roč. 67, 2020, č. 6

Vícejazyčnost v současné české poezii. Několik úvodních postřehů z korpusové perspektivy

Karel Piorecký – Michal Škrabal

PIORECKÝ, K. – ŠKRABAL, M.: Multilingualism in Contemporary Czech Poetry. A Few Introductory Observations from a Corpus Perspective
SLOVENSKÁ LITERATÚRA 67, 2020, No. 6, p. 568 — 583
DOI: https://doi.org/10.31577/slovlit.2020.67.6.4
Karel Piorecký – ORCID ID: 0000-0001-5727-289X
Michal Škrabal – ORCID ID: 0000-0001-9358-3361

Key words: multilingualism, xenisms, Corpus of contemporary [Czech] poetry, literary servers, quantitative-corpus approach

The paper is an attempt at a quantitative corpus related approach to the subject of multilingualism in contemporary Czech poetry (published both in books and on literary servers). The authors of the paper examine the frequency and distribution of foreign (i.e., non-Czech) lexical units, raising questions about the forms and functions of individual lexemes. Three selected poets (T. Kafka, M. Šanda, M. Torčík) are then analyzed more in-depth. The paper is also a report about a currently developed database – The Corpus of Contemporary Czech Poetry – and possibilities of using it. It suggests how beneficial the quantitative data analysis in the first phase of linguistically oriented literary research can be, pointing to the necessity of interconnecting the quantitative and qualitative approaches. It is only the researcher´s interpretative competence that can define the boundaries of the research field and the significance of its elements. When conducting text-centered analyses, language corpora should begin to play a role similar to other scientific infrastructure tools, such as bibliographic databases.

Kľúčové slová: viacjazyčnosť, xenizmy, Korpus súčasnej [českej] poézie, literárne servery, kvantitatívno-korpusový prístup

1. Záměry a východiska[1]

S pádem komunismu v listopadu 1989 a spolu s transformací české společnosti a kultury v duchu liberální demokracie vstoupila do české literární kultury dosud nebývalá mnohost a různost. Nad tímto faktem panuje odborná shoda (což dokládají například už názvy dvou rozsáhlých kompendií k literatuře polistopadové doby: *V souřadnicích volnosti*, 2008; *V souřadnicích mnohosti*, 2014).[2]

Mnohost a různost jsou dobře popsány pro oblast sociálního života soudobé literatury (bujení knižního trhu, nárůst počtu nakladatelství, časopisů, literárních akcí apod.). My se ovšem hodláme ptát, zda a jak tento živel různosti pronikal do samotných textů, resp. jejich jazyka, v našem případě jazyka básnických textů. Zde totiž literárněhistorický popis zatím absentuje. Rádi bychom proto tento popis doplnili a postavili ho na analýze reprezentativních dat, která jsou shromažďována v Korpusu současné poezie (dále KSP), a souběžně tuto datovou základnu také představili.

Za reprezentanta oné různosti jsme si zvolili různost jazykovou, tedy vícejazyčnost. Budeme sledovat četnost a distribuci jinojazyčných (tedy ne-českých) lexikálních jednotek ve zmíněném korpusu a ptát se po jejích formách a funkcích. Pojmové zázemí nám při tom poskytnou v první řadě práce Petra Mareše, zejména pak monografie *„Also: nazdar!" Aspekty textové vícejazyčnosti*.[3] Inspirativní je pro nás rovněž koncept xenismů Wolfganga Mosera,[4] neboť přesněji postihuje *záměrné* užívání jinojazyčných prvků, na rozdíl od prosté vícejazyčnosti, která nemusí se záměrností (neřkuli uměleckou záměrností) nutně souviset.

Dobře si uvědomujeme, že xenismy mohou být reprezentovány na různých jazykových rovinách: foneticky, graficky, slovotvorně, lexikálně, syntakticky. Naše práce se ovšem zaměří výhradně na xenismy lexikální. Důvodem k tomu je základní heuristická pozice, z níž naše práce vychází – dataset tvořený plnými texty 250 básnických sbírek současné české poezie opatřenými specifickým značkováním, které umožňuje kvantitativní výzkum xenismů právě na lexikální rovině.

Uvědomujeme si jistou reduktivnost tohoto přístupu, proto budeme kvantitativní metodu, resp. statistický přístup kombinovat s kvalitativní analýzou vybraných básnických sbírek, tedy sbírek, které statisticky vykazují největší frekvenci jinojazyčných lexikálních prvků.

Třetí – a možno říci doplňkovou – linií této analýzy bude dílčí sonda do vícejazyčnosti přítomné v textech amatérských autorů zveřejňovaných na internetových literárních fórech (dále používáme rozšířenější pojmenování „literární servery"). Tento subkorpus je kvantitativně neobyčejně bohatý, na druhou stranu

1 Při práci na publikaci byly využity zdroje výzkumných infrastruktur Česká literární bibliografie (http://clb.ucl.cas.cz) a Český národní korpus (LM2018137) financovaných Ministerstvem školství, mládeže a tělovýchovy ČR v rámci aktivity Projekty velkých infrastruktur pro VaVaI. Text vznikl s podporou z Akademické prémie udělené Akademií věd ČR prof. Pavlu Janouškovi. Za pomoc s datovou základnou děkujeme kol. Pavlu Procházkovi, za cenné podněty pak oběma recenzentům a také dr. Romanu Kandovi.

2 HRUŠKA, Petr – MACHALA, Lubomír – VODIČKA, Libor – ZIZLER, Jiří (eds.): *V souřadnicích volnosti: česká literatura devadesátých let dvacátého století v interpretacích*. Praha : Academia, 2008; FIALOVÁ, Alena (ed.): *V souřadnicích mnohosti: česká literatura první dekády jednadvacátého století v souvislostech a interpretacích*. Praha : Academia, 2014.

3 MAREŠ, Petr: *„Also: nazdar!" Aspekty textové vícejazyčnosti*. Praha : Karolinum, 2003. Srov. MAREŠ, Petr: *Nejen jazykem českým: studie o vícejazyčnosti v literatuře*. Praha : Filozofická fakulta Univerzity Karlovy, 2012.

4 MOSER, Wolfgang: *Xenismen. Die Nachahmung fremder Sprachen*. Peter Lang : Frankfurt am Main, 1996.

představuje souhrn mnohem hrubšího, syrovějšího výzkumného materiálu (manuální tagování zde vzhledem k rozsahu nepřipadá v úvahu). Proto se zde omezíme na ověření a zpřesnění jediné hypotézy, která se týká přítomnosti slovenských textů na českých internetových fórech, jejich četnosti a distribuce mezi jednotlivé weby. (Zde chceme demonstrovat, že KSP skýtá možnosti výzkumu jak *intratextuální*, tak *intertextuální* vícejazyčnosti, jak Georg Kremnitz nazývá situaci, kdy jeden autor – v našem případě ovšem jedno médium – používá v různých textech různé jazyky.)[5]

2. Metodologická rozprava: formy, funkce a význam xenismů v básnickém textu

P. Mareš ve své monografii dobře vystihuje stereotypní přístup k vícejazyčnosti v poezii: v básnických textech nejsou xenismy očekávány, protože báseň je tradičně vnímána jako sebevyjádření subjektu, který se přirozeně vyjadřuje mateřským jazykem a nemá důvod přepínat kód. Tato představa o subjektivitě v lyrice přežívá dodnes, byť textovou oporu má spíše v poezii 19. století – kdežto moderní, postmoderní, resp. současná poezie pracuje se subjektivitou již poněkud komplikovaněji a diferencovaněji. Homogennost konfesijního subjektivismu je postupně nahrazována subjektovým konstruktivismem – toto je, domníváme se, premisa, s níž je nutné pracovat při analýzách současné poezie, v opačném případě hrozí dezinterpretace, k níž dnes už zcela běžná jazyková a stylová heterogennost básnického textu může vést, pokud bude nadále zjednodušeně vnímán jako zpověď lyrického já.[6]

