

Proměny prózy v letech 1992 až 2018

Petra Poukarová — Václav Cvrček

The changing face of prose between 1992 and 2018

This study summarizes a corpus-based analysis of tendencies in register variation of Czech-written fiction texts in the period from 1992 to 2018. The analysis is based on projection of the results from a large sample of Czech prose texts (1070 texts, 12.7 mil. words) on a general register model (established by previous research using multidimensional analysis). The major tendencies found in the material are a decrease of cohesion level, addressee coding and retrospective narration, and increased polythematicity/lexical richness. These findings are supplemented by additional analyses of the role of translation, the position of a text excerpt in the original text (beginning, middle and end) and type of text in the results.

Klíčová slova / Keywords

beletrie — próza — registr — multidimenzionální analýza — korpus — překlad
fiction — prose — register — multidimension analysis — corpus — translation

Kontakt

Filozofická fakulta Univerzity Karlovy; Petra.Poukarova@ff.cuni.cz —
vaclav.cvrcek@ff.cuni.cz

Petra Poukarová — Václav Cvrček. „Proměny prózy v letech 1992 až 2018“; *Česká literatura* LXX, č. 6, s. 678–710

<https://doi.org/10.51305/cl.2022.06.02>

I. Úvod

Cílem této studie je ukázat na rozsáhlých datech z Českého národního korpusu (ČNK) proměny registrů česky psané prózy v posledním čtvrtstoletí. Vůdtkem přitom pro nás nebude hledisko literárněvědné, které by sledovalo imanentní vývojové tendence v rámci české beletrie, ale spíše perspektiva čtenáře a jeho možné vnímání prozaického textu. Budeme texty hodnotit podle toho, jaké jazykové prostředky v úhrnu užívají, což v současné české lingvistice nejlépe zachycuje termín *registr/rejstřík* (v tradici české stylistiky je mu nejbližší *funkční styl*). Zároveň se neomezíme pouze na českou produkci nebo na produkci kriticky či literárněvědně oceňovanou, ale sledované období budeme mapovat v celé šíři (tj. díla kvalitní i spotřební, originální i překladová) tak, jak současnou tvorbu recipují čtenáři.

Ve své podstatě modeluje náš přístup situaci, kdy běžný čtenář pravidelně vstupuje do beletristického oddělení v průměrném knihkupectví, náhodně vybere několik úryvků aktuální produkce v daném roce a snaží se udělat udělat si z nich představu o tom, jak se píše beletrie (jak bude zjevné z popisu dat, půjde nám o prózu, především o romány, a přihlížet budeme i ke kratším povídkám).

Ačkoli takové vymezení může znít jako poměrně extrémní extenze toho, co bývá označováno jako *literární kultura*, jsme přesvědčeni, že z lingvistické perspektivy dává smysl; cílem je na základě empirických metod problematizovat způsoby, jimiž jazykové ztvárnění literárních děl (v rámci funkčního přístupu označované jako *umělecký styl* či *styl krásné literatury*) uchopujeme (k vymezení těchto pojmů viz odd. 2.)

Horizont našeho výzkumu utváří na jedné straně dostupnost prozaických dat v korpusech ČNK a na straně druhé významné historické předěly. Beletristické texty se na rozdíl od jiných textových typů dostávají do korpusu se značným zpožděním, v zásadě pouze při aktualizaci reprezentativních korpusů řady SYN, což se děje jednou za pět let. V době psaní této studie tak poslední texty, které jsou našemu bádání dostupné, pocházejí z roku 2019 (vzhledem k jejich malému zastoupení je však do studie nezařazujeme, nebylo by to srovnatelné s datovou základnou pro ostatní roky, a sledování tedy končíme rokem 2018). Na opačné straně časové osy se jako přirozený zlym jeví rok 1989, resp. 1990, v němž se radikálně proměňuje beletristická produkce (domácí i překladová).¹ Vzhledem k tomu, že v období bezprostředně

1 O tom, že překladová literatura tvoří velkou část knižní produkce, svědčí údaje Národní knihovny, která podle L. Chlumské (CHLUMSKÁ 2017) registrovala v roce 2005 zhruba 33 % překladových titulů neperiodických publikací, v roce 2014 to bylo více než 40 % a na konci sledovaného období, v roce 2018, to bylo už více než 45 % (viz https://www.nkp.cz/soubory/ostatni/vykaz_dd2018.pdf). Je přitom zjevné, že zdaleka nejde o beletrii, či dokonce jenom o prózu.

po sametové revoluci se knižní tvorba zčásti orientovala na vydávání starších děl, která nebyla dříve dostupná, máme silněji zastoupenou aktuální produkci až od roku 1992.

Výsledky a jejich interpretace, které představíme v odd. 5, jsou přirozeně limitovány použitou metodou (viz odd. 3) a dostupnými daty (viz odd. 4). Ještě předtím je ovšem třeba zmapovat, jakým způsobem se k registru či stylu prozaických textů stavěla funkční stylistika (viz odd. 2).

2. Próza a stylistika

Styl umělecké literatury je tradiční stylistikou charakterizován jako styl vyznačující se dominancí estetické funkce (a potlačením funkce sdělné, ovšem ne její úplnou absencí). Její naplnění je spojeno s uplatněním jazykové aktualizace, tedy účelné odchylky od jazykového standardu (viz např. JELÍNEK 1995: 769).

Pestrost a heterogenost textů, jež jsou označovány jako umělecké, proměnlivost chápání toho, co má estetický účinek na recipienta, největší možnosti při stylizování, variabilita komunikace, rozkročení mezi komunikací veřejnou a privátní, psanou a mluvenou, dialogickou a monologickou, vymykání se normám (HOFFMANNOVÁ ET AL. 2016: 396, 403), to vše činí v současných kvalitativních přístupech tento styl velmi specifickým, hůře uchopitelným a podléhajícím zobecnění. Proto se jako účelné jeví uvažovat v této oblasti o více stylech, a to prozaickém, poetickém a dramatickém (IBID.: 397).

Definice stylu umělecké literatury na jednu stranu předestírá možnosti, z jakého úhlu pohledu na analyzované umělecké texty pohlížet (především jde o popisy aktualizací, odchylek od spisovné normy, viz dále), na druhou stranu pestrost žánrů (vedle románů například memoáry, deníky, autobiografie) má dopad na výběr jazykových prostředků — není neobvyklé, že autoři užívají výrazy zcela neutrální, bez jakéhokoli „atraktivního ozvláštňení“ (IBID.: 405). S tím souvisí i to, že repertoár těchto prostředků, tzv. stylovou vrstvu, nelze popsat (IBID.: 397). Na základě přítomnosti určitých jazykových prostředků není možné text automaticky charakterizovat jako umělecký; naopak při popisu musí docházet k velké míře abstrakce a analýzy uměleckého stylu se opírají převážně o vybrané konkrétní texty či autory (MAREŠ 2017). Tradičně bývají jako nositelé estetické funkce označovány poetismy, ale jejich výskyt v textech je omezený (viz HOFFMANNOVÁ ET AL. 2016: 397); toto tvrzení podporuje i výzkum zaměřený na multidimenzionální analýzu češtiny, kde jazykový rys poetismus významně statisticky nefiguroval (viz CVRČEK ET AL. 2020).

Jak už bylo zmíněno výše, u stylových analýz uměleckých textů je patrné, že nejčastěji se badatelé zaměřují na odchylky od spisovného standardu a na všech jazykových rovinách si všímají užívání jazykových prostředků spojených primárně s obecnou češtinou, případně nářečím, tedy výrazů a konstrukcí reflektujících mluvenost. Například Neil Bermel (BERMEL 2000,

2001) se věnuje dialogu v české beletrii z hlediska diglosie; beletristický dialog zvolil pro bližší popis z toho důvodu, že se jedná o útvar na pomezí spisovné češtiny (jde o texty psané) a češtiny obecné (charakter dialogu je neformální, soukromý) (IDEM 2001: 16). Karen Gammelgaardová (GAMMELGAARD 1997) si klade otázku, zda se obecná čeština stává spolutvůrcem estetického významu. Obecnou češtinu v literatuře studuje i Concetta Maglioneová (MAGLIONE 2001), jež se zabývá prvky spisovnými a nespisovnými v povídkách Petra Šabacha. Petr Mareš (např. MAREŠ 2012) se pak věnuje vícejazyčnosti v literatuře. Autorským záměrem užívání nespisovných prostředků může být charakterizace postav, zachycení expresivity (ČMEJRKOVÁ — HOFFMANNOVÁ 2011), modelace každodennosti či kreativita (HOFFMANNOVÁ ET AL. 2016: 416).

Konkrétní jazykové rysy, jež bývají v odborné literatuře označovány jako typické pro umělecký styl, nejsou vlastní jedné jazykové rovině, ale mají své místo v celém inventáři češtiny. Na hláskové úrovni se zmiňují např. krácení samohlásek, stylizující výslovnost běžně mluvené češtiny, na rovině morfologie obecněčeské tvary či tvary nářeční, frekvence slovesné osoby, vydělující pásmo postav a vypravěče, slovesný vid (IBID.: 416), k dalším relevantním rysům se řadí užití ukazovacích zájmen, částic zesilujících nebo naopak oslabujících význam, přítomnost hodnotících adjektiv a adverbii. Také syntax se opírá o mluvenost — k rysům mohou náležet málo členěný řečový proud, juxtaopozice, vsuvky, dodatky, fragmentárnost, parcelace (NAPŘ. ČECHOVÁ ET AL. 2008: 311–312; HOFFMANNOVÁ ET AL. 2016: 402, 411; JELÍNEK 1995: 770). Na úrovni lexikonu je pozornost věnována neologismům a vulgarismům, argotickému vyjadřování, frazeologii, ale i výskytu cizích slov, expresivních výrazů (např. ČECHOVÁ ET AL. 2008: 321) apod.

Terminologická poznámka

V tomto textu budeme používat termíny *registr*, *žánr*, *vnětextová* a *vnitrotextová analýza*.

