	ť ^{tie}	v	v ^{vé}	w	x	z	ž ^{žet}	ch ^{ché}	d
đ	r	r	r ^ř	r ^{er}	s	š ^{eš}	t	y										
a	b	c	č	d	d	e	f	g	h	i	j	y	k	l	í	m	n	ň
A	bude	celé	čeledi	dáno	dědictvie	ej,	farář,	jenž	hospodin									
i	y	k	l	l	m	n	ň	o	p									
ili	i	král	lidi,	lákán	mnoho,	nás	někdy	on	pokoj,									
r	ř	f	f	t	ť	v	ve	w	x									
rád	řádem	slúžil	šlechtný,	tak	tělesný	ukázav	velikost	v sobě	ksil									
z	ž	ch																
za ny	životem	chtě,	jenž jest konec	i počátek požehnaný na věky.														

Obrázek 29 – Abeceda bratrského pravopisu vytvořený Mistrem Janem Husem.

Maltař:

Ohvalia budu pana poyšie ga
 si widycky chwala geho w v
 stech mych. Wpam chwalena
 bude dusie slyšte tisy a wesele
 budu. Welebece pana seymu a powšime
 gmena geho wespolek. Hledal sem pana
 a wšyřal me ge. a zewšech zamuckow
 mych wyswobodil me ge. Wstupte k
 memu a osiwette sie. a twari wasie nebu
 du pohameny. Tento chudy wofal gest.
 a pan wšyřal ge ger. a zewšech zamuckow
 geho spasil ge ho. Wypustie andiel pame
 wokol bogiacz sie geho. a zprostie ge. O
 kusie a wize zet ge chutny pan. Blazeny
 mur ktery vsa weni. Dorte sie pana wšie
 kni swieci geho. negoc neme nedostatku
 bogiem sie geho. Bohati zewšestihua zlac
 melisu. ale ktery sie bogie pana. nebudu
 vmeusem wewsem dobre. Podte syno
 we poslučajte mne. bami pamenauant
 was. ktery ge sloueti gesto chce zwotu
 mluge dny vidietu dobre. Zbani yariku
 twenu otzeho. a rtowe twogi nemiduge
 sta. Dtwari sie otzeho a cim dobre. Hle
 day pokoge a nasledug geho. Oci pame
 nasprawedliwe. a wšy geho k wyřba gich.
 Ale oblicy pame naty ktery cime zlee
 wia. aby zatratil zenne pamer gich.
 Wolali su sprawedliwy. a pan wšyřal gest
 ge. a zewšech zamuckow gich wyswobodil
 ge. Hlisko ge pan tien ktery gsu zamuce
 nego sda. a pokone duchem spasi. Mo
 zu zamuckowe sprawedliwy. a zewšech
 tich wyswobodil ge pan. Ostrcha pan
 wšech kosty gich. qedna mch nebu po
 tiena. Smer hieřnych uayhorie. a ktery
 su nenanideli sprawedliwego shtie.
 Wytupi pan dusie slyš slyřba neštie sie
 wšickm ktery vsagy wineho. **Chwala du**
 nd pane. **indom wškonam. xxxiii**
 cy ktery s kodie mne. wyřoug
 cy ktery bogie stogie proti mne.
 Popadim odieme a stie. a pow
 stan napomoc mne. Wytahm meq. a o
 przi proti mny. ktery sie protime mne.
 rza dusi me spaseme ya gsem. Pohame
 ni bude a zastyde sie ktery sledagij dusie
 me. Odwincem bude zpiet a pohamem
 ktery wyslece omnie zlee wia. Bude ya
 ko pirač před wari vnetu. a andiel pame
 yndie ge. Bud a esta gich tny aplikost
 a andiel pame sřahage ge. Neb su darmo
 sřerli zabubu ofyda siveho. nadar mot su
 hameli dusi mny. Wad genu ofydo geho
 neme. a geie ktery ge sřerl popadim ger.
 a w ofydo wpadim wyř. Ale dusie ma udy
 wari sie bude wpam. a kochati sie bude z

spaseme siveho. Wšicky kosti me diegi
 pane kco ge towen tobie. Geni wyřahar
 nedostatecneho zruky sil nesyřch ge. po
 trebného a chudeho odiech ktery layagi
 geho. Wstawie swieci neprawy ktery
 sem wia nerieciel tazachu mne. Wdpla
 cowaču mi zlee wia zadobce yalowost du
 sie me. Ale ya kdyz mntiechu mne obšane
 ch sie wim. Pomozowach wyřostu dusie
 me. modřeba ma wšono me sie obřati. Jak
 to bšimeho. yakito bntia najeřo tak obř
 bowach. yakozto sřagia a zamuceny tak
 sie pomozowach. A ywta mne wšestihua sie
 a festsu sie. sebiam gsu namie bigore a
 nerieciel sem. Rozperlem gsu am su mie
 li zleme. pok uřieřisu mne posimewali
 su sie mne posimewanim. sřerli su na
 mne zuby sřyřim. Pane kdy wzeřis na
 wriac dusi mny od zřisřtwa gich obřow
 gedmku mny. Wymawati tie budu wšo
 tu welikem. wřidu tiekem chwala tie bu
 du. Zedugte sie nadenim ktery sie pro
 time mne neprawie. ktery su mne nena
 uideli darmo. a mndagj oćama. Neb za
 giste mne pokorme mšumecu. a wšime
 wzenne mšumecu sři mšestech. Wřst
 řili su namie vřta swa. rekli su ha ha. ha ha.
 videli su oći nati. Wdielei hořpodne ne
 mly. pane neochazeg oćenne. Wřtan
 a piri křidu mny boje moy. a pane
 moy křy me. Sud mne podle sprawedl
 nosti twe pane boje moy. at sie nendugj
 nadenim. At nediegy wřidach sřyřba
 ha. ha ha. dusi nati. am. at diegy porzeme
 ger. Zastyde sie. a zahanbiece sie spo
 ktery sie tadugj řim viceem mny. Ob
 sećeni bude pohamem a studem ktery
 zlořne wia mšume namie. Ndadugte sie
 a wseře sie ktery chce sprawedlnost me
 a rziete. wlebebi bud pan. ktery chiegy
 pokogi slyřy geho. A yarik moy pćeny
 řeti bude sprawedlnost twy. a cly den řy
 řa tu. **Ja wřestie slyřy bořeho čwa**
Ekł. ja dauidowa. řalim xxxv
Rgt ne sprawedliwy. aby sřieřil
 řam w řobie neme bami boje
 před oćama geho. Neb řtme ge
 ćiml před obřiegem geho. aby nateřena
 byla neprawost geho k nenanisti. Slown
 řst geho neprawost a leř. nečuel ge řto
 řumeti aby dobre ćiml. Neprawost ge
 mšřil naloř sřem. řal ge přikazde ae
 řtie nedobře. a zloř ge nennel w nenan
 řa. Pane wnebi mřofidensie twe. a
 prawda twa az do obřakow. Sprawedl
 wost twa yak o hory boje. sudowe twogj
 propařt mnoha. řidi yřowadu spasię pa

.Žaltarz.