Dobře je to vidět už na příkladech, které uvádí P. Mareš. Ještě ve Šrámkově sbírce *Modrý a rudý* (1906) je patrná zřetelná distance subjektu a jinojazyčného lexika. Xenismy zde slouží jako vehikulum hodnotových postojů, které nejsou subjektem nijak problematizovány, ale prostě předávány. Jakmile se však jeho výklad přesune směrem k básnické postmoderně (Andrej Stankovič, Tomáš Kafka, Jan Křesadlo), musí konstatovat, že zde se vícejazyčnost uplatňuje „komplexněji a komplikovaněji", protože „bezprostředně vrůstá do slovesné ‚tkáně' poetického textu".[7] Lyrický subjekt v těchto případech si vůči xenismům nevytváří žádnou distanci, ba naopak, xenismy slouží k demonstraci jeho komplikované utvářenosti – a vedle toho mají (jak upozorňuje P. Mareš) funkci „výrazovou a kreativní" nebo též „funkci kulturní". My bychom spíš řekli, že xenismy jsou zde nositeli subjektového konstruktivismu (promlouvající „já" nedovolují recipientovi jednoduše přijmout jako hlas fikční postavy, ale nutí vnímat „já" jako konstrukt utvářený uměle, resp. s uměleckým záměrem). Výrazová a kreativní funkce jsou pak spíše důsledky, které má tento způsob stylizace subjektu na poetice básnického

5 KREMNITZ, Georg: *Mehrsprachigkeit in der Literatur. Ein kommunikationssoziologischer Überblick.* Wien : Praesens Verlag, 2015, s. 18.

6 Navazujeme zde na starší práci jednoho ze spoluautorů této studie, v níž – v polemice s pojetím lyrického subjektu u Miroslava Červenky – stanovil pozici lyrického subjektu ve významové struktuře díla jako fluidní: lyrický subjekt se pohybuje někde mezi krajními póly ztotožnění se subjektem díla (tj. básníkem, resp. fikčním konstruktérem textu) a vyloučením tohoto ztotožnění (tj. v případě akcentu na konfesijní ráz výpovědi). Intencionální vícejazyčnost potom vnímáme jako symptom inklinace k onomu „konstruktérskému" pólu, kdy už není sugerována jednota zpovídajícího se já, ale je naopak exponována jeho „udělanost", umělost, konstruovanost – a tedy proces stylizace subjektu má i svou sebereflexivní polohu. Srov. PIORECKÝ, Karel: Červenkova teorie lyrického subjektu. In: *Česká literatura*, roč. 54, 2006, č. 6, s. 30 – 55.

7 Mareš, „Also: nazdar!" Aspekty textové vícejazyčnosti, c. d., s. 200.

textu. „Kulturní funkce“ nepochybně do hry vstupuje, resp. kulturní významy vstupují do textů spolu s xenismy nevyhnutelně (těžko si představit situaci, kdy by tomu tak nebylo) – kulturní významy tedy vnímáme jako stabilní, inherentní sémantickou vrstvu xenismů, na kterou se vrství další funkce. Souhlasíme s P. Marešem, že „prožitek cizosti“ přítomný v dílech autorů, kteří prošli zkušeností pobytu v jinojazyčném prostředí, nemusí být z literárního, resp. interpretačního hlediska automaticky významný a zajímavý (v kvalitativní části naší studie se tímto hlediskem tudíž neřídíme). Nicméně Marešovy sondy do díla Ivana Blatného a Jiřího Gruši (tedy autorů disponujících zmíněnou zkušeností a pro Mareše představující „hloubkové sepětí vícejazyčnosti a prožitku cizosti“), které tvoří závěr jeho monografie, jsou mimořádně inspirativní i pro nás.

P. Mareš se v závěru své monografie ospravedlňuje, když píše, „že všechny výklady zde předložené se mohly dotknout jen některých aspektů vícejazyčnosti“.[8] Ani tato naše drobná studie pochopitelně nepostihne vícejazyčnost básnických textů komplexně, ale chtěli bychom alespoň ukázat, jak k tomuto cíli či ideálu může přispět kvantitativní korpusový přístup k materiálu v kombinaci s kvalitativní interpretací textů. Metodologicky se zde opíráme o rámec tvořený dnes již klasickými díly Franca Morettiho či Matthewa L. Jockerse (obě 2013).[9]

Marešova práce je velmi cenná svým systematickým přístupem k formám a funkcím vícejazyčnosti. Jistou její slabinou je ovšem oblast heuristiky. Příklady vícejazyčných textů jsou v jeho knize vybírány spíše nahodile, rozhodně tedy nepředstavují reprezentativní vzorek zkoumaného materiálu.

S podobnou nahodilostí bychom mohli připojovat další jména a díla. Ze současných autorů, u nichž lze hovořit (podobně jako u J. Gruši a I. Blatného) o „hloubkové“ vícejazyčnosti, můžeme ještě zmínit např. Wandu Heinrichovou, žilinskou rodačku (*1968), která ve své sbírce *Jehla sestupuje* užívá četné slovakismy (ale i germanismy a anglicismy) a do českojazyčného celku umisťuje text kompletně psaný slovensky (báseň *Vec*). V letos vydaném výboru z básnického díla Vikiho Shocka (*1975) je značná část jednoho oddílu (*Pátý šok: Ego sum poeta*) složena z bilingvních básní: zrcadlově (vlevo originál, vpravo překlad, případně volnější parafráze) tu defilují postupně verše latinské, starořecké, německé, anglické a italské. Vícejazyčnost je klíčová rovněž v díle již zesnulého básníka Miroslava Koryčana (1938 – 2016) *Kormidla (ne)paměti*, kde autor užívá lexikální výpůjčky z více než šesti desítek jazyků a čtenáři odhaluje jejich význam pomocí vysvětlivek na vnějším okraji

8 Tamtéž, s. 210.

9 MORETTI, Franco: *Distant reading*. London : Verso, 2013; JOCKERS, Matthew Lee: *Macroanalysis: digital methods and literary history*. Urbana : University of Illinois Press, 2013. Srov. též uplatnění tohoto přístupu v jiných současných českých literárněvědných pracích, např. ZMĚLÍK, Richard: *Kvantitativně-korpusová analýza a literární věda: model a realizace autorského korpusu a slovníku Jana Čepa v kontextu zahraniční a české autorské lexikografie*. Olomouc : Univerzita Palackého, 2015; ZMĚLÍK, Richard: *Konceptualizace barev v naratívní fikci na pozadí kvantitativních modelů*. Olomouc : Univerzita Palackého, 2019; PLECHÁČ, Petr – KOLÁR, Robert: *Kapitoly z korpusové versologie*. Praha : Akropolis, 2017.

strany. Těchto vysvětlivek je v celé knize o rozsahu 116 stran a 7 584 textových slov celkem 449,[10] tzn. autor opatřuje vysvětlující glosou v průměru každé 17. slovo.

3. Korpus současné poezie: od nahodilosti ke statistice

Výše uvedený výčet jmen by se jistě dal rozšířit o další básníky a básnířky, naše ilustrativní příklady vyplynuly do značné míry z nahodilosti[11] a mohly by být dost dobře nahrazeny jinými. Máme však za to, že jakkoliv je takováto nahodilost nedílnou součástí badatelova úsilí, je pro vědce neocenitelnou pomocí a úlevou, může-li potřebné informace snadno a rychle získat ze solidního a reprezentativního zdroje dat. Možnost pohotově se zorientovat v určité problematice, zjistit, zda je to které téma dostatečně nosné, a tudíž hodno jeho zájmu a úsilí, získat pomyslné předpolí pro svůj výzkum, jakýsi „inkubátor" vhodný pro formulaci svých úvodních hypotéz a výzkumných otázek – to vše nabízí **Korpus současné poezie**, projekt, na němž od roku 2015 spolupracují Ústav pro českou literaturu AV ČR, v. v. i., a Ústav Českého národního korpusu Filozofické fakulty Univerzity Karlovy.