Vzhledem k tomu, že zmíněné pojmy jsou v multidimenzionální analýze užívány komplementárně, a proto že v české tradici je zaveden pojmový aparát odlišný, je zde stručně charakterizujeme.

Termín *registr* chápeme jako funkčně a situačně podmíněnou varietu jazyka, přičemž funkce je zde pojmána jako relace mezi složkami komunikační situace a jazykovými prostředky, jež s těmito složkami korelují (srov. CVRČEK ET AL. 2020: 19–20; MRÁZKOVÁ 2016). V české stylistické tradici tomuto termínu asi nejvíce odpovídá *funkční* či *objektivní styl*. Registr je tedy de facto výsledek analýzy konkrétních jazykových prostředků, jež prostupují celým textem; tato analýza se opírá o vnitrotextové hledisko. Naopak převážně hledisko vnětextové je uplatněno zejm. při stanovení žánru, který se vymezuje na základě konvenční, ustálené podoby textu (např. dělení do kapitol, prolog, epilog atd. budou vlastnosti očekávané u románu) a typických

jazykových prostředků objevujících se v zásadě jednou na určitém místě v textu (např. oslovení na začátku dopisu) (BIBER — CONRAD 2009: 16). V české stylistické tradici jej lze též ztotožnit s termínem *slohový útvar* (nejde tedy o pohled genologický). Vnětextová perspektiva se uplatňuje i při dělení na textové typy, což je klasifikace používaná při anotaci korpusů ČNK, která zahrnuje obecnější kategorie než žánr (např. textový typ *román* se v korpusové klasifikaci může dělit na žánry *milostný, detektivní...*; textový typ *vědecko-naučná práce* se dále podle žánru klasifikuje na jednotlivé vědecké obory — *chemie, ekonomie, lingvistika* apod.).²

3. Multidimenzionální analýza češtiny a registrový model

Abychom mohli popsat analytickou metodu, kterou jsme dospěli ke zjištěním o aktuálních posunech v podobě česky psané prózy, je nejprve třeba vysvětlit základní principy korpusové multidimenzionální analýzy (MDA), jak ji pro angličtinu navrhl a implementoval Douglas Biber (BIBER 1988, 1995; BIBER — CONRAD 2009).

3.1 Jak se dělá MDA

Biberova MDA vzniká v reakci na dosavadní popisy stylu, s jakými se setkáváme i v českých stylistických příručkách z devadesátých let. Ty jazykový styl popisují jako výsledek výběru a spojování jazykových prostředků, které jsou adekvátní komunikační situaci a zároveň naplňují komunikační potřeby původce textu. Z toho plyne několik poznatků: jednu a tutéž materii (obsah sdělení) lze popsat různě vzhledem k situaci, jednotlivé varianty textů nejsou libovolně zaměnitelné (jiná bude forma přednášky pro odborné publikum a jiná pro laickou veřejnost), a autor textu má tedy možnost volit z konkurenčních prostředků (např. obecněčeské *hysme* budeme očekávat v běžné komunikaci, zatímco *bychom* je typické pro spíše oficiální texty).³

Pro výše zmíněný přístup i pro přístup Biberův je společná základní úvaha, že jazyku je vlastní variabilita a že jazykové prostředky uživatelé při vzniku textů neužívají náhodně, ale v souladu s komunikační funkcí a adekvátně dané komunikační situaci. Primát Biberova přístupu tkví v tom, že jako výchozí disko nezvolil právě situační rozdíl mezi texty (např. mluvený mód jazyka je obecně spojen s větší interaktivitou, ale jednodušší nominální syntaxí než mód psaný atp.), ale jazykové prostředky (rysy), které jsou dle existujících výzkumů schopny signalizovat konkrétní komunikační situaci. Jako příklad

2 Přehledný popis textové klasifikace používaný pro synchronní psané korpusy je k dispozici zde: https://wiki.korpus.cz/doku.php/cnk:klasifikace_textu_syn2015.

3 Srov. s pojmem *konkurenční množiny*, které Jelínek definuje jako „seskupení takových prostředků, které realizují základní komunikační záměr původce jazykového projevu, a mohou se proto zaměřovat“ (JELÍNEK 1995: 732).

nám může posloužit třeba přechodník, jehož výskyt v textu obvykle naznačuje odbornost, formálnost apod. (odhlížíme — zde i u jiných rysů — od různých atypických, třeba ironických použití, která jsou ve svém úhrnu málo frekventovaná).

Takovéto funkčně využívané jazykové rysy Biber zkoumal z hlediska jejich provázanosti, tedy jak pravděpodobné je, že když původce textu užije jazykový rys A, použije zároveň i rys B. Pravidelné souvýskyty jazykových rysů lze identifikovat statistickými metodami a následně jsou interpretovány z hlediska společné funkce. Tato interpretace se opírá o text (či jeho úryvek) jako výsledný produkt a charakterizuje jej právě z hlediska užitých jazykových prostředků (jedná se tedy o hledisko vnitrotextové). Liší se tak od tradičních popisů, jež text sledují de facto při jeho vzniku, přičemž výběr konkrétních jazykových prostředků je do jisté míry záměrný a cíleně vedený právě s ohledem na vnější faktory (několikrát zmíněné funkce a situace) (např. MINÁŘOVÁ 2017).

Ke komplexnosti analýzy zásadní měrou přispívá také korpusová metoda, opřená o široké spektrum textů; výsledkem analýzy tak není pohled zkreslený malým vzorkem, kde klíčovou roli hraje to, jaké texty byly vybrány k analýze a porovnání. Značná textová šíře zároveň umožňuje určit více charakteristik podložených souvýskytem jazykových rysů — tyto charakteristiky Biber nazývá *dimenze variability* a jednotlivé texty se na těchto dimenzích umísťují podle toho, do jaké míry jsou v nich určité souvýskyty jazykových prostředků obsaženy.

Výsledky této studie vycházejí z modelu MDA, jenž byl vytvořen pro češtinu (CVRČEK ET AL. 2018, 2020, 2021). Stejně jako v případě analýz provedených D. Biberem jsme i v „české“ MDA dodrželi základní postup, sestávající z vytvoření korpusu, seznamu relevantních jazykových rysů a jejich operacionalizace, z aplikace faktorové analýzy a následné interpretace. Jednotlivé body nyní popíšeme podrobněji.

3.1.1 SESTAVENÍ KORPUSU KODITEX

Pro účely vytvoření českého modelu registrové variability byl sestaven specializovaný reprezentativní asynchronní korpus Koditex (ZASINA ET AL. 2018), obsahující 9 milionů textových slov (není započítána interpunkce, i s ní se jedná o korpus o velikosti více než 10 milionů slov). Jeho název je zkratkovým slovem pro **korpus diverzifikovaných textů**, zároveň odkazuje k osobě Viléma Kodýtky, který se jako první pokusil MDA aplikovat na češtinu (KODÝTEK 2007). Základním požadavkem při sestavování korpusu bylo co nejpestřejší složení (ve smyslu vnětřtextovém), jež by reflektovalo variabilitu současné češtiny, a to ve třech komunikačních módech: psaném, mluveném a internetovém. Módy jsou uspořádány hierarchicky a dělí se do dvou a více divizí (*division*; např. v rámci psaného módu se jedná o beletrii, oborovou literaturu, noviny a časopisy a soukromou komunikaci). Divize se člení do 45 tříd (*class*; například sci-fi literatura), s cílovou délkou přibližně 200 000 slov na třídu

(v závislosti na dostupnosti dat). Pro psaný mód byl ještě zaveden mezistupeň tzv. nadtrždy (*superclass*), která seskupuje některé třídy textů (podrobnosti o složení viz ZASINA — KOMRSKOVÁ 2019).

Text z každého módu vstupoval do korpusu ne jako celek (jak je běžné např. pro korpusy řady SYN), ale jako kratší souvislý textový úsek; hranice vět přitom byly zachovány. Tím bylo dosaženo velké pestrosti jak z hlediska textů samotných, tak i jejich autorů a s tím související rozmanitosti žánrů a registrů. Důležitou úvahou při designu korpusu byla velikost textového úseku, který by do korpusu měl vstupit. Primárně jsme usilovali o to, zahrnout do korpusu úryvky o přibližně stejné délce, arbitrárně byla zvolena velikost v rozmezí dva až pět tisíc slov. To se však ukázalo jako nerealistické vzhledem k tomu, že některé typy textů požadované velikosti nedosahovaly. Týkalo se to především soukromé korespondence a administrativních textů, jež jsou obecně kratší. V těchto případech jsme se rozhodli snížit počet slov na 1 000. U facebookových postů, které se skládaly z velkého množství fragmentů kratších než tisíc slov, jsme zvolili řešení sloučení textů podle autora a času a až pak texty vzorkovali.

Vedle rozmanitosti korpusu posloužila výzkumnému záměru i bohatá anotační vrstva, která zjednodušila operacionalizace rysů. Všechny texty byly lemmatizovány⁴ a lingvisticky značkovány (JELÍNEK 2008; SPOUSTOVÁ ET AL. 2007; STRAKOVÁ ET AL. 2013, 2014), byly v nich identifikovány a označeny frazémy, rozpoznány pojmenované entity, tedy výrazy či jejich posloupnosti, jež označují jména osob, míst, výrobků apod.

Zdrojem textů v psaném módu je převážně korpus SYN2015 (KŘEN ET AL. 2016), v některých případech bylo kvůli nedostatku textů nutno využít i jiné korpusy řady SYN. Dopisy byly do korpusu převzaty z Korpusu soukromé korespondence (HLADKÁ 2006). V mluveném módu je zdrojem neformální komunikace korpus ORAL2013 (BENEŠOVÁ ET AL. 2013), elicitované komunikace formální části korpusů PMK (ČERMÁK ET AL. 2001) a BMK (HLADKÁ 2002), doklady mediální komunikace pocházejí z korpusu DIALOG (KADERKA — PETREK 2012), připravené a převážně čtené projevy z korpusu SPEECHES (CVRČEK ET AL. 2015) a z české části zápisů v korpusu EUROPARL. Data pro mód internetové komunikace byla vybrána z Korpusu české Wikipedie, jenž vznikl v Centru zpracování přirozeného jazyka v Brně, blogy z korpusu Araneum Bohemicum Maius (BENKO 2014), texty pro skupinu příspěvků z diskuzních fór, internetových diskuzí a veřejných statusů ze sociální sítě Facebook poskytl Josef Šlerka a společnost Socialinsider.