hwaliti budu pana powfie cza-
fi wždycky chwala gehu w v-
ftech mych.W panu chwalena
bude duſie.flyfte tifi j a wefeli

budte.Welebte pana femnu.a powyfme
gmena gehu wefpolec.Hledalfem pana
a wflyfal mie gt.a zewfiech zamutkow
mych wyfwbodil mie gt.Prifupte k-
niemu a ofwiette fie.a twafi wafie nebu-
du pohanieni.Tento chudy wolal gefť.
a pan wflyfal gt.gey.a zewfiech zamutkow
geho fpafil gt.ho.W puťtie andiel panie
w okol bogicich fie gehu.a zproťfi gie.O-
kufte avizte zet gt chutny pan.Blažený
muž kterýž wfa wen.Boyte fie pana wfic-
kni fwietí gehu.nebot nenie nedoftatku
bogicím fie gehu.Bohaťi zepfielifu a zlač-
nielifu.ale kteriž fie bogie pana.nebudu
vmenſeni wewfem dobrém.Podte fyno-
wé pofluchayte mne.bazni panie naučímť
was.Ktery gt elouiek gefťo chce žiwotu.
miluge dny widieti dobré.Zbran yaziku
twenu otzleho.a rtowé twogí nemilugte
lfti.Otwrat fie otzleho a čin dobree.hle-
day pokoge a nafledug gehu.Oczy panie
nafprawedliwé.a wfi gehu k probā gich.
Ale obliceg panie naty kteriž činie zlee
wieci.aby zatratil z zemie pamiet gich.
Wolalifu fprawedliwij.a pan wflyfal gefť
gie.a zewfiech zamutkow gich wyfwbodil
gie.Blizko gt pan tiem kteriž gfu zamuce-
neho frdcie.a pokorne duchem fpafi.Mno-
zij zamutkowe fprawedliwych.a zewfiech
tiech wyfwbodij gie pan.Ofťrieħa pan
wfiech koťij gich.gedna znich nebude po-
ťrieħa.Smrt hrieřnych nayhorſie.a kteri
fu nenauidieli fprawedliwego ſhrieřie.
Wykupí pan duſie fluh fwych.a neſhrieřie
wfickni kteriž wfagij wnieho. Chwala da

ud pane. uidowa wfkonani.xxxiiij.
ty kterižť ſkodie mnie.wyboiug
ty kterižť bogie ſťogie proti mnie.
Popadni odienie.aťit.a pow-

ťtan napomoc mnie.Wyrhni mecz ao-
przi proti ním.kterižť fie protiue mnie.
rzi duſi me ſpaſenie ya gfem.Pohanie-
ni budte.a zaťyde fie.kterižť hledagij duſie
me.Odwraceni budte zpiet a pohanieni.
kterižť myflee omnie zlee wieci.Budte ya-
ko prach předtwari vietru.a andiel panie
pudie gie.Bud cieřta gich tmy a plzkofť.
a andiel panie ſtihage gie.Neb fu darmo
ſkryli zahubu ofydla fweho.nadarmot fu
hanieli duſi mu.Přid gemu ofidlo gehožť
neuſe.a getie kterežť gt ſkryl popadni gey.
a w ofidlo wpadni wnye.Ale duſie ma rado-
wati fie bude wpanu.a kochati fie bude z

ſpaſenie fweho.Wfiecky koťi mé diegi.
pane kto gt rowen tobie.Genž wytrħas
nedoftateczneho zruky filnieyfich ge po-
ťřebneho a chudeho odtiech kterižť lapagi
geho.Wftawſe ſwiedci neprawij kterych
fem wieci newiediel tazachu mne.Odpla-
cowachu mi zlee wieci zadobře.yalowofť du-
ſie me.Ale ya když mutiechu mie.oblacie-
ch fie wźini.Ponizowach wpoftu duſie
mé.modlitba ma wlono meſie obrati.Yakž-
to bliźnieho.yakžťo bratra nařeho tak obli-
bowach.yakožťo lkagici a zamuceny tak
fie pomzowach.A proti mnie wefelilifue.
a ſellifu fie.febrani gfu namie biczowé a
newiedielſem. Rozptylení gfú aniž fu mie-
li žielenie.pokufielifu mne poſmiewali
fu fie mnie poſmiewaním.ſkripielifu na-
mie zuby fwymi.Pane kdy wzezřiš na-
wrat duſi mu odzloſinſtwa gich odlwow
gedínku mú.Wyznawati tiebudu wfbo-
ru welikem.wludu tiežkem chwalti tiebu-
du.Neradugte fie nademnú kterižť fie pro-
tiue mnie nepraue.kterižť fu mne nena-
uidieli darmo.a mrdagij očima.Neb za-
gíté mnie pokoynie mluuiechu.a w hnie-
wu zemie mluuiece lfti myfſlechu.A rozfi-
řilifu namie wfa fwa.reklifu ha ha.haha.-
vidielifu oczí naſi.Vidielſi hořpodine ne
mlcz.pane neotchazeg otemne.Wftan
a prizri k fudu memu bože moy.a pane
moy kprime.Sud mie podle fprawedl-
noſti twe pane bože moy.at fie neradugij
nademnu.At nediegij w frdcich fwych.ha-
ha.haha duſi naſi.aniz at diegij pozreme
gey.Zaťyde fie.a zahanbiete fie ſpolu.
kterižť fie radugij zlym wiecem mym.Ob-
lečeni budte pohaniením a ſtudem kterižť
zloſťne wieci mluuie namie.Radugtefie
a wefelte fie kterižť chie ſprawednoſti me.
a rzcieite zweleben bud pan.kterižť chtiegij
pokogi fluhy gehu.A yazik moy premy-
ſleti bude ſprawednoſť twu.ciely den chwa-
lu twu. Za witezſťue fluhy božieho chwa
Ekl. la dauidowa.zalm xxxv.
gt neſprawedliwy.aby ſhrieřil

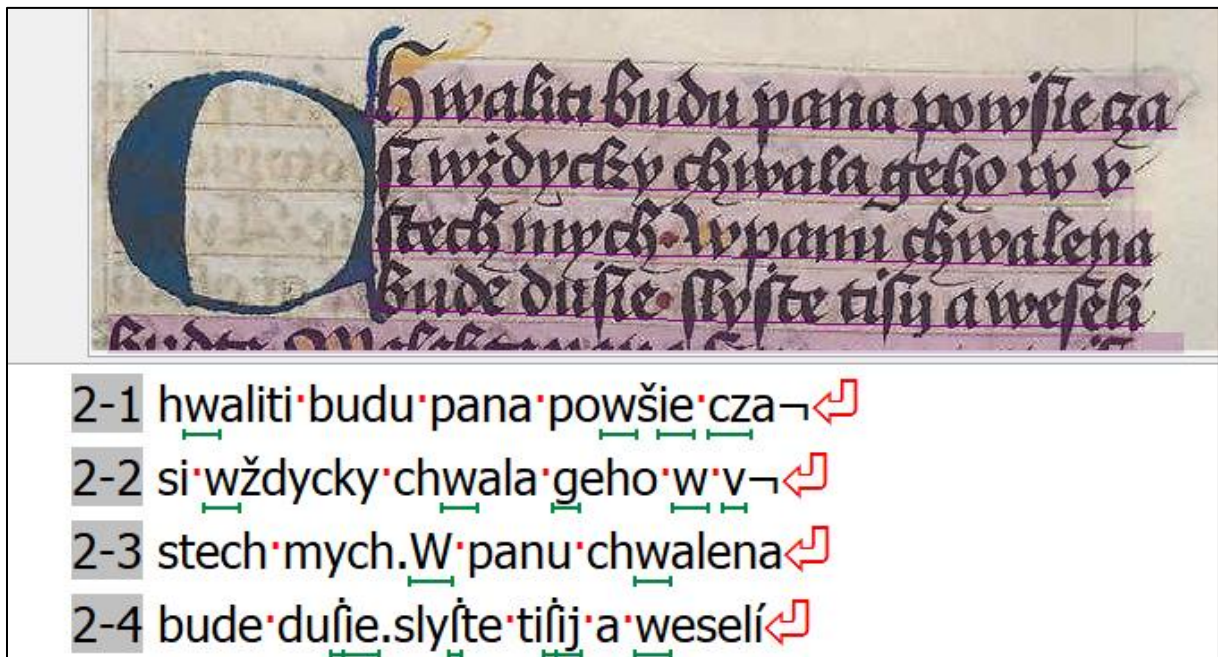
fam w fobie.nenie bazni bozie
před očima gehu.Neb lftiue gt
činil před oblicegem gehu.aby nalezena
byla neprawofť gehu k nenauiſti.Slowa
wft gehu neprawofť a leſť.nechtiel gt fro-
zumieti aby dobre činil.Neprawofť gt
myſlił naloži ſwem.ftal gt prikažde cie-
ťtie nedobré.a zloſťi gt nemiel w nenaui-
ťti.Pane wnebi milofrdenſťue twe.a
prawda twa až do oblakow.Sprawedli-
woſť twa yako hory bozie.fudowe twogi
popaťt mnoha.Lidi yhowada ſpařiš pa-