Jedná se, jak už název napovídá, o korpus současných básnických textů české provenience vymezené lety 1990 a 2020, tj. reprezentativní vzorek tuzemského básnictví za poslední tři dekády. Co je podstatné, tento vzorek zahrnuje nejen texty oficiálně knižně vydané, a tedy i prošlé klasickým edičním procesem, ale i tvorbu amatérskou,[12] koncentrovanou především na tzv. literární servery.[13] Toto metodologické rozhodnutí není dáno snahou o jakousi demokratizaci poezie; věříme, že bez webových textů by obraz současné české poezie nebyl úplný, postihoval by pouze jeden její segment, navíc co do proporce dosti malý (viz níže tabulka 2).[14] To by neodpovídalo realitě (tudíž by šlo i o vědeckou nepoctivost), protože literární servery sehrály v českém literárním kontextu významnou roli,[15] mimo jiné jako platforma pro publikační začátky některých dnes již etablovaných básníků. Tato základní dichotomie nám mimochodem otevírá možnost

10 Pokud jde o výchozí jazyk přejímek, dominuje latina s rovnou stovkou případů, následuje angličtina (53), francouzština (41), španělština (34), řečtina s hebrejštinou (po 20), italština (17), němčina (15), arabština (8), hindština (7), ruština (6), slovenština s čínštinou (po 5) a další, méně obvyklé až raritní jazyky včetně indiánských (tunguzština, minangkabauština, ojibwština, jazyk Ongů) aj. V min. 38 případech není zdrojový jazyk uveden, často se totiž nejedná o přejímku, ale autor spíše parafrázuje slovo domácího původu.

11 Wandu Heinrichovou zná jeden z autorů z někdejšího společného působení v redakci literárního obtýdeníku *Tvar*, s Viki Shockem pojí téhož přátelství a výbor dostal přímo od autora. Koryčanova sbírka figuruje v interní databázi textů projektu Český národní korpus, z technických důvodů (dostupnost pouze ve formátu pdf, který není ideální pro korpusové zpracování dat) však dosud nebyla použita v žádném z korpusů synchronní češtiny z řady SYN (blíže https://wiki.korpus.cz/doku.php/cnk:syn).

12 Přívlastek amatérský tu používáme zcela neevaluativně, nevydělujíce proti amatérům pomyslné profesionály, „parnasisty" apod. Jde nám primárně o data samotná a je vcelku jedno, ke kterému z těchto základních dvou módů poezie patří.

13 Při vymezení této amatérské literární scény se opíráme zejména o tyto starší práce k tématu: NÁBĚLKOVÁ, Mira: Literárne servery a amatérsky česko-slovenský literárny kontext. In: SUWARA, Bogumiła – HUSÁROVÁ, Zuzana – ŠVELCH, Jaroslav – NÁBĚLKOVÁ, Mira: *V sieti strednej Európy: nielen o elektronickej literatúre*. Bratislava : SAP, 2012, s. 281 – 298; PIORECKÝ, Karel: Literary work as a leisure activity. Amateur literary forums on the Czech internet. In: CARNICELLI, Sandro – McGILLIVRAY, David – McPHERSON, Gayle (eds.): *Digital Leisure Cultures: Critical perspectives*. New York : Routledge, 2017, s. 193 – 206.

14 V budoucnu by se KSP měl rozrůst také o texty čerpané z dalších médií: časopisů, osobních blogů, sociálních sítí apod., čímž ovšem nastane nutnost řešit novou řadu technických i interpretačních otázek (např. problematiku duplicit, variant apod.).

15 PIORECKÝ, Karel: *Česká literatura a nová média*. Praha : Academia, 2016.

oba módy, v KSP rozlišované pomocí atributu *doc.medium* (print vs. web), vzájemně konfrontovat a srovnávat.

Začněme tedy od subkorpusu tištěné poezie: ten zahrnuje celkem 10 790[16] básní (strukturní atribut <*div2*>) z 256 básnických sbírek (<*doc*>) od 146 autorů (<*doc.author*>) a jeho celková velikost činí 944 346 textových slov (včetně interpunkce jde o 1 085 274 textových pozic). Básnické knihy byly postupně naskenovány, digitalizovány pomocí technologie OCR a převedeny do standardního formátu xml, prošly též kontrolou a ručním značkováním: bylo především zapotřebí označit začátky a konce jednotlivých básní či oddílů sbírky, ke sbírkám doplnit metadata aj. Tagset jsme s minimálními úpravami přejali ze staršího projektu Korpus českého verše.[17] Jeho součástí je i tag <*foreign*>, sloužící k označení cizojazyčného prvku či pasáže – tj. xenismu, jak jsme jej vymezili v úvodu naší studie. Této značky se užívá primárně kvůli metrickému výzkumu: disproporce mezi grafickou a fonetickou realizací slova by mohla zkreslit metrickou analýzu textu (např. *Shakespeare*, *Jules*). Lze ji však rovněž využít i při našem výzkumu vícejazyčnosti v současné poezii, jakkoliv jsme si vědomi, že tímto způsobem nemusí být nutně podchyceny zdaleka všechny xenismy a že, naopak, některá slova cizího původu už jsou v češtině natolik etablovaná, že je lze za xenismy těžko považovat (*jazz*, *intermezzo*, *blues*, *ghetto*, ale i neologismy typu *selfie*). Mimoto značnou část hodnot strukturního atributu <*foreign*> tvoří propria, jejichž spojitost s cizojazyčností je sporná.

I přes všechny tyto námitky zůstává pohled na inventář nejčastějších hodnot atributu <*foreign*> inspirativní: *Goethe* 28, *eddie* 19, *Jacob* 17, *villy* 11, *jazz* 11, *Strymm* 10, *blues* 10, *Blume* 10, *billboard* 9, *ghetto* 9, *New York* 8, *villa* 8, *Wulfild* 7, *Léthé* 7, *superstar* 7, *intermezzo* 7, *Franz* 7, *Beethoven* 7, *notebook* 6, *genius loci* 6, *forbes* 6, *rock* 6, *selfie* 6, *haiku* 6. Celkově bylo označkováno 2 843 různých textových segmentů (vyskytujících se v korpusu celkově 3628 krát), z nichž však jen prvních osm má celkovou frekvenci deset a výš, kdežto 2 441 segmentů (necelých 86 %) je v našich datech doloženo pouze jedinkrát – jde tedy o unikátní prvky, z čehož je patrná vysoká variabilita tohoto prostředku. Nicméně – jakkoliv jde o prostředek nápadný a výrazně ovlivňující čtenářskou recepci básně – z hlediska čistě frekvenčního se k němu autoři neuchylují vůbec často, naopak, cizojazyčné segmenty tvoří pouhých 0,38 % veškeré slovní materie v našem datasetu.