4 Lemmatizace je jedna ze standardních fází automatického zpracování textu při vytváření korpusu. Je při ní každému slovu v textu přiděleno lemma, tj. základní slovníkový tvar, např. lemma *kočka* je přiřazeno všem tvarům tohoto slova (*kočka, kočkou, kočkami, kočko* ad.).

Většina textů (s pokrytím 76 % všech tokenů) zahrnutých v korpusu představuje české originály.

3.1.2 JAZYKOVÉ RYSY

Dalším krokem je sestavení seznamu relevantních jazykových rysů, u nichž lze předpokládat, že budou v textu plnit nějakou funkci a budou se podílet na variabilitě. Můžeme například očekávat, že protetické *v-* se častěji objeví v každodenní neformální komunikaci nebo v textech, které ji napodobují, zatímco vyšší frekvence v užívání temporálních či lokálních adverbii může být mimo jiné motivována snahou o snadnější orientaci v textu a jeho strukturaci, jež jsou typické pro texty odborné. Protože je ale dopředu těžko odhadnutelné, jakou měrou se konkrétní rysy budou na této variabilitě podílet, výsledný seznam může obsahovat de facto libovolné jazykové charakteristiky.

Při práci na výchozím soupisu rysů jsme čerpali z lingvistických prací českých i zahraničních (např. BIBER 1988, 1995; CVRČEK ET AL. 2010; HOFFMANOVÁ ET AL. 2016; KARLÍK ET AL. 1995; KODÝTEK 2008), ale i z vlastní intuice. Výsledkem bylo celkem 137 jazykových rysů vztahujících se k 7 jazykovým rovinám (zevrubný popis včetně operacionalizace rysů a určení jejich chybovosti viz CVRČEK ET AL. 2020: 33–78): rovina fonologická (např. rys diftongizace *y > ej* v koncovce i kmeni slova), rovina morfologická (např. rys frekvence genitivu, rys slovesná osoba), rovina slovtvorná (např. rys deminutiva), rovina lexikální (např. rys inventář spojek, rys výplňková slova), rovina syntaktická (např. rys pasivní tvary sloves, rys řetězce substantiv), rovina textová, kam lze řadit i rysy prostupující několik rovin (např. rys frazémy, rys tematická koncentrace).⁵ Následně byly všechny jazykové rysy operacionalizovány, tedy převedeny na korpusový dotaz, a to tak, aby zvolený dotaz vyhledal v korpusu správně co nejvíce relevantních dokladů. Pro kontrolu byly dané operacionalizace testovány a byla určena míra jejich chybovosti; na jejím základě byly některé rysy ze seznamu odstraněny nebo byla opravena jejich operacionalizace. Výsledný počet rysů, které vstupovaly do statistické faktorové analýzy, byl 122.

3.1.3 STATISTICKÉ VYHODNOCENÍ POMOCÍ FAKTOROVÉ ANALÝZY

Princip MDA je založen na předpokladu, že na interpretaci textu je možné nahlížet z vnitrotextové perspektivy, tedy charakterizovat jej na základě společného výskytu konkrétních jazykových prostředků, jež jsou v něm užity. Vstupní seznam rysů obsahuje více než 120 položek (viz výše), které se ovšem

5 Tematická koncentrace je index používaný v kvantitativní lingvistice na měření tematické kompaktnosti, resp. diverzity (ČECH 2016). Měření tematické koncentrace spočívá ve zjištění, kolik plnovýznamových slov se ve frekvenčním seznamu všech jednotek textu nachází nad tzv. h-bodem, tedy v oblasti nejfrekventovanějších slov, kde se vyskytují převážně slova gramatická.

nevyskytují nezávisle, jejich užití je provázané, a je proto účelné jejich počet pro přehlednost zredukovat.

Faktorová analýza (VOLÍN 2007: 307) umožňuje různorodá data na vstupu (v našem případě se jedná o frekvenci jednotlivých jazykových rysů v textech) zastoupit několika málo dimenzemi, jež jsou zjednodušeně řečeno tvořeny těmi rysy, které spolu výrazně korelují. Vezměme si tento ilustrativní příklad: pokud se v textu s vysokou frekvencí objevuje rys I. osoba plurálu, mohli bychom předpokládat, že se v něm zároveň objeví častěji souřadící spojka *a* či juxtapozice. Užití těchto rysů bývá typické pro popis instrukce, respektive návodu; tato funkce textu se stává tzv. latentním faktorem. Faktorová analýza uvedený předpoklad dokáže empiricky identifikovat (vyčíslit korelaci mezi zmíněnými prostředky), objevit další jazykové rysy, jež s těmito dvěma souvisejí (i když bychom jejich souvyskyt třeba neočekávali), a tyto jednotlivé rysy nahradit jednou dimenzí. Tu pak lingvista interpretuje, v našem případě by se tedy mohlo jednat třeba o návod, instrukci.

Na základě 122 jazykových rysů jsme zobecnili po nejrůznějších statistických testováních celkem 8 dimenzí variability, jejichž charakteristiku zmíníme v následujícím oddílu. Při popisu jednotlivých jazykových prostředků, jež spolu korelují a které vytvářejí jednotlivé dimenze variability, budeme užívat termín *loading*, jenž je vyčíslením příslušnosti rysu k dané dimenzi. Více informací o statistickém testování viz CVRČEK ET AL. 2020: 78–90.

3.2 Dimenze variability v češtině

Jednotlivé dimenze variability byly ustaveny na základě toho, jaké jazykové rysy v nich prokázaly nejvyšší loading, tedy největší příslušnost k danému latentnímu faktoru. Loadiny rysů mohly nabývat jak pozitivních, tak negativních hodnot — vytvořily tak opačné extrémy, jejichž interpretace stály za názvem dané dimenze. Pozitivní loading rysu přitom znamená, že čím častěji se rys v textu objevuje, tím výrazněji se posouvá směrem k pozitivnímu extrému dané dimenze. Při interpretaci jsme se vedle empirických hodnot naměřených u jednotlivých rysů opírali i o konkrétní texty, jež při faktorové analýze dosáhly v dané dimenzi extrémních hodnot; využili jsme tedy pro podporu charakteristiky dané dimenze i hledisko vnětextové. Zevrubný popis dimenzí (tzv. multidimenzionálního prostoru) společně s ukázkami textů viz CVRČEK ET AL. 2020: 91–105.

Dimenze I — dynamický (+) × statický (–): v této dimenzi nejvyšších kladných hodnot dosahovaly slovesné rysy (konkrétně minulý čas, obecně slovesa, verba finita apod.), na opačném pólu, tedy záporných extrémů, pak rysy jmenné (například substantivní přívlastky neshodné postponované, adjektiva, abstrakta). Při interpretaci jsme s verbálností, tj. nadprůměrnou frekvencí rysů souvisejících se slovesy, spojili právě charakteristiku dynamičnosti oproti staticčnosti.

Nejvyšších hodnot dosahují romány s převahou narativní složky (romantické či kriminální), verbálnost zde ale není pouze otázkou narace, nýbrž i mluvenosti či kontemplance (např. slovesný tvar *myslím*). Na opačném (statickém) pólu pak stojí texty administrativní povahy. S určitou mírou zjednodušení bychom mohli shrnout, že tato dimenze rozlišuje obecně vyprávěcí a popisný slohový postup.

Dimenze 2 — spontánní (+) × připravený (–): pro druhou dimenzi je typický výskyt kontaktních výrazů a výplňkových slov, ukazovacích zájmen či citoslovcí (na kladném pólu), na druhé straně pak přítomnost předložkových pádů, vět s interogativními a vztažnými adverbii nebo předložek. Obecně je rys spontánnosti připisován každodenní, neformální komunikaci. To podporuje i fakt, že z hlediska textových typů vyniká na kladném pólu škály interaktivní mluvený jazyk. Do této dimenze by tedy nejspíše patřily i rysy tradičně chápané jako charakteristiky uměleckého stylu (viz odd. 2), ale její rozsah je širší (jak je patrné, nejde jen o obecnou češtinu, ale i o rysy kontaktní a různé dysfluency, opakování slov, deixi apod.). Na záporném konci škály stojí texty administrativní a odborná literatura, tedy texty, kde jsou příprava a pregnanost vyjádření velmi důležité.

Dimenze 3 — vyšší (+) × nižší (–) stupeň koheze: třetí dimenze vykazuje nejvýraznější rozdíly ve výskytu spojovacích a odkazovacích prostředků. Nejvyšších kladných loadingů dosahují rysy korelativa typu *tím, že*, jmenný přísudek substantivní či vztažné věty typu *kteřý*. Z textů se zde shlukují především čtené projevy, dále texty z oboru společenských věd, texty vědecké a popularizační a televizní diskuzní pořady. Naopak záporné hodnoty vykazují pouze číslovky; negativní pozice na této dimenzi dosahují např. encyklopedické texty, běžná neformální komunikace či soukromá korespondence. Hlavní rozdíl mezi oběma konci škály spatřujeme v tom, zda je pro daný text charakteristický důraz na informaci, nebo na usouvztážnění těchto informací. Texty koheznější jsou složitěji strukturované, provázané, naopak např. u encyklopedií je primárně kladen důraz na fakta a jejich syntaktické uspořádání je většinou jednoduché, juxtaponované.