.Žaltarz.

hwaliti budu pana powŕie czaŕi wŕdycky chwala gehu w vftech mych. W panu chwalena bude duŕie. ŕlyŕte tiŕij a wefelí budte. Welebte pana ŕemnu. a powyŕme gmena gehu weŕpolek. Hledaŕŕem pana a vŕlyŕal mie ġt. a zewŕiech zamutkow mych wyŕwobodil mie ġt. Pŕiŕtupte kniemu a ofwiette ŕie. a twari waŕie nebudu pohanieny. Tento chudy wolal geŕt. a pan vŕlyŕal ġt gey. a zewŕiech zamutko^w gehu ŕpafil ġt ho. W puŕtie andiel panie w okol bogicích ŕie gehu. a zproŕtí gie. Okulte avizte ŕet ġt chutny pan. Blaŕeny muž kterýž vfa wen. Boyte ŕie pana wŕickni ŕwietí gehu. nebot nenie nedoŕtatu bogicím ŕie gehu. Bohatí zepŕielifu a zlaćnielifu. ale kteríž ŕie bogie pana. nebudu vmenŕeni wewŕem dobrém. Podte ŕynowé poŕluchayte mne. bazni panie naućimt was. Ktery ġt ćlouiek geŕto chce ŕivotu. miluge dny vidieti dobré. Zbran yaziku twenu otzleho. a rtowé twogi nemilugte lŕti. Otwrat ŕie otzleho a ćzin dobree. hleday pokoge a nalledug gehu. Oćzi panie naŕprawedliwé. a vŕlí gehu k proŕbā gich. Ale obliceg panie naty kteríž ćzinie zlee wiece. aby zatratil z zemie pamiet gich. Wolalifu ŕprawedliwij. a pan vŕlyŕal geŕt gie. a zewŕiech zamutkow gich wyŕwobodil gie. Blizko ġt pan tiem kterzí gfu zamucenoho ŕrdcie. a pokorne duchem ŕpafi. Mnozij zamutkowe ŕprawedliwych. a zewŕiech tiech wyŕwobodij gie pan. Oŕŕieha pan wŕiech koŕtij gich. gedna z nich nebudie poŕŕiena. Smrt hŕieŕnych nayhorŕie. a kterí ŕu nenauidielí ŕprawedliwego hŕieŕŕie. Wykupí pan duŕie ŕluh ŕwych. a neŕhŕieŕie wŕickni kteríž vŕagij wnieto. Chwala da ud pane. uidowa wŕkonaní. xxxiiij. ty kterížt ŕkodie mnie. wyboiug ty kterížt bogie ŕtrogie proti mnie. Popadni odienie. aŕŕit. a powŕtan napomoc mnie. Wytrhni mećz aoprzi proti ním. kterížt ŕie protiue mnie. rzci duŕie me ŕpaŕenie ya gŕem. Pohanieni budte. a zaŕtydte ŕie. kterížt hledagij duŕie me. Odwraceni budte zpiet a pohanieni. kterížt myŕlee omnie zlee wiece. Budte yako prach pŕedtwari wietru. a andiel panie pudie gie. Bud cieŕta gich tmy a plzkoŕt. a andiel panie ŕtihage gie. Neb ŕu darmo ŕkryli zahubu ofyďla ŕwego. nadarmot ŕu hanieli duŕie mu. Pŕid gemu ofidlo gehoŕt neuie. a getie ktereŕt ġt ŕkryl popadni gey. a w ofidlo wpadni wnye. Ale duŕie ma radowati ŕie bude wpanu. a kochati ŕie bude z

ŕpaŕenie ŕwego. Wŕiecky koŕti mé diegí. pane kto ġt rowen tobie. Genŕ wytrhaŕ nedoŕtatećznego zruky ŕilnieyŕich ge^o poŕŕebneho a chudeho odtiech kterížt lapagi gehu. Wŕtawŕe ŕwiedci neprawij kterych ŕem wiecei nevidiel tazachu mne. Odplacowachu mi zlee wiecei zadobré. yalowoŕt duŕie me. Ale ya kdyŕ mutiechu mie. oblaćiech ŕie wŕzini. Poniŕowach wpoŕtu duŕie mé. modlitba ma wlono meŕie obrati. Yakoŕto bliŕnieho. yakoŕto bratra naŕeho tak obli bowach. yakoŕto lkagící a zamuceny tak ŕie pomŕowach. A proti mnie wefelililufie. a ŕellifu ŕie. ŕebraní gfu namie bicŕowé a nevidielŕem. Rozpytlení gŕí aniŕ ŕu mieli ŕielenie. pokuŕielifu mne poŕmiewali ŕu ŕie mnie poŕmiewaním. ŕkripielifu namie zuby ŕwymí. Pane kdy wzeŕŕiŕ nawrat duŕie mu odzloŕŕŕtwa gich odlwow gedínku mú. Wyznawati tiebudu wŕboru welikem. wliđu tieŕkem hwaliti tiebudu. Neradugte ŕie nademnú kterížt ŕie protiue mnie nepraue. kterížt ŕu mne nenauidieli darmo. a mrdagij oćima. Neb zagíŕté mnie pokoyne mluuiechu. a w hniewu zemie mluuieće lŕti myŕlechu. A rozŕŕirilifu namie vŕta ŕwa. reklifu ha ha. haha. vidielifu oćzi naŕí. Vidielŕi hoŕpodine ne mlć. pane neotchazeg otemne. Wŕtan a pŕiŕzi k ŕudu memu boŕe moy. a pane moy kprime. Sud mie podle ŕprawednoŕti tve pane boŕe moy. at ŕie neradugij nademnu. At nediegij w ŕrdcích ŕwych. haha. haha duŕie naŕí. aniŕ at diegij poŕieme gey. Zaŕtydte ŕie. a zahanbiete ŕie ŕpolu. kterížt ŕie radugij zlym wiecem mym. Oblećeni budte pohaniením a ŕtudem kterížt zloŕtne wiecei mluuie namie. Radugteŕie a weŕte ŕie kterížt ćhtie ŕprawednoŕti me. a rzciete zweleben bud pan. kterížt ćhtiegij pokogi ŕluhy gehu. A yazik moy pŕemyŕleti bude ŕprawednoŕt twu. ciely den chwalu twu. Za wítieŕŕtuie ŕluhy boŕiego chwa Ekl. la dauidowa. ŕalm xxxv. ġt neŕprawedliwy. aby ŕhŕieŕil ŕam w ŕobie. nenie bazní boŕie pŕed oćima gehu. Neb lŕtíue ġt. ćzínil pŕed oblicegem gehu. aby nalezena byla neprawoŕt gehu k nenauiditi. Slowa vŕt gehu neprawoŕt a leŕt. nechtiel ġt frozumieti aby dobře ćzínil. Neprawoŕt ġt myŕlil naloŕi ŕwem. ŕtal ġt pŕikaŕde cieŕtie nedobré. a zloŕti ġt nemiel w nenauiditi. Pane wnebi miloŕŕdenŕtuie tve. a prawda twa aŕ do oblakow. ŕprawedliwoŕt twa yako hory boŕie. ŕudowe twogi propaŕt mnoha. Lidi yhowada ŕpafiŕ pa

Transliterovaný text Bible padeřovské ve formátu Word – výstup z aplikace Transkribus.



Obrázek 30 – Transkribovaný model pomocí zkratek v aplikaci Transkribus.