Co do délky segmentu dominují tzv. unigramy (jednotlivá slova) – z téměř dvou třetin, další pětinu tvoří bigramy (segmenty o délce dvou slov), trigramů je 251, tetragramů necelá stovka atd. (viz níže tabulka 1); s rostoucí délkou cizojazyčného textu frekvence výskytů úměrně klesá. Průměrná délka cizojazyčného

16 Pro srovnání uveďme statistiky pro Korpus českého verše, pokrývajícího 19. a začátek 20. století: 1 689 básnických sbírek, 76 699 básní, 2 664 989 veršů, 14 592 037 textových slov. Cit. podle PLECHÁČ, Petr – KOLÁR, Robert: Korpus českého verše. In: HLAVÁČOVÁ, Jaroslava (ed.): *Sborník semináře o digitálních zdrojích a službách ve společenských a humanitních vědách (WDH 2015). Praha, 24. září 2015*. Praha : Univerzita Karlova, Matematicko-fyzikální fakulta, 2015, s. 74 – 77. Pokud jde však o současnou poezii (nejen tu českou), omezují se korpusové analýzy na ad hoc pořízené vzorky malého rozsahu, a tudíž i reprezentativnosti – viz např. výzkum J. A. Hardyho a E. Rothenbergové, reflektující tři období moderních britských dějin (viktoriánskou epochu, meziválečné období a poválečná léta do roku 1990), pracuje s 393 básněmi o celkové velikosti přes 125 tisíc slov (*Evolution of modern British poetry: A keyword analysis.* Přednáška na International Corpus Linguistics Conference, 22. – 26. 7. 2019, Cardiff, UK).

17 Plecháč – Kolár, Korpus českého verše, c. d. Viz též http://versologie.cz/v2/web_content/tagset.php?lang=cz.

segmentu činí 1,92 slov (bez započtení hapaxů, tj. fragmentů s pouze jedním výskytem, by klesla na 1,8).

slov	četnost	%	15	4	0,1
1	2329	64,2	17	1	0,0
2	726	20,0	21	3	0,1
3	251	6,9	22	2	0,1
4	97	2,7	23	1	0,0
5	66	1,8	25	1	0,0
6	42	1,2	26	1	0,0
7	32	0,9	27	1	0,0
8	24	0,7	36	1	0,0
9	8	0,2	44	1	0,0
10	16	0,4	49	1	0,0
11	6	0,2	54	1	0,0
12	4	0,1	65	1	0,0
13	7	0,2	67	1	0,0

Tabulka č. 1: Frekvenční distribuce atributu <*foreign*> podle slovní délky

Drtivá většina případů atributu <*foreign*> se objevuje v těle básně, méně pak už na výlučných místech, jako je titul básně, resp. dedikace: v titulu je to 445 různých segmentů (z celkového počtu 10 790 titulů básní včetně netitulovaných), jinými slovy: titul každé v průměru 24. básně obsahuje jinojazyčný prvek. Tituly celých sbírek jsou převážně ze slov domácího původu, byť i zde najdeme výjimky (*À la thèse*, *Villa diabolica*, *Fantasía*).

Dále lze hledat výskyt značky <*foreign*> u jednotlivých autorů – a zjistit tak, kteří s cizojazyčnými prvky pracují ve své poezii nejvíce. Takto vypadá frekvenční špička autorů, seřazená podle relativní[18] frekvence.

	Filtr	doc.name	Freq	i.p.m. ▼	
1	p / n	Kafka, Tomáš	244	22206.04	
2	p / n	Torčík, Marek	44	21307.51	
3	p / n	Šanda, Michal	134	19539.22	
4	p / n	Hobza, Štěpán	58	17846.15	
5	p / n	Voda, David	41	15287.1	
6	p / n	Hájíček, Jiří	22	15130.67	
7	p / n	Zrno, Jan Šimon	162	14265.59	
8	p / n	Fridrich, Radek	61	14123.64	
9	p / n	Bušta, Pavel	30	11990.41	
10	p / n	Šimon, Patrik	59	10433.24	

Obrázek č. 1: Frekvenční špička autorů podle četnosti výskytu tagu <*foreign*> v rozhraní KonText[19]

18 S relativní frekvencí, nikoliv absolutní (sloupeček Freq), je tu potřeba pracovat proto, že velikosti textů od toho kterého autora jsou různě velké (např. Šanda 5 781 textových slov ze dvou sbírek, Torčík 1 883 slov z jediné sbírky, Kafka 9 098 slov ze dvou sbírek atd.). Zkratka *i.p.m.* (*instances per million* – česky „výskyty na milion“) normalizuje četnost vzhledem k pomyslnému korpusu o velikosti 1 milion textových slov.

19 MACHÁLEK, Tomáš: *KonText – rozhraní pro vyhledávání v korpusech*. Praha : FF UK, 2014. Dostupné z WWW: <http://kontext.korpus.cz/>.

Rozkliknutím pozitivního filtru (sloupec Filtr, odkaz p) se snadno dostaneme do seznamu konkordancí, v nichž u daného autora figuruje značka <*foreign*>. V něm už pak lze dané výskyty tagu (ale i konkrétních slov či sousloví) podrobit důkladnější analýze:

1	□	Metro	bydlet , čuromedán . celej <foreign>	squat	</foreign> nás nenávidí . nebejt mě
2	□	Metro	noci v depu na 95 <foreign>	Street Ft Hamilton	</foreign> spicha s inspirací . když
3	□	Metro	máti je v pohodě . <foreign>	Allen	</foreign> ? na muj vkus příliš
4	□	Metro	<foreign> Franz </foreign> K . ehm ! <foreign>	you realy haven't read his Process ?	</foreign> monděnně vyhlížející dáma s kloboukem
5	□	Metro	</pageNum> Londýn je úchvatnej . <foreign>	Tower Piccadilly	</foreign> byli jsme i v muzeu
6	□	Ošklivé příběhy z krásných slov	za americký hospodářský krize ho <foreign>	C .	</foreign> <foreign> Hyansey </foreign> z <foreign> RCA </foreign> vyhrabal v
7	□	Metro	Modiglianimu a <foreign> Picassovi </foreign> . pokud <foreign>	Picassovi	</foreign> stojí zakřenil se mně poletí
8	□	Ošklivé příběhy z krásných slov	Tyršáku odlesky slunce kreslily bizarní <foreign>	surrealistický	</foreign> obrazy . odsunul jsem se
9	□	Metro	keře šípků na pláních u <foreign>	Little Big Hornu	</foreign> " povídá lakotštinou sladce se
10	□	Metro	pendrek zezadu mezi nohy . <foreign>	Spray Haring	</foreign> vystříknul . <pageNum> 28 </pageNum>
11	□	Metro	hnědém koženém kabátu spěchajícího do <foreign>	Reichstagu	</foreign> : "nemáte oheň ? <pageNum>
12	□	Metro	pouti jak jinak pomyslel si <foreign>	G .	</foreign> a případ si jako hošík
13	□	Ošklivé příběhy z krásných slov	7 <foreign> noir </foreign> JAKO BY KDOSI <foreign>	Impair	</foreign> PŘIPIL RŮŽI a vypil jí
14	□	Metro	po něm vyšplhat třeba na <foreign>	Mt .	</foreign> . - dj - <pageNum>
15	□	Ošklivé příběhy z krásných slov	zapálil jsem si a pustil <foreign>	Blind Carenbecka	</foreign> . za americký hospodářský krize

Obrázek č. 2: Konkordance výskytu tagu <*foreign*> u M. Šandy

4. Kvantitativně-kvalitativní analýza vybraných autorů

Tento přístup k heuristice nám otevírá validní cestu k dominantám zkoumaného jevu, jak je přítomen ve výzkumném materiálu. Proto i následující pasáž bude zaměřena na kvantitativně-kvalitativní analýzu a interpretaci vícejazyčnosti v díle autorů, u kterých se xenismy vyskytují v nejvyšší relativní frekvenci: jsou jimi Michal Šanda, Marek Torčík a Tomáš Kafka.