Dimenze 4 — polytematický (+) × monotematický (–): dimenze číslo 4 reflektuje charakteristiky související s bohatostí lexikonu, což je v mnoha případech spojeno se zapojením většího počtu témat. Na kladném pólu dimenze oscilují rysy jako velikost inventáře lemmat (lexémů), bigramů (dvojice slov), frekvence toponym, inventář předložek či zájmen. Se zápornými hodnotami se vyskytují tematická koncentrace, Yuleův koeficient,⁶ verbální substantiva

6 Index vyjadřující míru repetitivnosti, tedy pravděpodobnost, že ve dvou náhodných pokusech vyberu z textu stejná slova.

či slovesný rod (pasivum). Na kladném pólu škály se objevují texty novinové a časopisecké, tematicky různorodé (od kultury přes vědu a techniku až po bulvár), a tedy lexikálně pestré; na záporném pak jde o administrativu a odborné texty, zaměřené na jedno téma a s obecnou platností.

Dimenze 5 — vyšší (+) × nižší (-) míra explicitní adresnosti: v páté dimenzi nejvíce korelují s kladným pólem tyto jazykové rysy: otázky, 2. slovesná osoba, otázky doplňovací, zájmena pro 2. slovesnou osobu, vokativ apod. Společným jmenovatelem zmíněných rysů je přítomnost komunikačního partnera. Naopak na záporném pólu škály relevantních hodnot dosáhly pouze rysy průměrná délka věty a výskyt frekventovaných n-gramů (ustálených sekvencí slov). Nejvyšších hodnot v 5. dimenzi dosahují dramata, poezie, sci-fi či detektivky. Jak se tedy ukazuje, mluvený mód, který bychom zde očekávali primárně, v této dimenzi vysokých skóre nedosáhl. Interaktivnost, explicitní obracení se ke komunikačnímu partnerovi je naopak typické pro beletrii, nejvíce pro dramata. Stereotypní vnímání mluveného jazyka tak nemusí odpovídat skutečnosti. Například vysvětlení nižší frekvence vokativu v mluvené komunikaci může spočívat v tom, že její účastníci jsou v jeden čas na jednom místě, a tyto kontaktové či adresné prostředky tak k úspěšnému průběhu dialogu nepotřebují. Pátá dimenze tedy vyděluje texty na jedné straně dialogické, na straně druhé spíše monologické (převažují texty z Wikipedie).

Dimenze 6 — obecný (+) × konkrétní (-): kladných korelací dosahují v šesté dimenzi tyto jazykové rysy: koordinace, sémanticky vyprázdněná adjektiva a spojky. Na záporné straně se pak výrazněji objevují antroponyma, numeralia, časové výrazy, shluky substantiv či toponyma. Opora o vnětextovou charakteristiku není v tomto případě příliš velká, na obou pólech se jedná o heterogenní texty z různých oblastí bez zjevného společného jmenovatele. Na základě samotných jazykových rysů ale lze říci, že šestá dimenze je spojena s obecnou, nadčasovou platností informací (kladný pól) a na druhé straně s jejich konkrétností, situační zakotveností (vlastní jména, časové výrazy) a proměnlivostí.

Dimenze 7 — prospektivní (+) × retrospektivní (-): v sedmé dimenzi se na obou stranách škály objevují slovesné rysy, jako například pro kladný pól slovesný čas přítomný a budoucí, jmenný přísudek adjektivní, imperativ či 2. slovesná osoba. Na záporné straně korelují s touto dimenzí nejvíce slovesný čas minulý, zájmena pro 3. osobu, přivlastňovací adjektiva ad. Pokud jde o kategorii textů, nejvýrazněji se na záporném pólu vyskytují beletristické texty s převažujícím retrospektivním vyjadřováním. Tato dimenze je významná, protože odlišuje dynamické retrospektivní texty (naraci) od

ostatních dynamických (a zároveň přispívá k charakterizaci beletrie jako vyprávění).

Dimenze 8 — postojovost (+) × faktualnost (–): osmá dimenze vyděluje texty vyjadřující postojovost (subjektivnější) od faktualních (objektivní). Pozitivně korelují rysy částice oslabující význam, restriktory, částice zesilující význam, částice členící text či adverbia. Na záporném pólu najdeme pouze rys koordinace. Z textů se výrazně profilují (na kladné straně) internetové diskuze, fóra, blogy, tedy útvary, kde je vyjádření názoru hlavním důvodem, proč texty vůbec vznikly. Zápornou část spektra obsadily texty administrativní či webová encyklopedie Wikipedie, u nichž je důraz na fakta primární.

Dimenze modelu nejsou stejně významné (jejich pořadí je arbitrární). Nejvyšší podíl variability vysvětlují dimenze 1 a 2, dále následují dimenze 5 a 8, které charakterizují texty v těch nejobecnějších rysech, zbylé dimenze zachycují specifičtější vlastnosti textů.

V grafech v obr. 1 jsou jednotlivé texty korpusu Koditex znázorněny tečkami, beletristická část (tj. nejen próza) je zvýrazněna černou elipsou. Jejich pozice na jednotlivých dimenzích značí míru vlastnosti, které jsou dimenzemi reprezentované.

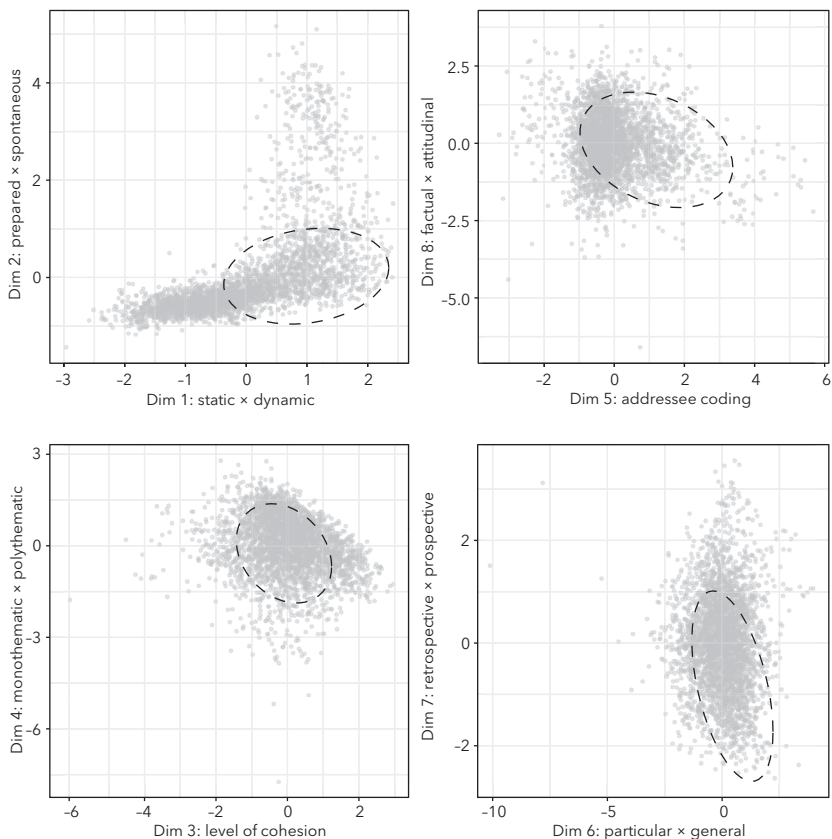
Z pozice beletristické části korpusu Koditex v obr. 1 můžeme dovodit, že beletrie je obecně tvořena texty dynamickými (dimenze 1), které vykazují charakter spíše připraveného komunikátu (dimenze 2), je relativně hodně adresná (dimenze 5) a značně retrospektivní (dimenze 7). V ostatních dimenzích je beletrie nevyhraněná.

3.3 *Registry*

Jak je zřejmé z obr. 1, texty v multidimenzionálním (MD) prostoru nejsou rozptýleny rovnoměrně, vyskytují se ve shlucích. Tyto shluky (klastry) pak představují typické konfigurace vlastností zastupovaných dimenzemi. Na jejich identifikaci je založeno vymezení registrů,⁷ kterých na základě dat v Koditexu identifikujeme 10 (podle první, nejvýznamnější dimenze se dělí na pět dynamických a pět statických):

1. analýza: statický monotematický registr
2. popularizace: statický polytematický obecný registr
3. žurnalistika: statický nevyhraněný registr
4. fakta: statický polytematický konkrétní registr
5. argumentace: statický kohezní registr

7 K vymezení registrů a jejich detailnímu popisu viz CVRČEK ET AL. 2020: 118n.



Obr. 1. Pozice beletristických textů v rámci korpusu Koditex na 8 dimenzích MD prostoru.

6. anketa: dynamický neadresný registr
7. konverzace: dynamický spontánní registr
8. komentář: dynamický postojový registr
9. scénář: dynamický adresný registr
10. narace: dynamický retrospektivní registr

3.4 Projekce textů na MD model

Vzorek prózy z posledních třiceti let budeme analyzovat na podkladě právě představeného obecného modelu. Každý z textů, který detailně představíme v následujícím oddílu, bude podroben stejné analýze jako texty

Koditexu, abychom zjistili jeho pozici na každé z 8 dimenzí. Následně pak budeme sledovat, jestli dochází v jednotlivých dimenzích k diachronním posunům.

4. Popis vzorku

Při konstruování vzorku jsme vycházeli z dat dostupných v největším korpusu psané češtiny SYNv9 (KŘEN ET AL. 2021). Prvotní výběr byl vymezen těmito podmínkami:

- díla, která byla poprvé vydána po roce 1990;
- textové typy NOV: próza (typicky romány) a COL: kratší próza (typicky soubory povídek);
- pouze texty určené obecnému publiku, tedy nikoli texty označené jako JUN, tj. pro děti a mládež;
- texty, které vyšly knižně, nikoli časopisecké texty;
- texty s více než 10 tisíci slovy.