.Žaltař.

hvaliti budu pana povše ča-
si vždycky chvala jeho v u-
stech mych.V panu chvalena
bude duše.slyšte tíši a veselí

buďte.Velebte pana semnu.a povyšme
jmena jeho vespolek.Hledal sem pana
a uslyšal me jest.a zevšech zamutkov
mych vysvobodil me jest.Přistupte k-
nemu a osvíte se.a tvaři vase nebu-
du pohaneny.Tento chudy volal jest.
a pan uslyšal jest jej.a ze všech zamutkov
jeho spasil jest ho.V pušte andel pane
v okol bojících se jeho.a zproští je.O-
kuste a vizte žet jest chutny pan.Blažený
muž kterýž vsa ven.Bojte se pana všic-
kni sveti jeho.nebot není nedostatku
bojícím se jeho.Bohatí zepselisu a zlač-
nelisu.ale kteříž se bojí pana.nebudu
uměnění vevsem dobrém.Podte syno-
vé posluchajte mne.bazni pane naučím
vas.Ktery jest člověk ješto chce životu.
miluje dny videti dobré.Zbran jaziku
tvemu otzleho.a rtové tvoji nemiluje
lsti.Otvrat se otzleho a čin dobré.hle-
daj pokoje a nasleduj jeho.Oči pane
naspravedlivé.a vši jeho k prosbam jich.
Ale obličej pane na ty kteříž činí zlé
veci.abý zatratil z zeme pamet jich.
Volalisu spravedliví.a pan vslyšal jest
je.a zevšech zamutkov jich vysvobodil
je.Blizko jest pan tem kteříž jsu zamuce-
neho srdce.a pokorne duchem spasi.Mno-
zí zamutkově spravedlivých.a zevšech
tech vysvobodí je pan.Ostříha pan
všech kostí jich.jedna z nich nebude po-
třína.Smrt hřešných najhorše.a kteří
su nenavideli spravedliveho shřeší.
Vykupí pan duše sluh svých.a neshřeší
vsickni kteříž vsají vneho.Chvala da-

Sud pane.vidova vskonaní.xxxiif.

ty kteříž škode mne.vyboiuj
ty kteříž boje strojí proti mne.

Popadni odení.aštit.a pov-

stan napomoc mne.Vytrhni meč ao-
při proti ním.kteříž se protive mne.
řci duši me spasení ja jsem.Pohaní-
ni buďte.a zastýdte se.kteříž hledají duše
me.Odvraceni buďte zpeta pohaneni.
kteříž myslé omne zlé veci.Buďte ja-
ko prach před tvaři vetru.a andel pane
pude je.Bud cesta jich tmy a plzkost.
a andel pane stihaje je.Neb su darmo
skryli zahubu osydlu sveho.nadarmot su
haneli duši mu.Přid jemu osidlo jehožt
neví.a jete kterežt jest skryl popadni jej.
a v osidlo vpadni vnye.Ale duše ma rado-
vati se bude vpanu.a kochati se bude z

spasení sveho.Všecky kosti mé dejí.
pane kto jest roven tobe.Jenž vytrhaš
nedostatečného zruky silnějších jeho po-
třebného a chudeho od tech kteříž lapají
jeho.Vstavše svedci nepraví kterých
sem vecí nevedel tazachu mne.Odpla-
covachu mi zlé veci zadobré.jalovost du-
še me.Ale ja když mutíchu me.oblace-
ch se vžíní.Ponižovach vpostu duše
mé.modlitba ma vlono mese obrati.Jakž-
to bližního.jakžto bratra našeho tak obli-
bovach.jakožto lkající a zamuceny tak
se pomžovach.A proti mne veselilisuse.
a sešlisu se.sebraní jsu name bičové a
nevedelsem. Rozptýlení jsú aníž su me-
li želení.pokušelisu mne posmívali
su se mne posmíváním.škřipelisu na-
me zuby svými.Pane kdy vzezříš na-
vrat duši mu odzlosinstva jich odlvov
jedínku mú.Vyznavati teбудu vsbo-
ru velikem.vlídú težkem chvaliti tebu-
du.Neradujte se nademnú kteříž se pro-
tíve mne neprave.kteříž su mne nena-
videli darmo.a mrdají očima.Neb za-
jisté mne pokojne mluvíchu.a v hne-
vu zeme mluvíce lsti myslechu.A rozší-
řilisu name vsta sva.reklisu ha ha.haha.-
videlisu oči naši.Vidísi hospodine ne
mlč.pane neotchazej otemne.Vstan
a příží k sudu memu bože moj.a pane
moj kpríme.Sud me podle spravedl-
nosti tve pane bože moj.at se neradují
nademnu.At nedejí v srdcích svých.ha-
ha.haha duši naší.aníž at dejí pozřeme
jej.Zastýdte se.a zahanbete se spolu.
kteříž se radují zlym vecem mym.Ob-
lečení buďte pohaním a studem kteříž
zlostne veci mluve name.Radujtese
a veselte se kteříž chte spravednosti me.
a řcete zveleben bud pan.kteříž chtejí
pokoji sluhý jeho.A jazik moj přemy-
šleti bude spravednost tvu.cely den chva-
lu tvu.Za vítězství sluhý božeho chva-
Ekl.la davidova.žalm xxxv.
jest nespravedlivy.abý shříšil

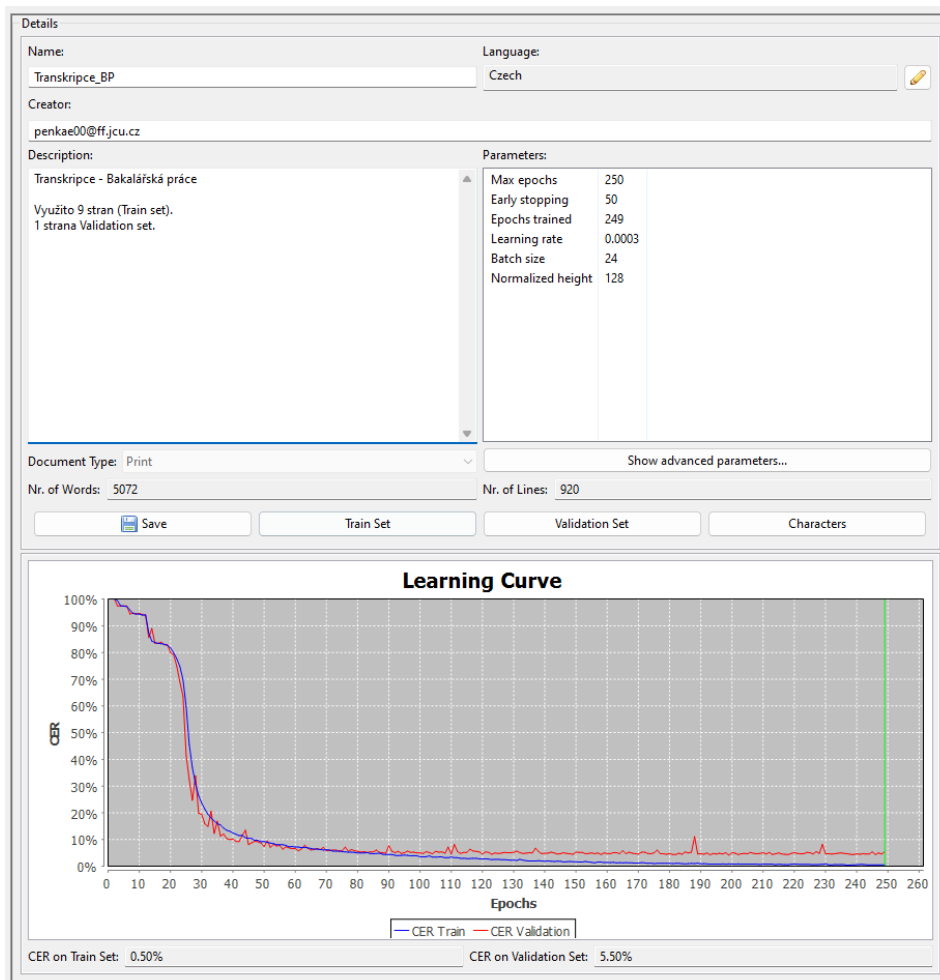
sam v sobe.není bazní bože

před očima jeho.Neb lstíve jest.

činil před obličejem jeho.abý nalezena
byla nepravost jeho k nenavisti.Slova
vst jeho nepravost a lest.nechtel jest sro-
zumeti aby dobře činil.Nepravost jest
myslil naloží svem.stal jest příkažde ce-
ste nedobré.a zlostí jest nemel v navi-
sti.Pane vnebi milosrdenství tve.a
pravda tva až do oblakov.Spravedli-
vost tva jako hory boží.sudove tvoji
propast mnoha.Lidi yhovada spasiš pa-

Obrázek 31 – Transkribovaný model v aplikaci Transkribus.