Začněme posledně jmenovaným, protože poezií **Tomáše Kafky** se zabýval už Petr Mareš ve výše zmíněné monografii *„Also: nazdar!"*; bude tak hned zkraje zřetelná diference našich metodologických přístupů. Mareš – byť Kafkovým básnickým sbírkám věnuje celé dvě strany svého výkladu – prakticky pouze konstatuje (a na řadě příkladů dokládá), že vícejazyčnost u tohoto autora je především nástrojem jazykové hry a humoru a že tato hravost je založena na asociativních vazbách mezi prvky různých jazyků. Marešův výklad je tedy založen na metodě *close reading* vybraných básní obsahujících různojazyčné prvky a popis jejich funkce v rámci autorské poetiky. Zcela zde chybí pokus stanovit podíl xenismů na úhrnném významu jednotlivých textů, celé sbírky, resp. ve finále na jejím smyslu.

Cestou k tomu může být kvantitativní přístup, který předřadíme kvalitativní textostředné analýze. (Kvantitativní přístup zde tedy připravuje terén pro přístup kvalitativní, který ve výsledku směřuje k interpretaci zkoumaného fenoménu.) U multijazyčných textů, kterými Kafkovy básnické sbírky nepochybně jsou, je – domníváme se – zkraje nutné stanovit podíl jednotlivých uplatněných jazyků a eventuelně vytěžit z textových dat další okolnosti jejich uplatňování. V případě Kafkovy poezie se uplatňují tyto jazyky v následujícím poměru: němčina 37,4 %, angličtina 27,5 %, francouzština 18 %, ostatní jazyky (latina, španělština ad.) 17,1 %. Při konfrontaci s vágním čtenářským dojmem, na jehož základě se Kafkova poezie jeví jako výrazně „německá", jsou tato čísla poměrně překvapivá, neboť němčina se podílí na vícejazyčnosti Kafkovy poezie pouze zhruba třetinově. Vysvětlení poskytuje detailnější náhled do konkordance výskytů xenismů německé provenience. Němčina se totiž z největší míry podílí na úhrnné vícejazyčnosti víceslovnými sekvencemi – v řadě případů jde o kompletní syntaktické struktury. Totéž, avšak v mnohem menší míře, se vyskytuje u angličtiny, u dalších jazyků už nikoli. Němčina tedy konkuruje základnímu uplatněnému jazyku nikoli pouze

jako zdroj příležitostných lexikálních výpůjček, ale jako gramatický systém. Oproti tomu francouzština se jako zdroj xenismů uplatňuje především prostřednictvím vlastních jmen (antroponyma, toponyma) odkazujících k tradici francouzské literatury a umění – s P. Marešem tedy můžeme říci, že má funkci kulturní. Francouzština (na rozdíl od němčiny) u T. Kafky nepůsobí jako konkurence češtiny, ale spíše figuruje v rámci stylizace lyrického subjektu.

Stylizace subjektu je ostatně další problematikou, jež je u P. Mareše podceněna a v našem přístupu sehrává klíčovou, syntetizující funkci (viz výše o subjektovém konstruktivismu v současné poezii). Ludismus akcentovaný v Marešově výkladu nezpochybňujeme, ale nepovažujeme herní uplatnění xenismů za jejich základní funkci ve struktuře textů – hra s různojazyčným jazykovým materiálem je spíše prostředkem, jehož funkce spočívá právě v participaci na stylizování lyrického subjektu. P. Mareš pouze konstatuje, že T. Kafka využívá „náhodných mezijazykových podobností",[20] jako např. ve verších „*Come, I see your leg / kam? já se tě lek*",[21] ale zcela rezignuje na interpretaci. Kafkův subjekt je stylizován jako vzdělanec, jako ten, kdo zná jazyky a je mimořádně jazykově kompetentní. Zmíněné mezijazykové podobnosti potom vyznívají jako projev ironie, resp. jako záměrná komičnost ironicky předstírající neschopnost relevantního překladu. A právě toto bylo na počátku devadesátých let, kdy sbírka vznikala, součástí životní praxe velké části české populace. Jde vlastně o smích nad tím, jak se jazyková krajina otevřela, rozrůznila a jak se v ní Češi, dlouhá desetiletí izolovaní od západní jazykové kultury, ztrácejí.

Xenismy mají u T. Kafky, jak už jsme naznačili, často také podobu vlastních jmen. V těchto případech zřetelně na povrch vystupuje Kafkův základní kompoziční postup, který spočívá v konfrontaci vysokého a nízkého (tento postup se projevuje už na jazykovém stylu, kde je do těsných sousedství zasazována exkluzivní slovní zásoba a vulgarismy). Některá jména spadají do kulturní encyklopedie vzdělaného člověka (např. Lorca, Brecht), jiná už nikoli – neboť jde nejčastěji o jména sportovců (ta autor vysvětluje v poznámkách za texty). Tato skutečnost ovšem neoslabuje intelektuálskou stylizaci subjektu, naopak je jejím potvrzením: jako by nás i tímto mluvčí chtěl přesvědčit o tom, že jeho rozhled je tak velký, že zahrne i osobnosti, které intelektuálská komunita většinou nezná, nebo dokonce jejich činností pohrdá.

Stylizaci Kafkova subjektu a zejména podobu a intenzitu uplatnění vícejazyčnosti lze v tomto případě považovat za dobově příznačnou a reagující na proměny jazykové kultury v raně transformačním období. Ostatně po dvou sbírkách z devadesátých let (*Kvaše*, *Ze světa*) následovalo u T. Kafky dlouhé básnické odmlčení a ve sbírce třetí, vzniklé a vydané už ve zcela jiné atmosféře (*Verše v roce*, 2005), je podíl xenismů mnohem menší. Tuto kontextovou interpretaci, vycházející z empirických korpusových dat, podporují i „sousední" básnické výkony, v nichž se zejména podíl němčiny jeví rovněž jako významný a dobově příznačný, byť autorská intence je poněkud odlišná.[22] Němčina zde figuruje jako jazyk nejbližšího západního souseda a funguje jako znak otevírání se Západu. Pro jedny

20 Mareš, „Also: nazdar!" Aspekty textové vícejazyčnosti, c. d., s. 201.

21 KAFKA, Tomáš: *Kvaše.* Praha : Nakladatelství Q, 1994, s. 7.

22 Na mysli zde máme zejména sbírku *Zbrusu nové jarmareční písně českotuzemské* Jiřího Žáčka (1995), která vyšla rovněž v polovině devadesátých let a rovněž tematizuje interakci české jazykové kultury s kulturou německou – byť ve značně jiném, varovném gardu.

je to příležitost (byť spojená s pocitem vlastní nedostatečnosti), pro druhé nebezpečí. Potenciál vzbuzovat antigermánské averze je u němčiny vždy latentně přítomen, neboť trauma druhé světové války, resp. německého protektorátu nad českými zeměmi je na začátku devadesátých let u značné části populace dosud živé. I v podtextu Kafkových básní je toto trauma přítomno, byť je ironicky podrýváno a zlehčováno hravostí a humorem.

Uplatnění xenismů je v Kafkově poezii evidentně (s Marešem řečeno) hloubkové – vícejazyčnost je jedním ze základních strukturních segmentů autorské poetiky a významně se podílí na stylizaci lyrického subjektu. Mohli bychom očekávat, že podobně tomu bude také v poezii **Michala Šandy**, protože jeho texty mají podle KSP vůbec nejvyšší relativní frekvenci xenismů. Skutečnost je ovšem poněkud jiná.

Na vícejazyčnost Šandovy poezie se dominantně podílí angličtina: anglojazyčné sekvence představují 90,2 % celkového počtu xenismů v jeho verších. O multijazyčnosti zde tedy nemůže být řeč, jakož ani o konkurenci mezi angličtinou a češtinou. Téměř polovinu (46 %) xenismů tu totiž představují vlastní jména osobností z oblasti umění, kultury, popřípadě sportu. Xenismy zde tedy nesou především kulturní funkci. Nutné je ovšem podívat se blíže na to, jaké provenience tyto kulturní xenismy jsou. Z velké části se totiž jedná o jména rockových či bluesových hudebníků či názvy kapel. Víceslovné jinojazyčné sekvence jsou zase významně tvořeny citacemi anglojazyčných písňových textů nebo jejich názvů.