V druhém kole výběru byl jako základní postulován požadavek, aby byl ve výsledném vzorku každý autor zastoupen maximálně jedním dílem. Texty jednoho autora, které splňovaly kritéria prvovýběru, byly následně vybírány podle těchto zásad:

- pokud byla k dispozici díla NOV i COL, dávali jsme přednost NOV;
- pokud bylo k dispozici více děl NOV, dávali jsme přednost nejstaršímu;
- pokud bylo takových děl víc, zvolili jsme text, který byl první v abecedním pořadí.

Takto jsme dospěli ke vzorku, který obsahoval 1 070 titulů (jejich rozložení v letech shrnuje tab. 1).

Zastoupení titulů podle textového typu (txtype) a podle překladovosti shrnuje tab. 2. Z hlediska zdrojových jazyků je podle očekávání nejvíce zastoupena angličtina (500 titulů), s velkým odstupem následují němčina a francouzština (po 63 titulech) a švédština (36 titulů).

Vzhledem k tomu, že model, na nějž budeme prozaické texty z našeho vzorku projektovat, je kalibrován na úseky textu o rozsahu 2–5 tisíc slov, nebylo by korektní, kdyby do analýzy vstupovaly vybrané tituly v celé své délce. Z každého dostupného titulu proto byly vybrány 3 souvislé úryvky reprezentující jeho začátek, prostředek a konec. Algoritmus vytváření těchto úseků nejprve rozdělil každý titul na třetiny a pak z každé třetiny vybral část, která obsahovala 2 000–5 000 slov (s preferencí okolo 4 000). Při dělení byly respektovány hranice odstavců, pokud takové dělení nebylo možné provést, byly alespoň respektovány hranice vět.

Tab. 1. Počet titulů ve vzorku v jednotlivých letech.

Rok	Počet titulů	Rok	Počet titulů
1992	22	2006	35
1993	36	2007	48
1994	24	2008	40
1995	30	2009	34
1996	37	2010	42
1997	43	2011	22
1998	25	2012	51
1999	42	2013	69
2000	35	2014	68
2001	41	2015	38
2002	37	2016	50
2003	32	2017	62
2004	41	2018	30
2005	35	Celkem	1070

Tab. 2. Podíl titulů v rámci vzorku v závislosti na textovém typu a překládovosti.

Txtyp	Překládovost		Celkem
	originál	překlad	
COL	85	119	204
NOV	188	678	866
Celkem	273	797	1070

Ve výsledku jsme tedy dospěli ke vzorku, který v 3 210 úryvcích z 1 070 titulů obsahuje 12,657 milionů slov. Průměrná délka úseku dosahovala 3 942 slov, medián byl 3 969 slov, přičemž mezi úryvky ze začátku, prostředku a konce textu nejsou žádné signifikantní odchylky z hlediska jejich délky (podle párového Wilcoxonova testu, což je test schopný určit, zda mají dva soubory statisticky signifikantně odlišné mediány).

5. Výsledky

Každý text ve vzorku prózy byl podroben stejné analýze jako texty v korpusu Koditex, na němž byl vybudován registrový MD model. Konkrétně to znamená, že v každém textu bylo identifikováno stejných 122 jazykových rysů a z jejich hodnot pak byla vypočítána poloha na každé z 8 dimenzí modelu. V následujících oddílech budeme sledovat, jaké charakteristiky, které vyplýva-

Tab. 3. Příslušnost textů v beletristickém vzorku k jednotlivým registrům.

Registr	Všechny texty		Podle pozice			Podle textového typu	
	Počet textů	v %	Začátek (v %)	Prostředek (v %)	Konec (v %)	COL (v %)	NOV (v %)
Narace	2574	80,19	77,85	82,62	80,09	76,8	80,99
Žurnalistika	157	4,89	6,73	3,93	4,02	6,05	4,62
Komentář	147	4,58	5,7	4,11	3,93	5,07	4,46
Scénář	128	3,99	2,43	4,21	5,33	4,41	3,89
Argumentace	71	2,21	3,18	1,96	1,5	2,94	2,04
Fakta	70	2,18	2,06	1,31	3,18	2,12	2,19
Popularizace	52	1,62	1,78	1,4	1,68	1,63	1,62
Konverzace	11	0,34	0,28	0,47	0,28	0,98	0,19

jí z pozice v MD prostoru, prozaické texty vykazují a zda lze z jejich posunů v průběhu let vysoudit nějaký převažující vývojový trend.

5.1 Převažující registry

Nejprve se ovšem zaměříme na příslušnost prozaických textů k registrům popsaným výše (viz 3.3). Pozice každého z textů v osmidimenzionálním prostoru dává možnost spočítat vzdálenost od pomyslného středu jednotlivých registrů. S přijatelnou mírou zjednodušení lze konstatovat, že text je z hlediska zapojení konkrétních jazykových prostředků nejpodobnější tomu registru, k němuž má nejbliž. Ačkoli se v našem vzorku jedná o texty prozaické, je očekávatelné, že budou obsahovat pasáže různého typu. Podle výsledků v tabulce 3 se nepřekvapivě na nejvyšší příčce umísťuje narace (u 80 % úseků):

Ve vzorku se naopak z registrů vůbec neobjevují analýza a anketa. Rozložení registrů ve vzorku se mírně liší v závislosti na textovém typu — úryvky pocházející z textů označených jako COL inklinují mírně k argumentativnosti a žurnalistickému registru na úkor narativity, ale přes statistickou signifikanci (χ^2 test, $p = 0.04$) rozdíl není výrazný (srov. s 5.3).

Naproti tomu umístění úryvku ve výchozím textu má poměrně výrazný a statisticky signifikantní vliv na rozložení registrů. Citelný je zejména úbytek argumentace, komentářů (tj. polemických registrů) a žurnalistického registru ve srovnání počátečních, prostředních a koncových úryvků, pokles zastoupení faktografického registru ve středních úsecích (nová fakta se v textech objevují buď na začátku, nebo na konci) a naopak nárůst úseků se scénářem jako převažujícím registrem (srov. s 5.5).

5.2 Posuny v jednotlivých dimenzích

Diachronní analýzu proměn registrů česky psané prózy v posledních 30 letech zahájíme sumarizujícím grafem v obr. 2. Pro každou dimenzi je v něm zobrazen vývoj od roku 1992 do roku 2018 tenkou linií. Její hodnota pro každý rok je vypočítána jako medián hodnot vzorků v dané dimenzi. Nadto je tlustší čarou naznačen převažující trend (vypočítaný pomocí metody lokální polynommické regrese) spolu s konfidenčním intervalem (šedá oblast), jehož šíře ukazuje na spolehlivost odhadu. V interpretaci se budeme soustředit převážně na trend, který umožňuje odhlédnout od výkyvů v konkrétních letech.

Pro korektní interpretaci grafů v obr. 2 je dobré připomenout, že měřítko vísle osy v grafech není shodné a u každé dimenze zaujímá jiný rozsah hodnot. Ve všech dimenzích tak můžeme sledovat nějaký pohyb, ne u všech lze ale konstatovat zásadní vývojový trend. Kupř. v dimenzi 2 — ačkoli zde vidíme z roku na rok různé posuny — se počáteční a koncový stav zásadněji neliší.

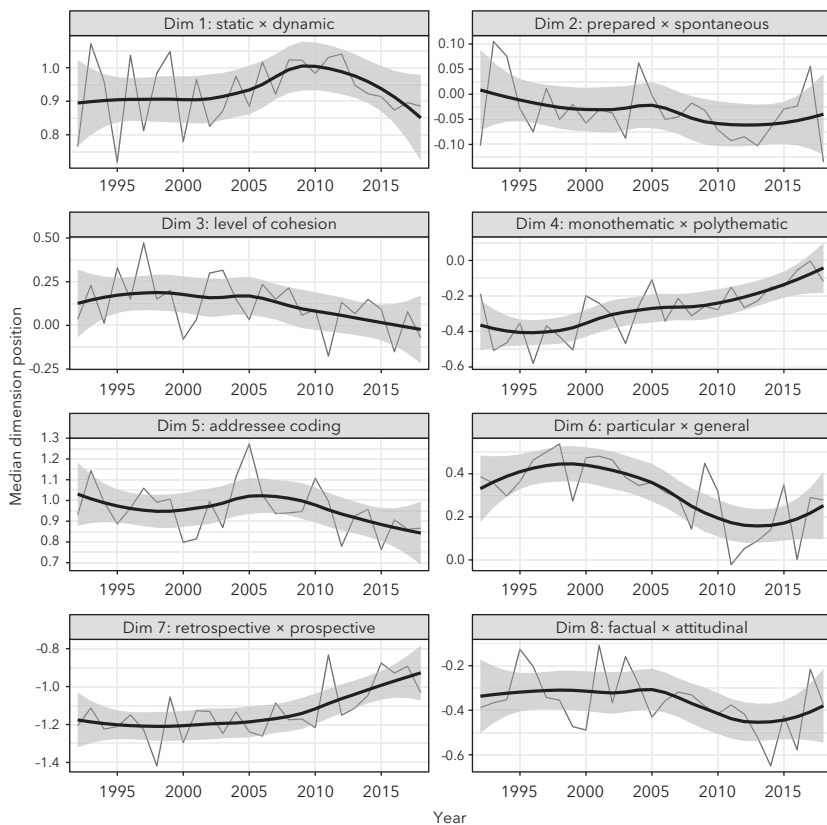
V každé dimenzi jsme srovnali hodnoty pro počáteční a koncové období reprezentované vždy dvěma roky (tedy průměr let 1992 a 1993 versus průměr let 2017 a 2018). Ze srovnání pomocí Wilcoxonova testu a po aplikaci statistické korekce pro násobné testování (metodou FDR) plyne, že statisticky signifikantní rozdíl mezi počátkem a koncem sledovaného období nacházíme u dimenzí 1, 3, 4, 5 a 7. Tento výsledek jsme doplnili informací o relevanci rozdílu počátečních a koncových let.⁸ Z těchto výsledků plyne, že nejvyšší odlišnost najdeme v dimenzích 3, 4, 5, 7 a 8.