Model zjednodušené transkripce Bible padeřovské



Obrázek 32 – Model v aplikaci Transkribus (automatická transkripce), který ukazuje, že i z malého počtu stran lze vytvořit funkční model a který potvrzuje možnost využití aplikace i pro transkripci.

si navysost vzalsi vazbu.vzalsi dary nali di.Zajiste y neveficii.abi přebyvali vpa-
nu bohu.Požehnany pan v den nakaždy den.štrastn cestu v činí nam.boh spasení našich.Bohnaš boh aby včimíl spaseny.a pane pane vychod smrti.Ale však boh ze-
tře hlavy nepřatel svych.vrch vlasu tech ještě chode v hřešech svych.Rzekl jest pan zbazan.obratím obratím v hlubokost morskú.Aby omocena byla noha tva vekt-
vi.jazik psov tvych znepřatel tvych ot-
nyeho.Vidílisu v chazení tva bože.vcha-
zeme boha meho.krale meho jenž jest v sva-
tem.Předešlisu knížata spojem jsut chva-
litebníkom.prostřed mladíc bubnujících.
Vsbořích chvalte boha.pana zstudnic izra-
helských.Tu bemaní mladenec.vmysli vystupení.Knížata vida vodce jich.knížata zabulon.knížata neptalm.Přikaž bože moji prve patve toho ko.čak v delal vnas.Od chramu tveho v jeruzaleme.
tobe obetovati budú kralove dary.Talaj zveří trestne.sebrane by kov mezi krava-
mi lidskými.abi vyvrhli ty kteříž pokušení jsu střebřím.Rozptylej pohany kteří chtějí boje.přijdut poslove zejipta.muřemnska zeme předejde rukama svyama k bohu.K ralovství zeme zpevajt bohu.chvalu vzdajte panu.Chvalu vzdajte bohu jež jest vstupil nanebe nebeš.kvychu slunce.
Aj dat hlasu svemu hlas moci.dajte chva-
lu bohu nad izrahelem.velebnost je a moc-
jeho v obláčích.Divny boh v svatých svy-
ch boh izrahelsky.ont da mot a sílu lidu svev požehnany boh.Vskonání zaty kteříž bu-
du prominem.Počma se zalm.lxvii.

spasena me včí-
bože.nebt su ve-
sly vody až do du-
se me.Vvezl
sem v hliní hlu-
bokosti.a není-
podstaty.Při-
šelsm v hlubo-
kost morskú.a
buře potopila

jest me.Pracovalsem volaje.zm lklysu
dasní mé.sesly oči moji.když vsam v bo-
ha meho.Rozmohlit su se nad vlasí hlavy-
me.kteříž nenavidelisu mne darmo.Po-
silnenit jsu kteříž su se protivili mne ne-
přetelé moji bezpravne.kterých sem ve-
ti nelapal tehdy vyplačovach.Bože ty
veš nemudrost mu.a hříchové moji ob-
tebe nejsou skryti.Nestyďte se nademnu.
kteříž čekají tebe pane.pane moci.Ne
budťe pohaním nalezyimi kteříž hledají

tebe bože izrahelsky.Neb sem prote trpel pohanení.přikryla jest hanba tvař mu.
Eizí včinen jsem bratřím mym.a put-
ník synom mateře mé.Neb horlivost do-
tveho smedla me jest.a ruhaní tech kte-
řisu te haneli.spadla su name.Y přikryl
sem v postu duši mu.y včineno jest v posme-
ch mne.A položilsem rucho me zim.a včí-
men jsem jim vpřislove.Proti mne mlu-
víchu kteříž sedichu vbrane.a omne
zpevachu kteříž pijíchu víno.Ale ja med-
litbu mu ktobe pane.čas dobrě líbosti bo-
že.vmnožství milosrdenstve tve vslyš
me.vpravde spasení tveho.Vvytrh
ni me z blata a t pych neuvazl.vysvobod-
me odtích kteříž nenavide mne.a zhlú-
bokostí vod.Nepohřížuj mne buře vo-
dy.aníž mne požeraj hlubokost.aníž za-
veraj nademnu propast vst svych.Vslyš
me pane neb dobro tve jest milosrdenství
tve.vedle množství smilovaní tvych v-
zeží name.Aneotvracuj tvař tve otdé-
tete tveho.neb se mutím.rychle vslyš me.
Přizí kduši me a vysvobod ji.pronepřa-
tely me zprost me.Ty veš hanení me.a
hanbu mu.a stud moj.Před obličejt tvym.
jsu všickni kteříž mutí me.hanení ča-
kalo jest srdce me a psoty.Y čekalsem ktoby-
se spolu semnu zamutí.a nebyl jest.a ktoby
vtešil a nenalezlsem.Y dalisu mi zapokrmí
zluč.a vžezni me napajelisu me otem.
Bud stol jich předními v osydo a v odpla-
ty a v pohoršení.Zatmeny budte oči jich
aby nevideli.a chřbet jich vždicky sklon.
Vyli nane hnev tvoj.a prchaní smevu-
tveho popadni je.Bud přebytek jich
pusty.a v stanovištích jich nebud ktoby byd
lil.Neb kohož si ty zbil protivilisu se.a k-
bolesti ran mych přidalisu.Přilož nepra-
vost knepravostí jich.anevychodte v spra-
vedlnost tvu.Shlazení budte s kníh živý-
ch.a s spravdivými nebudťe napsani.
.Žat jsem chudy a bolestny.spasení tve bo-
že přijalo jest me.Chvaliti budu jmeno
bože s zpevaním.a velebiti budu jeho v-
chvale.slibi se bohu nad telce mladeho.
jemuž se rohove vymítají a kopyta.
Vizte chudí a veselte se.hledajte pana.a
živa bude duše vaše.Neb jest vslyšal cha-
de pan.a okovanými svými nepohřel jest
Chvalte jeho nebesa.y zeme.more.y vši-
ni hyzove vnich.Neb boh spasenu včí-
ní syon.a vzdelana budu mesta vidova.
V budu bydlení tam.a dedictvými dosahnu
jeho.A seme sluh jeho vladnutí bude ji.
a ktož milují jmeno jeho přebyvatí bu-
du vním.Vskonání chvala davidova

Obrázek 33 – Výsledek automatické transkripce vytvořený pomocí modelu v aplikaci Transkribus.

Transkribovaný model (z aplikace „Automatická transkripce“) Bible padeřovské

Přidat pravidlo
Deaktivovat pravidlo
Zobrazit pravidla
Aplikovat pravidla

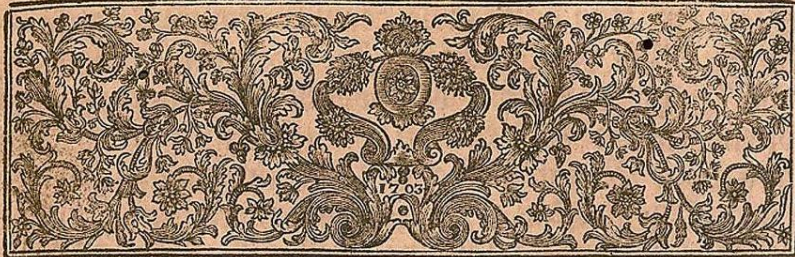
```

Z -> Ž
g -> ž
W -> V
W -> V
rz -> ř
Rz -> Ř
f -> š
(ale|lo|u|y -> $1)
(A|E|I|O|U|y -> $1)
r -> ř
z -> ž
f -> š
l -> l
g -> j
cz -> č
c -> č
W(y) -> j
š -> š
ie -> i
ie -> e/ě/í
(oj)(?l|w) -> új
cz -> č
š -> š
    
```

Transkribovaný text:

Zlatě, hvaliti budu pana povše/ě/í časi vždycky chvala jeho v vstech mych. V panu chvalena bude duše/ě/í, slyšte tiši a veselí budte. Velebte pana semnu a povyšme jména jeho vespolek. Hledalsm pana a vslyšal me/ě/í jest a zevše/ě/ích zamutkov mych vysvobodil me/ě/í jest. Přistupte kne/ě/ímu a osve/ě/íte se/ě/í a tvaři vase/ě/í nebudu pohane/ě/íny. Tento chudy volá jest, a pan vslyšal jest jej a zevše/ě/ích zamutkov jeho spasil jest ho. V pušte/ě/í ande/ě/í pane/ě/í v okol bojících se/ě/í jeho a zprostj je/ě/í. Okuste avizte žeť jest chutny pan. Blaženy muž kterýž vta ven. Bojte se/ě/í pana všickni sve/ě/íti jeho nebot není nedostatku bojícím se/ě/í jeho. Bohatí zepse/ě/ílisu a zlačne/ě/ílisu ale kteříž se/ě/í bůjji pana nebudu vmenšeni vevsem dobrém. Podte synové postuchajte mne. Bazni pane/ě/í naučim vas. Kteryž jest čloue/ě/ík ješto chce životu miluje dny vide/ě/íti dobré. Zbran jaziku tvenu otzeho a rtové tvoji nemiluje lsti. Otvrat se/ě/í otzeho a čin dobre hledaj pokoje a nasleduj jeho. Oči pane/ě/í naspravdivé a všji jeho k prosbá jich. Ale obličej pane/ě/í naty kteříž činí zlee ve/ě/íci aby zatratil z zeme/ě/í pame/ě/íti jich. Volalisu spravdiví a pan vslyšal jest je/ě/í a zevše/ě/ích zamutkov jich vysvobodil je/ě/í. Blizko jest pan te/ě/im kteříž jsu zamuceného srdce/ě/í a pokorne duchem spasi. Mnozí zamutkove spravdivých a zevše/ě/ích te/ě/ích vysvobodil je/ě/í pan. Ostřihá pan vše/ě/ích kostí jich, jedná z nich nebudu potřina. Smrt hře/ě/ísných najhorše/ě/í a kteříž su nenaude/ě/íli spravdivého shře/ě/ísi. Vykupi pan duše/ě/í sluh svých a neshře/ě/ísi vsickni kteříž vřaji vne/ě/ího. Chvala da ud pane. Uidova vskonaní xxxiii. ty kteříž škode/ě/í mne/ě/í vyboiuj ty kteříž boje/ě/í strůjji proti mne/ě/í. Popadni ode/ě/íni ašit a povstan napomoc mne/ě/í. Vytrhni meč aopř proti nim. kteříž se/ě/í protue/ě/í mne/ě/í. fci dusi me spasení ja jsem Pohaníni budte a zastydte se/ě/í kteříž hledaji duše/ě/í me. Odvraceni budte zpe/ě/íti a pohane/ě/íni. kteříž myslee omne/ě/í zlee ve/ě/íci. Budte jako prach předtvaři ve/ě/ítru a ande/ě/í pane/ě/í pude/ě/í je/ě/í. Bud ce/ě/ísta jich tmy a plzkost. a ande/ě/í pane/ě/í stihaje je/ě/í. Neb su darmo skryli zahabu osyda svého nadamot su hane/ě/íli duši mu. Přid jemu osidlo jehožti neui a je/ě/í kteříž jest skryl popadni jej, a v osidlo vpadni vnye. Ale duše/ě/í ma radovati se/ě/í bude vpanu a kochati se/ě/í bude z spasení svého. Vse/ě/ícky kosti mé de/ě/íji. pane kto jest roven tobe/ě/í. Genž vytrháš nedostatečného zrucky siine/ě/íjich je* potřebného a chudeho odtě/ě/ích kteříž lapaji jeho. Vstavše sve/ě/ídcí nepraví kterých sem ve/ě/íci neve/ě/íde/ě/íli tazachu mne. Odplacovachu mi zlee ve/ě/íci zádobře jalovost duše/ě/í me. Ale já když mutlichu me/ě/í oblace/ě/ích se/ě/í vřini. Ponižovach vpostu duše/ě/í mé modlitba ma viono mese/ě/í obrati. Jakžto blížního jakžto bratra našeho tak obli bovach. jakožto ikajici a zamuceny tak se/ě/í pomůzovach. A proti mne/ě/í veselilisu se/ě/í, a sešlisu se/ě/í sebrani jsu name/ě/í bičové a neve/ě/íde/ě/í íšem. Rozpřtyleni jsu aníž su me/ě/íli že/ě/ítení pokuse/ě/ílisu mne posmívali su se/ě/í mne/ě/í posmívání skřipe/ě/ílisu name/ě/í zuby svymi. Pane kdy vzeztíš navrat duši mu odziosinstva jich odlov jedniku mu. Vyznavati te/ě/íbudu vsboru velikem vřidu te/ě/ížkem chvaliti te/ě/íbudu. Neradujte se/ě/í nademnú kteříž se/ě/í protue/ě/í mne/ě/í nepraue/ě/í kteříž su mne nenaude/ě/íli darmo a mrdaji očima. Neb zajisté mne/ě/í i pokojne/ě/í mluuichu a v hne/ě/ívu zeme/ě/í mluuice lsti myslechu. A rozšřilisu name/ě/í vsta sva reklisu ha ha haha vide/ě/ílisu oči naši. Vidili hospodine ne mlč. pane neotchazej oteme. Vstan a přizvi k sudu memu bože můj a pane můj křime. Sud me/ě/í podle spravdivosti tve pane bože můj at se/ě/í neraduj nademnu. At nede/ě/íji v srdcích svých haha haha duši naši aníž at de/ě/íji pozřeme jej. Zastydte se/ě/í a zahanbe/ě/íte se/ě/í spolu. kteříž se/ě/í raduj zlym ve/ě/ícem mym. Oblečeni budte pohaním a studem kteříž zlostne ve/ě/íci mluue/ě/í name/ě/í. Radujtese/ě/í a veselte se/ě/í kteříž chte/ě/í spravdivosti me. a řce/ě/íte zveleben bud pan. kteříž chte/ě/íjji pokoji sluh jeho. A jazik můj přemýšleti bude spravdivost tvu. ce/ě/íly den chvalu tvu. Za vitizstui sluh bože/ě/ího chva Ekl. Ia dauidova. Zalm xxxv. jest nespravdivy aby shřřil sam v sobe/ě/í. není bazni bože/ě/í před očima jeho. Neb lstitue/ě/í jest. činil před obličejem jeho aby nalezena byla nepravost jeho k nenausti. Slova vst jeho nepravost a lest. nechte/ě/íli jest srozume/ě/íti aby dobrě/ě/í činil. Nepravost jest myslii naloži svem stal jest příkažde ce/ě/íste/ě/í nedobré a zlosti jest neme/ě/íli v nenausti. Pane vnebi milosrdenstui tve a pravda tva až do oblakov. Spravdivost tva jako hory boží sudove tvoji propast mnoha. Lidi jhovada spasiš pa

Obrázek 34 – Vlastní transkripční pravidla použita na transliterovaný text Bible padeřovské.



Přepis
Nedůstognějšimu a Wysoce
Oswjcenému Knížeti

Pánu/Pánu

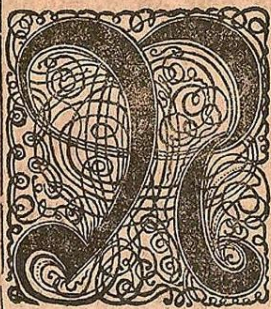
M A T A U S F E R D Y N A N D

z Boží Milosti/a Stolice Aposstolité

Arcy-Biskupu Pražskému/

Swatě Římské Říše Knížeti/a Hrabi z Rhin-
burku/Geho Chf. a Král. Mil. Stut. Tegné Radě/Pánu
na Mladý Wozicy/Semberku/Městicy/a Šebětowě zc. Stolice Swatě Apo-
stolské Římské Rozenému Leháru/a w Králowstwj Českém Prymasu/tež Král:
Univerf: Carolo-Ferdinandez Praž: stálemu a vřtawicnému Kancelji.

Pánu a Pánu neymilostiwějšimu.