Xenismy se u Šandy silně podílejí na stylizaci lyrického subjektu, na úhrnném významu či smyslu jeho děl už ovšem velmi málo. Angličtina, a především pak angličtina rockových skladeb jejich hvězdných interpretů, spoluutváří bohémskou a pijáckou stylizaci subjektu, který je autorem prezentován jako nonkonformní mladý muž užívající plnými doušky radosti života ve velkoměstě. Angličtina zde má tedy spíše statusovou roli, resp. roli identifikátoru s velkoměstskou (rockerskou) subkulturou.

Konečně, **Marek Torčík** (*1993), generačně nejmladší z trojice námi analyzovaných autorů, leccos naznačuje už zaumným titulem své sbírky (*Rhizomy*) a použitím řecké alfabety v jedné básni. Vzhledem k přihlášení se k odkazu francouzského myslitele Gillese Deleuzea[23] bychom v Torčíkových textech předpokládali větší zastoupení galismů, avšak kromě třikrát užitého titulu básně *mise en scène* a titulních *rhizomů* je tu už jediný výskyt: *dans le vrai*. Naopak suverénně dominuje angličtina: z 44 xenismů je 32 anglicismů (72,7 %), často se jedná o dlouhé syntaktické celky (*for over a decade meandering streets under the heavy april rain…* – 11 slov). Lze to přičítat jednak vlivu, jemuž je dnešní mladá generace tvůrců této nové *lingva franca* vystavena, jednak faktu, že sám M. Torčík vystudoval anglickou filologii.

Cizojazyčné elementy (zmiňme ještě osm německojazyčných prvků a jeden latinský citát) tu hrají zásadní roli ve své funkci metajazykové. Autorskému subjektu, snažícímu se dobrat významu a smyslu určitých slov či konstrukcí – aby je následně mohl coby básník organicky používat, už nestačí jeho mateřština;

23 Interpretačním klíčem k vícejazyčné povaze sbírky pak zřejmě má být její motto převzaté z *Dialogů II* Gillese Deleuzea a Claire Parnetové: „mluvit svým rodným jazykem jako cizinec." Jazyková praxe Torčíkových básní je ovšem spíše opačná – příležitostně používá cizí jazyk, jako by se nelišil od jeho mateřštiny.

slova z ní konfrontuje s jejich cizojazyčnými protějšky, která se mu zdají být v daném kontextu případnější a trefnější. Báseň se tu stává až jakýmsi meditativním metalingvistickým traktátem:

„jazyk je heterogenní realita“
měníš slova, jak se ti hodí
ohýbáš významy i sloh

když řeknu lehá a něha, v hlavě
hashtagy jemných nuancí
#signifier #signified

binární vztahy/rhizomy?

větná fragmentace:
slovo utterance představuje otevřený systém
neideální. rozmělněný stav
slovní entropii:

bezbarvé
zelené myšlenky
sní, furiously[24]

Mísení nejen jazyků, ale i prvků moderních (užití hashtagů) a těch ustavujících tradici (odkaz na Saussureovu dichotomii *označující – označované*, byť i zde bychom mohli očekávat původní, tj. francouzské termíny) a prolínání roviny jazykové s metajazykovou odkazuje jednoznačně k postmodernímu vnímání světa a sugerují také snahu o generační výpověď. Pro generaci narozenou po roce 1989 je už zcela běžné plynně během konverzace přecházet z češtiny do angličtiny a zpět. Přesně to dělá M. Torčík ve svých básních – nejzřetelněji pak tam, kdy do české syntaktické struktury plynule napojuje strukturu anglojazyčnou. Xenismy zde tedy slouží jako prostředek podílející se na stylizaci subjektu jako mladého intelektuála, příslušníka své generace, která se hrdě hlásí ke kosmopolitnímu, resp. postnacionálnímu vnímání světa (Torčík ho demonstruje právě prostřednictvím xenismů a jejich organického včleňování do veršů psaných většinově mateřštinou).

5. Vícejazyčnost na amatérských literárních serverech

Vedle subkorpusu tištěné poezie je v KSP zahrnuta i básnická tvorba, resp. reprezentativní vzorek z ní, soustředěná na šestici těchto literárních serverů (řazeno sestupně podle velikosti): *liter.cz*, *pismak.cz*, *totem.cz*, *libres.cz*, *psanci.cz*, *xxvi.cz* (webové stránky Skupiny XXVI).

Budování webových korpusů je obecně složité v mnoha ohledech, a to i v našem případě. Problematické není ani tak postahovat obsah jednotlivých serverů (koneckonců jedná se o pouhých šest portálů, nikoliv např. o celý český web),

24 TORČÍK, Marek: *Rhizomy*. Brno : Šimon Ryšavý, 2016, s. 38.

náročnost úkolu tkví ve zpracování a vyčištění syrových dat a jejich převedení do potřebného xml formátu. Ukazuje se totiž, že přispěvatelé na literární servery jsou kromobyčejně vynalézaví, co se ozvláštnění vlastních výtvorů týče – využívají prakticky všechny možnosti, jež jim daný systém nabízí (řezy písma, vkládání emotikonů, obrázků, tabulek aj.). Mimoto rádi experimentují s formou (výjimkou nejsou ani pokusy o kaligramy), ne vždy respektují žánrovou orientaci rubrik, samotné básnické texty nesystematicky doplňují o komentáře v paratextech – a to vše představuje nemalé komplikace při tvorbě korpusu.

Pro účely webové složky KSP byl stažen výchozí datový materiál o velikosti 37,8 milionu slov v bezmála půl milionu textů (celkem 374 MB). Ten je následně zapotřebí vyčistit, a to nejprve globálně (níže popsanou dvojicí filtrů), a posléze jednotlivé básnické texty zbavit nežádoucího obsahu, který není součástí básně, ale webové šablony (záhlaví, zápatí, navigační odkazy, případně i reklama aj.) a negativně by ovlivnil frekvenční statistiky. Takto bylo odfiltrováno cca 11 % textů. Údaje o velikosti zdrojových dat shrnuje následující tabulka:

server	textů před filtrací	z toho poezie	sk	sk [%]	textů po filtraci	velikost [MB]	slov
xxvi.cz	507	206	0	0	206	0,7	87 989
totem.cz	42 045	40 972	1 201	2,93	39 771	28,1	2 898 766
liter.cz	203 338	199 117	4 299	2,16	194 818	148,7	16 194 551
pismak.cz	199 837	176 264	17 350	9,84	158 914	122,5	10 847 660
psanci.cz	21 402	20 515	552	2,69	19 963	14,4	1 522 553
libres.cz	29 443	29 023	232	0,8	28 791	22,8	2 458 200
Souhrn	**496 572**	**466 097**	**23 634**	**5,07**	**442 463**	**337,1**	**34 009 719**

Tabulka č. 2: Statistiky webové složky KSP

Nejprve aplikujeme prozaický filtr. Ačkoliv jsme z daných serverů stahovali pouze texty z relevantních rubrik, celkem často se tu objevují vedle básní také prózy – a zdaleka ne jen básně v próze. Tyto texty musíme pochopitelně též odfiltrovat, aby nezkreslovaly metrickou statistiku. Prozaické texty detekujeme na základě průměrné délky řádku: ta nesmí být vyšší než deset slov (min. délka řádku je pak logicky jedno slovo), zároveň nesmí být nejdelší řádek delší než dvacet slov. Jinými slovy: pokud je průměrná délka menší nebo rovna deseti slovům na verš a maximální délka verše je dvacet slov, jedná se o poezii; v opačném případě text zahazujeme.[25] Takto bylo odfiltrováno cca 6,1 % textů.