Z hlediska signifikance i relevance rozdílu se jako nejvýznamnější ukazují a hodné bližšího prozkoumání jeví tedy dimenze 3, 4, 5 a 7. Zároveň můžeme konstatovat, že ve sledovaném období nedochází k výraznější změně v míře dynamičnosti (dimenze 1), spontánnosti (dimenze 2), konkrétnosti (dimenze 6) ani faktualnosti (dimenze 8).

V **dimenzi 3** (odlišující texty s **vyšší × nižší mírou koheze**) dochází k viditelnému poklesu hodnot v průběhu času (viz obr. 2), což ukazuje na snižování míry koheze v textech. Příkladem velmi kohezního textu obsahujícího množství rysů charakteristických pro tuto dimenzi je následující ukázka z románu *Kalibán* (první vydání v ČR z roku 1997):

Už jsem řekla, že každý lidský vynález je odrazem doby, kdy vznikl. Je-li tomu tak, jsou tři zákony odrazem velmi temného zrcadla. Odrážejí čas, kdy se lidé strojů obávali a nevěřili jim, kdy byla technologie celkem správ-

8 Abychom zajistili srovnatelnost, bylo třeba hodnoty v dimenzích nejprve normalizovat. K tomu jsme využili data z korpusu Koditex (medián a mezikvartilové rozpětí na každé dimenzi), s nimiž jsme hodnoty prozaických textů srovnávali. Hodnoty pro počátek (roky 1992 a 1993) a konec sledovaného období (roky 2017 a 2018) jsme odečetli od mediánu dané dimenze a vydělili jejím mezikvartilovým rozpětím, což



Obr. 2. Trendy prozaických textů v období 1992–2018 na jednotlivých dimenzích.

ně vnímána jako často zlovorná, kdy to, co se získalo za pomoci strojů, šlo na úkor lidí, kdy dokonce i ten nejbohatší člověk byl podle našich dnešních měřítek chudý a kdy chudí hluboce — a pochopitelně — nenáviděli bohaté. Řekla jsem a ještě řeknu mnoho negativních věcí o naší kultuře založené na robotech, ale ona má i mnoho jasných a zářivých kladů. Ztratili jsme nejen fakt chudoby, ale i schopnost představit si ji.
(ASIMOV — MACBRIDE 1997)

je postup napodobující převod na z-skóre, tedy hodnotu, která je nezávislá na absolutním průměru v dané dimenzi a která zohledňuje i její obvyklý rozptyl. Tímto způsobem získáváme rozdíl mezi počátečním a koncovým stavem, který je mezi dimenzemi srovnatelný, ačkoli jsou jejich absolutní hladiny i disperze odlišné.

Protipól, tedy text z konce sledovaného období a z opačného extrému na ose vyšší × nižší míra koheze, reprezentuje úryvek z povídkové knihy Petra Šámala *Záchranáři* z roku 2016, která obsahuje výrazné prvky literatury faktu:

Původně jsem přijela s určitým cílem, i když mám berle, mám jen 8 Kč + hlad + stres = ironie osudu. Upadla jsem a trochu si narazila levou ruku, nevolejte mě, vypínám mobil.

9. srpna

22.29 hod. Od 19.30 jsem tady u vás v Krkonoších, spala jsem na železničním mostě, tady chci dostat zápal plic či uštknutí.

22.32 hod. Přijela jsem sem, aby se mně něco stalo, nereaguju na zvonění mobilu, nemám peníze na cestu zpět + nemám právo žít.

23.57 hod. Nemohu přijmout hovor, mám už skoro vyb. bat. v mobilu, málo vydrží, jdu na Špindlerovku po zelené.

00.05 hod. Je to tu špatné, po pravé ruce je to strmé a teče tam potok, nevidím ani na 30 cm.

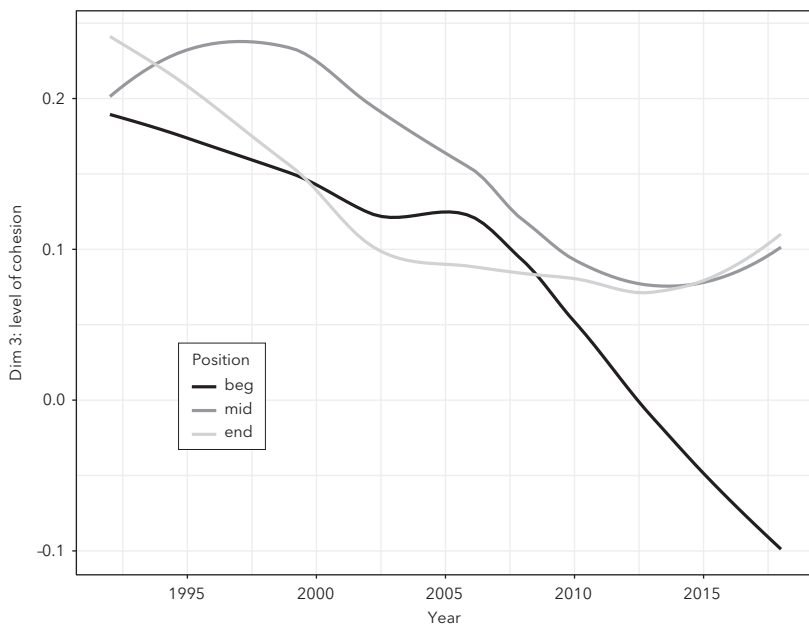
00.48 hod. Sedím na nějaké lavici asi 500m od Dívčí lávky (po trase autobusu), vpravo podle pramene.

01.34 hod. Už mám vybitej mobil (baterku), sedím tady, je mně docela zima, nemohu jít, nevidím na cestu + mám 2 berle.

(ŠÁMAL 2016)

Namísto propracovaného systému vztahů mezi větami a větnými členy, který je charakteristický pro první ukázkou, nacházíme v druhé prostý výčet událostí, často bez explicitně vyjádřeného syntaktického vztahu. Strohá výčtovost působí akčnějším dojmem rychlého střídání obrazů, dynamicky, zároveň úsečně, jakoby stroze.

Při vyšetřování trendů v jednotlivých dimenzích můžeme spekulovat o tom, že vývojové tendence jsou zesilovány (nebo naopak maskovány) spolupůsobením dalších faktorů (jako je textový typ NOV × COL, překladovost, případně pozice úryvku v původním textu). Tyto hypotézy můžeme ověřovat pomocí tzv. smíšených modelů (Mixed-effect models). V těchto modelech se snažíme závislou proměnnou (pozici textu v dimenzi) odhadovat pomocí fixních a náhodných (random) efektů. Těmi fixními jsou typicky charakteristiky, u nichž předpokládáme systémový vliv na závislou proměnnou (v našem případě to bude třeba textový typ, pozice textu či překladovost), náhodné (random) efekty pak představují nahodilou variaci (typicky vliv autoratextu nebo specifikum textu samotného). V případě třetí dimenze díky takovému modelování můžeme konstatovat, že pokles v čase se projevuje výrazněji u úryvků pocházejících ze začátku textů (viz



Obr. 3. Vývoj hodnot v dimenzi 3 (vyšší vs. nižší míra koheze) pro úryvky z počátku, prostředku a konce původního textu (beg = začátek, mid = prostředek, end = konec).

obr. 3). Můžeme spekulovat, že prozaický registr se vyvíjí z hlediska koheze radikálněji v úvodních částech textů (než v částech prostředních či koncových), v nichž se stává uzuální spíše nižší úroveň koheze, což může souviset s horizontálním členěním textu, tedy s expozicí postav a situací, jež budou v díle dále rozpracovány.

V **dimenzi 4** (texty **polytematické × monotematické**) je patrný poměrně strmý nárůst značící větší polytematickost textů. Jedná se o nejvýraznější diachronní trend v datech. Vzhledem k tomu, že nejvýraznější vliv na umístění textu v této dimenzi mají rysy celostně kvantifikující texty z hlediska lexikální bohatosti, lze jen obtížně demonstrovat rozdíl mezi jednotlivými extrémy pomocí úryvku. Srovnáme-li nicméně první a poslední tři roky sledovaného období pomocí ukazatelů, jako je standardizovaná velikost repertoáru slovních tvarů či bigramů (dvojic) lemmat (zTTR, viz CVRČEK — CHLUMSKÁ 2015), Yuleův koeficient repetitivnosti lexikonu (OAKES 1998: 204) nebo tematická koncentrace (ČECH 2016), ve všech případech dochází v průměru k posunům svědčícím o větší diverzitě lexikonu. Tu lze interpretovat buď jako větší vyjadřovací variabilitu, nebo jako sklon kombinovat větší počet témat v rámci jednoho textu.

Jako potenciální vysvětlení tohoto trendu se nabízí vliv kratších povídkových textů v textovém typu COL na celkové výsledky. Podrobná analýza pomocí smíšených modelů však prokázala, že vliv jejich zapojení do vzorku je zanedbatelný, resp. že úryvky z textů COL se z hlediska 4. dimenze zásadně neliší od textů typu NOV, a naopak odhalila, že trend rostoucí polytematicčnosti v čase je nejvíc patrný u úryvků překladů (pocházejících z konce textů, viz obr. 4). Lze tedy spekulovat, že překladatelé si buď vybírají lexikálně bohatší předlohy, které věrně převádějí do češtiny nadprůměrně bohatým slovníkem, nebo se jedná o trend v rámci překladatelské praxe, která směřuje k větší pestrosti vyjadřování.

Pokles hodnot v **dimenzi 5 (příklon k nižší míře explicitní adresnosti)** nevypadá tak významně, protože v průběhu sledovaného období dochází přechodně ke změně trendu (viz obr. 2). V porovnání počátečních a koncových bodů se ale jedná o druhý nejvýraznější rozdíl v datech.