Nedůstognějši a Wysoce Oswjcený Kníže/a
z Boží Milosti Pražský Arcy-Biskup Pán
Pán Matauš Ferdynand z Bilenberku/a
Jan Frydrych z Waldšteyna / dvě sta-
wné Paměti tohoto wzneseného Králowstwj
Českého někdy na oswjcený Cirkwe Swatě
Swjcen postawená a wyzdwižená neygasněg-
šji Swětla/a z Boží Milosti/tež y od Aposstol-

a

ské

Připis.

ské Stolice na trůn důstojnosti Arcy-Biskupské Pražské / jako
 v Římské Říše Knížetství welebně wywysšeni / a řádně přizeni
 Wrehnj Duchownj Pastýřowé / twogi / Nedyůstogněgšj a Wysoce
 Dswjcenj Knjze / a z Boží Milosti nyněgšj Pražskj Arcy-Biskupe /
 Pane a Pane Nymilostiwěgšj / (Titul) chwalitebnj Předlowé ;
 aby swau Apostolstau a Pastýřstau nevwraucněgšj horliwost
 w rozhogněj a wzdělanj prawé Křestanské Katolicé Římské
 a Samo - Spasytelné Wiry zgerwné Swětu na wěcnau Pa-
 mátku / a k zwelebenj Sláwy Boží / oznámili / a wsselité bludy
 Kachěské / a bludná Bčeni / (která až posawád wětšim dšlem / jako
 pod negakým gbelkem / se skrýwala / a strz Čestě rozličné naka-
 zliwé Knihy / a ty aneb w czých fraginách / aneb pokradmo kbe-
 šy wydané / sem y tam po Králowstwj Čestém / ano y w okol-
 njch fraginách / mezy sprostj lid obzwláště / roztrauffené / se
 zastjraka / a y pod dnes nebezpečně se zastjrags) s pomocj Nedy-
 wysššjho Pána / jakožto Otce wssěch prawých Swětěl / po ne-
 náhlu aučiml iwě vyplniti a wyfořeniti se wynasnažili / gal na
 swá Dcowská ustanoweni / tak y na nemalé wlastnj náklady
 Nowau Bibli Nowého y Starého Zákonu / w Čestém / a této
 nassj wlasti přirozeným gazylu / wydati milostiwé sau našj diti
 ráčili. Tak gal toho Nařizenj Knížetškj Dekret (kterj po-
 loženj se nacházy před Nowým Zákonem) Jana Frydrycha
 z Waldsteyna (Titul) wewš gmenowaného někdegššjho Arcy-
 Biskupa w Lětu 1676. Měsíce Listopadu / dne 9. wklazuge.
 Záčátek pak toho po wssěcky časy schwáleného a spasytelné-
 ho Ustanoweni / obau těch slawné Paměti / a nymilostiwěgššjch
 nassjch někdy duchownjch Wrehnostj / stal se gest od Nowého
 Zákonu / genž poneyprw swětlo widěl w Lětu. 1677.

Nynj / kdjž y také Starý Zákon / po tolika letech / ne s menššj prá-
 ch / ale stegnau / a s tau též pilnostj y snažnostj jako Nowý Zákon /
 podle Latinského od Cyrkwe Swaté potwřzeného textu (gme-
 nem Editio vulgata) gal neywlastněgi / a negčistěgi / s Miko-
 štj Boží / spořádan / složen / a zprawen gest / k Wassj Dswjcené Knj-

zet.

Přjpis

Neydůftogněgffimu a Wyfoce

Ofwjcenému Knjžeti

Pánu / Panu

FERDINANDOWI

Z Božj Milofti / a Stolice Apofftolfké

Arcy→Bifkupu Pražfkému /

Swate Ržjmfke Ržiffe Knjžeti / a Hraběti z Khůn→

burku / Geho Cýf. a Král. Mil. Skut. Tegné Radě / Pánu

na Mlady Wožicy Sfemberku / Měfficy / a Sfeběrowě sc. Stolice Swaté Apo =

fftolfké Řjmfke Rozenému Légátu / a w Králowftwj Čefkém Prymafu / též Král:

Univerf: Carolo-Ferdinandeæ Praž: ftálému a vftawičnému Kancljři.

Pánu a Pánu neymiloftiwěgffjmu.

Eydůftogněgffj a Wyfoce Ofwjcený Knjže / a
z Božj Milofti Pražfký Arcy→Bifkup Pán
Pan Matauš Ferdynand z Bilenberku / a
Jan Frydrych z Waldffteyna / dvě fla→
wné Paměti tohoto wzneffeného Králowftwj
Cžefkého někdy na ofwjcený Cýrkwe Swaté
Swjcen poftawená a wyzdwižená neygafněg→
ffi Swětla / a z Božj Milofti / též y od Apofftol→

Přjpis.

fké Stolice na trůn důftognofti Arcy-Bifkupfke Pražfké / gako y Ržjmfké Ržjffe Knjžetftwj welebně wywyffenj / g řádně zřjzenj Wrchnj Duchownj Paftýřowé / twogi / Neydůftognegffj a Wyfoce Ofwjcený Knjže / a z Božj Milofti nyněgffi Pražfký Arcy-Bifkupe / Pane a Pane Neymiloftiwěgffj / (Titul) chwalitebnj Předkowé ; aby fwau Apoftolfkau a Paftýřfkau neywraucněgffi horliwoft w rozhogněnj a vzdělánj prawé Křeftanfké Katolické Ržjmfké a Samo . Spafytedlné Wjry zgewně Swětu na wěchnau Pa-
mátku / a k zwelebenj Sláwy Božj / oznámili / a wffeliké bludy Kacýřfké / a bludná Včenj / (která až pofawád wětfjfm djlem / gako pod negakým gbeljkem/ fe fkrýwala / a fkrz Čzefké rozličné naka-
zliwé Knihy / a ty aneb w cyzých kraginach / aneb pokradmo kde-
fy wydané / fem y tam po Králowftwj Čzefkém / ano y w okol-
nich kraginách / mezy fproftý lid obzwláfftě / roztrauffené / fe zaftjrala / a y pod dnes nebezpečně fe zafkjragj) s pomocy Ney-
wyšffjho Pána / gakožto Otce wftech prawých Swětél / po ne-
náhlu aučinliwě wypleniti a wykořeniti fe wynafnažili / gak na fwá Otcowfká wftanowenj / tak y na nemalé wlaftnj náklady Nowau Biblj Nowého y Starého Zákona / w Čzefkým / a této naffj wlafti přirozeným gazyku / wydati miloftiwě fau nařjditi ráčili. Tak gak toho Nařzenj Knjžetfký Dekret (který po-
ložený fe nacházy před Nowým Zákonem) Jana Frydrycha z Waldfteyna (Titul) weyš gmenowaného někdegffiho Arcy-
Bifkupa w Létu 1676. Měfýce Liptopadu / dne 9. vkazuge. Začátek pak toho po wfcecky čafy schwáleného a fpafytedlné-
ho Wftanowenj / obau tech flawné Paměti / a neymiloftiwegffich naffich někdy duchownjch Wrchnoftj / ftal fe geft od Nowého Zákona / genž poneyprw fwětlo widěl w Létu. 1677. Nynj / když y také Starý Zákon / po tolika letech / ne s menffi prá-
cý / ale ftegnau / a s tau též pilnoftj y fnažnoftj gako Nowý Zákon / podlé Latinfkého od Cýrkwe Swaté potwrzeného textu (gmé-
nem Editio vulgata) gak neywlaftněgi / a negčiftegi / s Milo-
fti Božj / fpořádán / fložen / a zprawen geft / k Waffj Ofwjcené Knj-