Dále je nutno vyloučit *plně* cizojazyčné texty.[26] Na rozdíl např. od výše zmíněné básně W. Heinrichové *Vec* tyto texty do KSP nezahrnujeme. Báseň *Vec*

25 Automatickou metodou však nelze rozlišit básně v próze od obvyklých próz, na tuto distinkci bude nejspíše zapotřebí rezignovat a tato forma tak bude zastoupena pouze v subkorpusu tištěné poezie.

26 Toto „vyloučení" je pro nás zároveň ale ze sociologického hlediska zajímavé, jelikož dovoluje zpřesňovat úvahy nad internacionální povahou internetové literární komunikace. Z hlediska automatizované jazykové a textové analýzy jsou tyto texty ovšem nežádoucí. Jinou otázkou jsou však jednotlivé xenismy, které v textech pochopitelně zůstávají.

nelze z dané básnické sbírky (a tudíž ani z KSP) vyřadit, neboť by tím utrpěla celistvost sbírky. Oproti tomu básně na literárních serverech vstupují do komunikace jako jednotliviny.[27] K detekci nečeských textů je možné využít různé jazykové filtry, opírající se o specifické dvou- až tříznakové sekvence pro ten který jazyk – takto byly odhaleny četné jinojazyčné texty: nejčastěji slovenské, méně pak anglické či německé, polské a ruské.

Nás při analýze básnických textů na *českých* literárních serverech zajímala navíc specifická otázka sociolingvistická: kolik přispěvatelů na ně píše své výtvory *slovensky*? Za tímto účelem jsme vytvořili ad hoc diferenční heslář nejfrekventovanějších slovenských slov ve vztahu k češtině[28] a na jeho základě detekovali texty, které jsou potenciálně slovenské, nikoliv české provenience. Počet vyřazených textů bude ovšemže záviset na nastavení hranice poměru slovakismů vůči neslovakismům; stanovujeme ji dynamicky, podle počtu slov v básni: nad 100 slov 1 %, 51–100 slov 2 %, do 50 slov 5 %. Toto kritérium nám vytipovalo přes 23 tisíc potenciálních slovenských textů (viz sloupce sk a sk [%] v tabulce 2), tedy každý zhruba dvacátý text ve zdrojových datech. Nejčastěji jsou slovenské texty zastoupeny na serveru *Písmák* (téměř 10 %), na ostatních portálech se vyskytují o poznání řidčeji.

Po aplikaci obou výše uvedených filtrů bylo z výchozího datasetu vyřazeno přes 54 tisíc nevyhovujících textů, tedy necelých 11 %. Souhrnný aktuální objem webové složky KSP tak činí více než 442 tisíc básní čítajících přes 34 milionů slov (podrobněji výše v tabulce 2). Nelze však vyloučit další redukci webových dat na základě podrobnější heuristiky; zapotřebí budou ještě alespoň namátkové manuální kontroly, které odhalí problematické případy (včetně potenciálních duplicit), jež by v KSP být neměly. Všechna naše výše popsaná rozhodnutí jsou do značné míry arbitrární, a jako taková se mohou při finální fázi výstavby KSP změnit.

Je zjevné, že při takovémto rozsahu dat je plošná *manuální* anotace vyloučena. Lze však provést anotaci *automatickou*: aplikovat soupis hodnot tagu <*foreign*>, pořízený manuálně na vzorku tištěných sbírek (viz kap. 3), na data z webu, tedy nechat tímto tagem označkovat jednotlivé texty z webového subkorpusu. Stejně tak lze aplikovat na data jakýkoliv jiný seznam, jako např. výše

27 Literární servery většinou umožňují sdružovat básně do „sbírek" či cyklů, nicméně tato možnost je (na rozdíl od knižního média) fakultativní a v poměru k celkovému objemu zveřejňovaných textů je využívána spíše okrajově.

28 Za východisko jsme vzali seznam tisíce nejfrekventovanějších slovních tvarů (nikoliv lemmat) z dvou veřejně dostupných korpusů slovenštiny: NÁBĚLKOVÁ, Mira – VAVŘÍN, Martin: *Korpus InterCorp – slovenština, verze 12 z 12. 12. 2019*. Praha : Ústav Českého národního korpusu FF UK, 2019. Dostupný z WWW: http://www.korpus.cz; BENKO, Vladimír: *Araneum Slovacum Maximum, verze 15.04*. Praha : Ústav Českého národního korpusu FF UK, 2015. Dostupný z WWW: http://www.korpus.cz. Obě frekvenční špičky jsme sloučili, deduplikovali a smazali tvary, jež by mohly být homonymní s českými lexémy. Např. frekventované slovenské slovo *bol* v našem seznamu není, neboť v češtině jde o synonymum slova *žal* (a jeho užití v básnických textech se dá jistě předpokládat), zato tvar *boli* ano (hypotetickou bezdiakritickou verzi slova *bolí* ignorujeme, do KSP neplánujeme zahrnovat básně bez diakritiky, nebudou-li součástí tištěných sbírek); podobně slov. *budem* (1. osoba singuláru indikativu futura) je homonymní s obecněčeským futuritním tvarem 1. osoby plurálu, proto ho do našeho hesláře nezařazujeme, avšak tvar *budú* ano (moravismy taktéž ignorujeme); podobně též slovenské i obecněčeské *chcem*, kdežto tvary *chcel*, *chcela*, *chceli*, *chcú* jsou specifické pouze pro slovenštinu. Výsledný seznam čítá 662 slovních tvarů. (Za tento podnět děkujeme dr. Vladimíru Benkovi, jenž podobný heslář používá k snazšímu rozpoznávání slovenských webových stránek od českých.)

popsaný diferenční slovenský heslář (a texty automaticky značkovat např. novým tagem <*foreign_sk*> apod.).

6. Závěry

Byť je naše stať spíše zprávou o vznikající datové základně a možnostech jejího využití, domníváme se, že už na této pilotní studii je poměrně dobře vidět, jak přínosná může být kvantitativní datová analýza v první fázi jazykově zaměřeného literárního výzkumu (viz samotné setřídění materiálu podle relevance). Zároveň se ukázalo jako efektivní, resp. nutné propojování kvantitativního a kvalitativního přístupu, protože teprve badatelova interpretační kompetence může rozhodnout o hranicích zkoumaného pole a významu prvků v něm přítomných (hrubý materiál poskytnutý konkordancí je ovšemže nutno dále revidovat a kriticky hodnotit). Potvrdilo se tedy, že kvantitativní přístup není samospasitelný, že při jednostranném akcentu na něj může dojít k deformacím, zároveň se však zřetelně ukázalo, o kolik validnější jsou výsledky výzkumu opírající se zprvu o *distant reading* výzkumného materiálu jako celku, nikoli pouze o analýzu nahodile vybraných vzorků. Potvrdila se také (obecně platná) nezbytnost silného interpretačního zaměření k některému z klíčových významových uzlů (my jsme zvolili lyrický subjekt), v opačném případě – a při korpusovém přístupu zvláště – hrozí ulpívání na výčtech příkladů a jejich komentování.