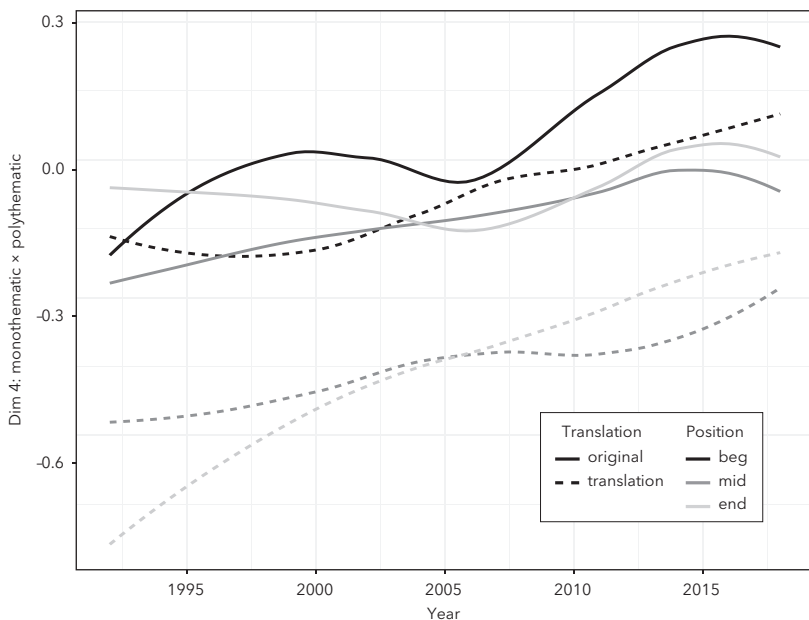
Vyšších hodnot v této dimenzi dosahují texty hojně využívající otázky, slovesa a zájmena vyjadřující druhou osobu, vokativ, imperativ a věty s vykřičníkem. Všechny tyto prostředky obsahuje následující ukázka z roku 1993, která přispívá k výraznému výkyvu směrem ke kladnému průměru v tomto roce.

... ech, Všiváku, Všiváku,
dokud vodám vládl Bůh Vod a větrům Bůh Větru, dokud v trávě přebýval
Bůh Trav a slunce hřálo všechny stejně, bylo volno na světě
— — ale k čemu to, Všiváku, řekni?
Jak poznáš světlo, neznaje tmy?
Jak poznáš dobro, neznaje zla? aha? nepoznáš — — nevíš.
Protože jsi blbej.
Dyť ty to ani poznat nechceš, Všiváku, přiznej to!
(BERKOVÁ 1993)

Styl úryvku sice nemá formu scénáře, ale sdílí s ním mnohé rysy, které se velkou měrou podílejí na výrazném umístění na pozitivním pólu 5. dimenze. Na opačném extrému najdeme úryvek z roku 2015:

Adolf Hitler má rád děti

Nebo si vzpomínám, jak jsme s otcem za domem sbírali šnečí ulity. Neohrabanými prsty čtyřletého dítěte jsem je opatrně vkládal do plastového kbelíku, jako by se mohly rozbít. Většina z nich už přitom byla děravá, některé se rozpadaly mezi prsty a v některých naopak ještě žil šnek. To jsem pak kbelík vždycky upustil a odmítal se k němu přiblížit, dokud otec šneka nepoložil do trávy alespoň tři kroky daleko. Měl jsem strach z tvorů menších, než jsem byl sám; cokoli, co nedosahovalo mých rozměrů, jsem



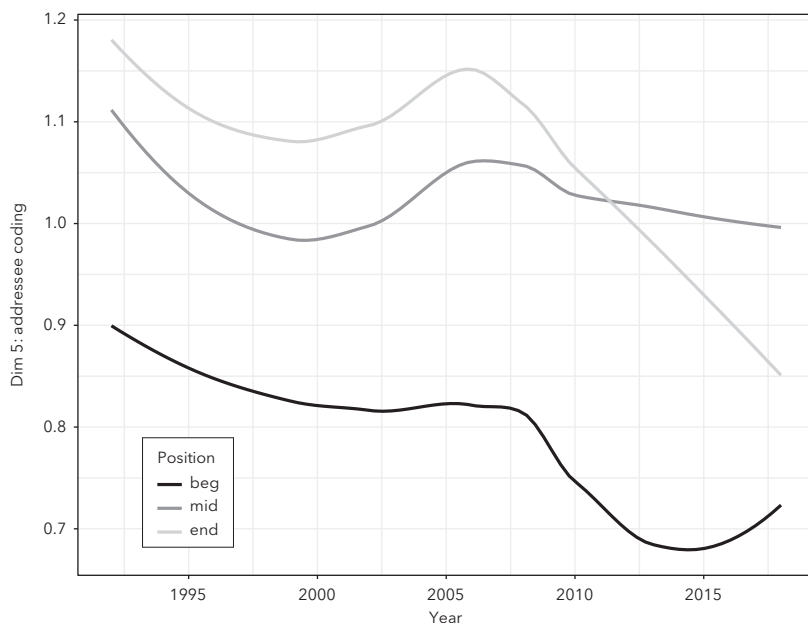
Obr. 4. Vývoj hodnot v dimenzi 4 (polytematičnost × monotematičnost) pro úryvky originální a překladové pocházející z počátku, prostředku a konce původního textu (beg = začátek, mid = prostředek, end = konec).

podezíral z potměšilých úmyslů. Takhle jsem tedy hleděl na svět živých bytostí. Matka zmizela krátce před mými třetími narozeninami.
(PETR 2015)

Text osciluje mezi narací a komentářem. Kromě toho, že se zde jevy typické pro adresnost, které můžeme sledovat v předcházející ukázce, neobjevují, je zde zjevná preference minulého času a první osoby.

V zásadě ale lze posun na této dimenzi zobecnit jako ústup od použití klasické přímé řeči, v níž je zapotřebí signalizovat adresnost. Můžeme tedy spekulovat, že tento odklon od adresnosti souvisí s tím, že děj neposouvá dopředu a nezžitovňuje pouze interakce postav, ale i sevřenější, možná subjektivnější pohled vypravěčů různého typu (jak je patrné i v analýze beletrie na dimenzi číslo 7 dále).

Pokles ve sledovaném období se přitom realizuje nejvíc v úryvcích z iniciálních pozic (expozice), které jsou celkově na nižších úrovních než úryvky z prostředku či konce textů (viz obr. 5).



Obr. 5. Vývoj hodnot v dimenzi 5 (vyšší × nižší míra adresnosti) pro úryvky z počátku, prostředku a konce původního textu (beg = začátek, mid = prostředek, end = konec).

Dimenze 7 reprezentuje variabilitu v rámci dynamického vyjadřování, využívajícího nadprůměrně sloves, a to rozlišením použití verb podle kategorie času na **prospektivní** (prézens a futurum) a **retrospektivní** (préteritum). Rostoucí trend, který v této dimenzi můžeme sledovat, svědčí o odklonu od tradiční retrospektivní narace, kterou v kombinaci s kontemplativními pasážemi rovněž zasazenými do minulosti nacházíme třeba v následující ukázce:

Pracoval ukryt v přítmi své pracovny a nikdo by se teď neopovážil ho rušit. Frederika se zcela oprávněně cítila být fungující součástí živého organismu KREMODRAMATU, jehož neviditelné nervy se chvěly vzrušením. Nezáviděla tomu slavnému muži, jehož kariéra možná visela na vlásku.

Kariéra zasloužená a dlouhé roky pracně budovaná. Hrůza státního pohřbu tkvěla v tom, že každá minuta života Miláčka národa jevila se detailně zmapovanou. Když mu bylo šest let, spadl na rodinném letním sídle do pstruhového potoka. Zachránil ho newfoundlandský pes, který ho vytáhl za kšandičky z vody, zatímco zmatkující chůva se neúspěšně probírala z mrákot. (VERNER 1994)

Nízkou hodnotu v 7. dimenzi má na svědomí především vysoká frekvence minulého času a třetí osoby. To poměrně ostře kontrastuje s ukázkou z fantasy románu z roku 2017:

S vrčením natáhnou ruku po zvonícím mobilu. Nadzvednu hlavu zabořenou do polštáře, otevřu pravý oko a zaostřím na svítící displej. Směje se z něj ženská. Hezká. Bohužel ji znám. Zrzavá je jako liška a taky tak prohnaná. Zmáčknu zelenou telefonku pro přijetí hovoru.

„Ghrrr?“ Nesnažím se ani správně artikulovat.

Je natolik chytrá, aby si uměla přeložit, že to znamená: Dobré ráno, má krásná, úžasná, dokonalá Naty. Doufám, že se ti daří výborně.

„Sig!?! Ty tam máš zase nějakou kurvu!“ zajecí.

S úšklebkem dám mobil dál od ucha. Prej zase.

(FUCHS 2019)

V ukázce je místo préterita výrazně častěji užito historický přezens, respektive přezens narační (ŠTÍCHA 2017), a první osoba (ta spolu s druhou osobou a s imperativem patří k nejméně výrazným rysům 7. dimenze). Užítí ich-formy, tedy přítomnosti tzv. přímého vyprávěče, a formálně přítomného času může přispívat k větší autentičnosti, živosti, která by mohla akcentovat či suplovat mluvenost, a plynulost. Vyprávění v I. osobě může být zdůrazněním toho, co je významné pro rozvoj děje, toho, co se přímo dotýká hlavní postavy příběhu.

Na trend ústupu od retrospektivnosti podle post-hoc šetření nemají vliv další zkoumané faktory, které by ho zvýrazňovaly nebo maskovaly.