Přjpis Neydůstogněgšlímú a Wyfoce Ofwjcenému Knjžeti Pánu / Panu FERDINANDOWI Z Božj Milofti / a Stolice Apofštolké Arcy-Bifkupu Pražškému / Swate Ržjmké Ržjše Knjžeti / a Hraběti z Khůnburku / Geho Cýf. a Král. Mil. Skut. Tegné Radě / Pánu na Mlady Wožicy Sfemberku / Měšicy / a Sfeběrowě sc. Stolice Swaté Apo štolké Řjmké Rozenému Léžátu / a w Králowštwj Čelkém Prymafu / též Král: Univerš: Carolo-Ferdinandeæ Praž: štalému a vřtawičnému Kancljři. Pánu a Pánu neymiloftiwěgšljmu. Eydůstogněgšlj a Wyfoce Ofwjcený Knjže / a z Božj Milofti Pražšký Arcy-Bifkup Pán Pan Matauš Ferdynand z Bilenberku / a Jan Frydrych z Waldšteyna / dvě šlawné Paměti tohoto wzněšeného Králowštwj Čzeškého někdy na ofwjcený Cýrkwe Swaté Swjcen pořtawená a wyzdwižená neygafněgšlj Swěta / a z Božj Milofti / též y od Apofštol-

Přjpis. šké Stolice na trůn důřtognofiti Arcy-Bifkupške Pražšké / gako y Ržjmké Ržjše Knjžetřtwj welebně wywyřšenj / g řádně zřjzenj Wrchnj Duchownj Pařtýřowé / twogi / Neydůstognegšlj a Wyfoce Ofwjcený Knjže / a z Božj Milofti nyněgšlj Pražšký Arcy-Bifkupe / Pane a Pane Neymiloftiwěgšlj / (Titul) chwalitebnj Předkowé ; aby řwau Apofštolškau a Pařtýřřkau neywraucněgšlj horliwořt w rozhogněnj a wzdělánj prawé Křeřtanšké Katolické Ržjmké a Samo . Špařytedlné Wjry zgewně Swětu na wěčnau Památku / a k zwelebenj Šlawy Božj / oznámili / a wřřeliké bludy Kacyřřké / a bludná Včenj / (která až pořawád wěřřljm djlem / gako pod negakým gbeljkem/ ře řkrýwala / a řkrz Čzešké rozličné nakazliwé Knihy / a ty aneb w cyřých kraginach / aneb pokradmo kdefy wydané / řem y tam po Králowštwj Čzelkém / ano y w okolnich kraginách / mezy řprořtý lid obzwlářřě / roztrauffené / ře zařřjala / a y pod dnes nebezpečně ře zalkjragj) s pomocy Neywyřřljho Pána / gakožto Otce wřřech prawých Swětel / po nenáhlu aučinliwě wypleniti a wykořeniti ře wynařnařžili / gak na řwá Otcowřká vřřtanowenj / tak y na nemalé wlařřnj náklady Nowau Biblj Nowého y Šarého Zákona / w Čzeškým / a této nařřj wlařři pořirozeným gazyku / wydati miloftiwě řau nařřditi ráčili. Tak gak toho Nařřzenj Knjžetřšký Dekret (který pořožený ře nacházy před Nowým Zákonom) Jana Frydrycha z Waldšteyna (Titul) weyř gmenowaného někdegšljho ArcyBifkupa w Létu 1676. Měřyce Liřtopadu / dne 9. vkazuge. Začátek pak toho po wřřecky čařy řchwáleného a řpařytedlného Vřřtanowenj / obau tech šlawné Paměti / a neymiloftiwegšljch nařřich někdy duchownjch Wrchnořřj / řtal ře geřř od Nowého Zákona / genž poneyprw řwěřřlo wiđěl w Létu. 1677. Nynj / když y také Šarý Zákou / po tolika letech / ne s menřři pořácy / ale řtegnau / a s řau též pořnořřj y řnařřnořřj gako Nowý Zákou / podle Latinského od Cýrkwe Swaté pořwřřzeného řextu (gménem Editio vulgata) gak neywlařřněgi / a negčifřegi / s Milofti Božj / řpořřádán / řložen / a zprawen geřř / k Wařřj Ofwjcené Knj-

Transliterovaný řext Bible řwatováclavské ve řormátu Word – výřřup z aplikace Transkribus.

Transkribovaný model (z aplikace „Automatická transkripce“) Bible svatováclavské

Přidat pravidloDeaktivovat pravidloZobrazit pravidlaAplikovat pravidla

j -> í
(a|e|i|o|u)y -> \$tj
f -> s
g -> j
w -> v
ss -> š
G -> J
RZ -> Ř
ü -> ú
au -> ou
s -> š
cý -> cí
(A|E|I|O|U)y -> \$tj
W -> V
cy -> ci
Cy -> Ci
Cý -> Cí
Ss -> Š
Cz -> Č
cz -> č
- -> -
ë -> ě
kí -> ký
b(ou) -> u
vulgata -> vulgata

Transkribovaný text:

Přepis Nejdůstojnějšimu a Vysoce Osvícenému Knižeti Pánu / Panu FERDINANDovi Z Boží Milosti / a Stolicé Apoštolské Arcibiskupu Pražskému / Svate Římské Říše Knižeti / a Hraběti z Khúnburku / Jeho Cis. a Král. Mil. Skut. Tejně Radě / Pánu na Mladý Vožici Šemberku / Měšici / a Šeběřové sc. Stolicé Svaté Apoštolské Římské Rozenému Légátu / a v Království Českém Prymasu / též Král: Univers: Carolo-Ferdinandae Praž: stálému a vstavičnému Kancilíři. Pánu a Pánu nejmilostivějšímu. Ejdůstojnější a Vysoce Osvícený Kniže / a z Boží Milosti Pražský Arcibiskup Pán Pan Matouš Ferdinand z Bilenberku / a Jan Frydrych z Valdštejna / dvě slavné Paměti tohoto vznešeného Království Českého někdy na osvěcený Církev Svaté Svícen postavená a vyzdvižená nejjasnější Světa / a z Boží Milosti / též y od Apoštol Připis. ské Stolicé na trůn důstojnosti Arcibiskupske Pražské / jako y Římské Říše Knižetství velebně vyvýšení / j řádně zřízení Vrchní Duchovní Pastýřové / tvoji / Nejdůstojnější a Vysoce Osvícený Kniže / a z Boží Milosti nynější Pražský Arcibiskupe / Pane a Pane Nejmilostivější / (Titul) chvalitební Předkové ; aby svou Apoštolskou a Pastýřskou nejrůdnější horlivost v rozhojnění a vzdělání pravé Křestanské Katolické Římské a Samo . Spasytedlné Víry zjevně Světu na věčnou Památku / a k zvelebení Slávy Boží / oznámili / a všeliké bludy Kacířské / a bludná Včení / (která až posavad větším dílem / jako pod nejakým jbelikem/ se skrývala / a skrz České rozličné nakazlivé Knihy / a ty aneb v cizích krajínách / aneb pokradmo kděsy vydané / sem y tam po Království Českém / ano y v okolních krajínách / mezy sprostý lid obzvláště / roztroušené / se zastírala / a y pod dnes nebezpečné se zaskýrají) s pomoci Nejvyššího Pána / jakožto Otce všech pravých Světél / po nenáhlou učinlivě vyplnění a vykoření se vynasnažili / jak na sva Otcovská vstanovení / tak y na nemalé vlastní náklady Novou Bibli Nového y Starého Zákona / v Českým / a této naší vlastní přirozeným jazyku / vydali milostivě sou naříditi ráčili. Tak jak toho Nařízení Knižetský Dekret (který položený se nachází před Novým Zákonem) Jana Frydrycha z Valdštejna (Titul) vejš jmenovaného někdejšího Arcibiskupa v Létu 1676. Měsíce Listopadu / dne 9. vkazuje. Začátek pak toho po všechny časy schváleného a spasytedlného Vstanovení / obou tech slavné Paměti / a nejmilostivějších našich někdy duchovních Vrchností / stal se jest od Nového Zákona / jenž ponejprv světo viděl v Létu. 1677. Nyní / když y také Starý Zákon / po tolika letech / ne s menší práci / ale stejnou / a s tou též pilností y snažností jako Nový Zákon / podlé Latinského od Církev Svaté potvrzeného textu (jmenem Editio vulgata) jak nejvlastnější / a nejčistější / s Milostí Boží / spořádan / složen / a zpraven jest / k Vaší Osvícené Kni

Obrázek 35 – Vlastní transkripční pravidla použita na transliterovaný text Bible svatováclavské.