Přínos korpusové optiky ovšem nekončí ve fázi heuristické, rešeršní – výsledky konkordancí jsou badateli pochopitelně k dispozici po celou dobu analytické práce s textem – a netřeba zdůrazňovat, o kolik efektivnější je to cesta oproti zhotovování a vytěžování vlastních excerpt. V neposlední řadě analytická práce realizovaná s oporou v korpusových datech stojí na mnohem pevnějších základech co do kontextualizace dílčích zjištění; jinými slovy řečeno: mnohem snadněji a validněji se stanovuje vztah mezi částí a celkem zkoumaného materiálu (např. zjištění o vysoké frekvenci cizích vlastních jmen u M. Šandy můžeme okamžitě konfrontovat s konkordancí celého korpusu a vidíme, že toto zjištění nemůžeme prezentovat jako diferenční znak, kterým se tento autor vymezuje vůči soudobému kontextu, ale že jde naopak o jev poměrně obvyklý a daný autor je tedy spíše konformní s obvyklou praxí).

Byť práce na segmentu textů převzatých z literárních serverů je zatím v počátcích, podařilo se nám už nyní zpřesnit, resp. opravit tradovanou hypotézu o četném zastoupení slovenských textů na českých literárních serverech: jejich distribuce napříč servery je velmi nevyrovnaná; slovenští přispěvatelé vyhledávají především server *pismak.cz* (až 10 % slovenského obsahu, ostatní servery mají slovenských textů do 3 %). Do budoucna tedy KSP může sloužit nejen jako opora při textostředných analýzách, ale může být uplatněn i v rámci projektů zaměřených spíše sociologicky.

Naším cílem do budoucna je další rozšiřování KSP o nová data (autoři starší generace, některé další texty z webu, ale též lepší filtrace stávajících webových textů), tak aby šlo skutečně o maximálně reprezentativní vzorek současné české poezie. V roce 2021 by měla být hotova verze 1.0, která už bude zpřístupněna široké veřejnosti. Pevně věříme, že tento datový zdroj – mimochodem v celosvětovém měřítku v takovémto rozsahu naprosto unikátní a pionýrský počin – zvláště uvítá veřejnost odborná.

Kvantitatívne-korpusový prístup by podle našeho názoru měl být vnímán stejně samozřejmě jako jiné nástroje vědecké infrastruktury, např. bibliografické databáze. Dnes už nikoho ani nenapadne sbírat bibliografická data od nuly, zato s naprostou samozřejmostí zahajujeme svou odbornou práci rešeršováním v digitálních bibliografických databázích. Při textostředných analýzách by měly začít sehrávat obdobnou roli jazykové korpusy.

Zdroje a nástroje

BENKO, Vladimír: *Araneum Slovacum Maximum, verze 15.04*. Praha : Ústav Českého národního korpusu FF UK, 2015. Dostupné z WWW: http://www.korpus.cz.

MACHÁLEK, Tomáš: *KonText – rozhraní pro vyhledávání v korpusech*. Praha : FF UK, 2014. Dostupné z WWW: <http://kontext.korpus.cz/>.

NÁBĚLKOVÁ, Mira – VAVŘÍN, Martin: *Korpus InterCorp – slovenština, verze 12 z 12. 12. 2019*. Praha : Ústav Českého národního korpusu FF UK, 2019. Dostupné z WWW: http://www.korpus.cz.

Pramene

KAFKA, Tomáš: *Kvaše*. Praha : Nakladatelství Q, 1994.

KAFKA, Tomáš: *Ze světa*. Brno : Petrov, 1996.

KAFKA, Tomáš: *Verše v roce*. Brno : Petrov, 2005.

ŠANDA, Michal: *Ošklivé příběhy z krásných slov*. Praha : Protis, 1996.

ŠANDA, Michal: *Dvacet deka ovaru*. Praha : Klokočí, 1998.

ŠANDA, Michal: *Metro: postmoderní jukebox*. Praha : Protis, 1998.

TORČÍK, Marek: *Rhizomy*. Brno : Šimon Ryšavý, 2016.

ŽÁČEK, Jiří: *Zbrusu nové jarmareční písně českotuzemské*. Mnichov : Obrys/Kontur – PmD, 1995.

Literatúra

FIALOVÁ, Alena (ed.): *V souřadnicích mnohosti: česká literatura první dekády jednadvacátého století v souvislostech a interpretacích*. Praha : Academia, 2014.

JOCKERS, Matthew Lee: *Macroanalysis: digital methods and literary history*. Urbana : University of Illinois Press, 2013.

KREMNITZ, Georg: *Mehrsprachigkeit in der Literatur. Ein kommunikationssoziologischer Überblick*. Wien : Praesens Verlag, 2015.

MACUROVÁ, Alena – MAREŠ, Petr: Soužití a střet. Ke stylu vícejazyčného textu. In: *Stylistika*, roč. 5, 1996, s. 164 – 182.

MAREŠ, Petr: *Nejen jazykem českým: studie o vícejazyčnosti v literatuře*. Praha : Filozofická fakulta Univerzity Karlovy, 2012.

MAREŠ, Petr: *„Also: nazdar!" Aspekty textové vícejazyčnosti*. Praha : Karolinum, 2003.

MORETTI, Franco: *Distant reading*. London : Verso, 2013.

MOSER, Wolfgang: *Xenismen. Die Nachahmung fremder Sprachen*. Peter Lang : Frankfurt am Main, 1996.

NÁBĚLKOVÁ, Mira: Literárne servery a amatérsky česko-slovenský literárny kontext. In: SUWARA, Bogumiła – HUSÁROVÁ, Zuzana – ŠVELCH, Jaroslav – NÁBĚLKOVÁ, Mira: *V sieti strednej Európy: nielen o elektronickej literatúre*. Bratislava : Slovak Academic Press, 2012, s. 281 – 298.

PIORECKÝ, Karel: Červenkova teorie lyrického subjektu. In: *Česká literatura*, roč. 54, 2006, č. 6, s. 30 – 55.

PIORECKÝ, Karel: *Česká literatura a nová média*. Praha : Academia, 2016.

PIORECKÝ, Karel: Literary work as a leisure activity. Amateur literary forums on the Czech internet. In: CARNICELLI, Sandro – McGILLIVRAY, David – McPHERSON, Gayle (eds.): *Digital Leisure Cultures: Critical perspectives*. New York : Routledge, 2017, s. 193 – 206.

PLECHÁČ, Petr – KOLÁR, Robert: Korpus českého verše. In: HLAVÁČOVÁ, Jaroslava (ed.): *Sborník semináře o digitálních zdrojích a službách ve společenských a humanitních vědách (WDH 2015). Praha, 24. září 2015*. Praha : Univerzita Karlova, Matematicko-fyzikální fakulta, 2015, s. 74 – 77.

PLECHÁČ, Petr – KOLÁR, Robert: *Kapitoly z korpusové versologie*. Praha : Akropolis, 2017.

HRUŠKA, Petr – MACHALA, Lubomír – VODIČKA, Libor – ZIZLER, Jiří (eds.): *V souřadnicích volnosti: česká literatura devadesátých let dvacátého století v interpretacích*. Praha : Academia, 2008.

ZMĚLÍK, Richard: *Kvantitativně-korpusová analýza a literární věda: model a realizace autorského korpusu a slovníku Jana Čepa v kontextu zahraniční a české autorské lexikografie*. Olomouc : Univerzita Palackého v Olomouci, 2015.

ZMĚLÍK, Richard: *Konceptualizace barev v narativní fikci na pozadí kvantitativních modelů*. Olomouc : Univerzita Palackého v Olomouci, 2019.

PhDr. Karel Piorecký, Ph.D.
Ústav pro českou literaturu AV ČR, v. v. i.
Na Florenci 1420/3
110 00 Praha 1
Česká republika
E-mail: piorecky@ucl.cas.cz

Mgr. Michal Škrabal, Ph.D.
Ústav Českého národního korpusu
Filozofická fakulta Univerzity Karlovy
Panská 890/7
110 00 Praha 1
Česká republika
E-mail: michal.skrabal@ff.cuni.cz