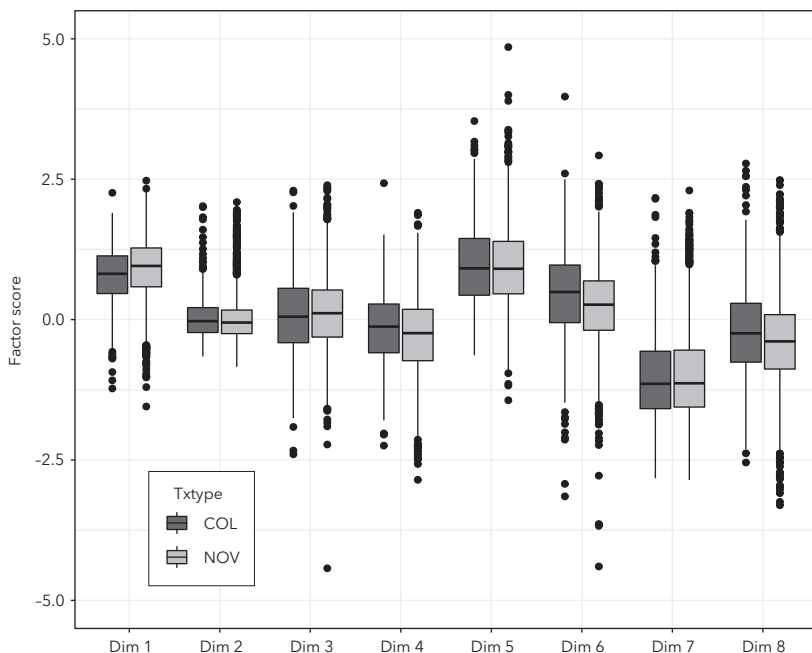
V následujících třech sekcích se krátce zastavíme u vlivu typu textu, ne/překládovosti a pozice úryvku ve výchozím textu na hodnoty v dimenzích. Opouštíme tak diachronní perspektivu a v těchto částech budeme na náš vzorek reprezentující téměř 30 let nahlížet jako na synchronní soubor (jímž z pohledu dějin moderní české literatury bezesporu je).

5.3 Analýza typu textu

Do vzorku vstupovaly dva prozaické textové typy — COL (kratší texty, povídky) a NOV (delší texty, romány). Tyto charakteristiky byly převzaty z anotace psaných textů v rámci ČNK, jak je uplatňována od roku 2015.⁹

Důvodem pro zařazení obou textových typů byl především pragmatický ohled na dostatek dat, která by pokrývala celou časovou osu. V této souvislosti je ovšem třeba prověřit, jestli tím nebyly do souboru vneseny odlišnosti dané těmito různými typy textů, resp. zda je možné oba tyto typy textů

9 Viz https://wiki.korpus.cz/doku.php/cnk:klasifikace_textu_syn2015.



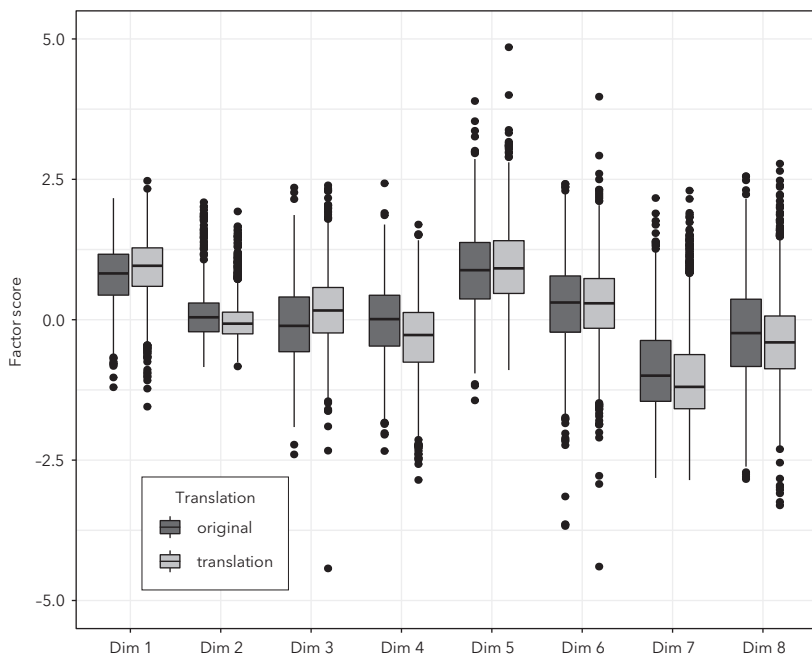
Obr. 6. Pozice textů v jednotlivých dimenzích podle jejich textového typu (COL = kratší próza, NOV = delší próza).

ve sledovaném období považovat za rovnocenné reprezentanty česky psané prózy z hlediska zastoupení registrových charakteristik. Typické pozice textů NOV a COL na jednotlivých dimenzích prezentuje graf v obr. 6.¹⁰

Ze statistického porovnání hodnot pro jednotlivé textové typy pomocí Wilcoxonova testu (opět po korekci metodou FDR) plyne, že signifikantní rozdíl nacházíme v dimenzích 1, 4, 6 a 8. Zdá se tedy, že povídky jsou oproti románům statictější, mírně polytematičtější, méně konkrétní a více signalizují postojovost.

Pokud bychom tedy ve sledovaném období měli nabídnout diferenční charakteristiku delších a kratších próz, můžeme ve shodě s výsledky uvedenými výše (tab. 3) konstatovat, že povídky oproti románům neposkytují konden-

¹⁰ Pro zobrazení střední hodnoty a naznačení disperze se obvykle používá tzv. boxplot (např. obr. 6). Tlustší čára uprostřed barevného obdélníku reprezentuje medián (prostřední hodnotu), okraje barevného boxu pak 1. a 3. kvartil (25 % a 75 % dat). Linie vedoucí svisle z obdélníků reprezentují 1,5násobek mezikvartilového rozpětí. Odlehlé hodnoty jsou naznačeny samostatnými body.



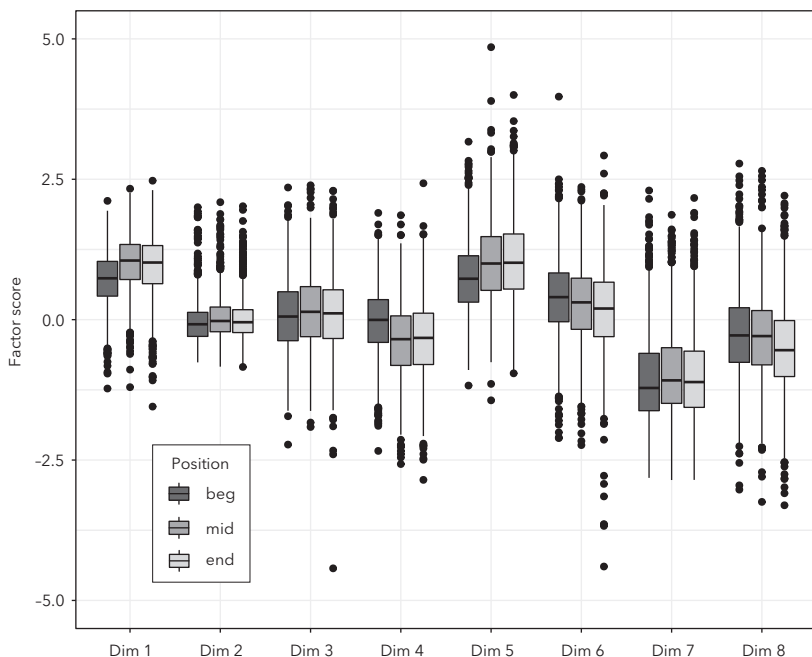
Obr. 7. Pozice textů v jednotlivých dimenzích podle toho, zda se jedná o překlady, či české originály.

zovanější naraci, kterou bychom si mohli představovat jako vyšší relativní četnost sloves. Povědky jsou naopak statictější (1. dimenze) a postojovější (8. dimenze), což ukazuje na úvahový či argumentační charakter. S tím souvisí i posun na ose od konkrétního k obecnému, který lze vykládat tím, že v povídkách není prostor pro detailní charakterizaci všech postav a míst, což je aspekt popisovaný 6. dimenzí. Naproti tomu posun na 4. dimenzi směrem k polytematičnosti lze asi přičíst na vrub kondenzovanému vyjadřování (v delším textu lze jen obtížně udržet stejně vysokou míru lexikální či tematické diverzity jako v textu krátkém).

5.4 Analýza překládovosti

Rozdíly mezi texty, které vznikly překladem z cizího jazyka, a texty originálně českými shrnuje graf v obr. 7.

Ačkoli rozdíl mezi překlady a nepřeklady v grafu není příliš nápadný, statisticky signifikantní (po korekci metodou FDR) je u všech dimenzí s výjimkou páté a šesté. Z výsledků tohoto synchronního zkoumání plyne, že překladové prozaické texty bývají dynamičtější, s menším počtem rysů spontánního



Obr. 8. Pozice textů v jednotlivých dimenzích podle jejich umístění ve výchozím textu (beg = začátek, mid = prostředek, end = konec).

vyjadřování, s větší mírou koheze, celkově tematicky zaostřenější (ačkoli polytematičnost v čase roste, viz obr. 4), víc retrospektivní a faktuální.¹¹

V zásadě nelze rozhodnout, jestli výsledky jsou dány tím, že překladatelé si (vědomě či nevědomky) vybírají předlohy odlišné od domácí prozaické produkce, které pak věrně převádějí do češtiny způsobem, který se promítá do rozdílů v grafu 7, nebo zda jsou odlišnosti v dimenzích důsledkem procesu překládání.

Podrobná analýza výsledků přesahuje možnosti tohoto příspěvku, nabízí se nicméně alespoň letmé srovnání s některými staršími výsledky, které vznikly analýzou heterogennějšího materiálu, a ze srovnání s nimi můžeme

¹¹ Za zmínku stojí, že v současném vzorku nacházíme statisticky signifikantní rozdíly mezi překlady a nepřeklady, přičemž v dřívější analýze provedené na datech z korpusu Koditexu se takové odlišnosti neobjevovaly. Tento rozpor nejspíš způsobily odlišnosti v datech; korpus Koditex obsahoval vedle beletrie i celou řadu textů z oblasti non-fiction, přičemž překlady v odborné literatuře a publicistice se od překladů uměleckých výrazně liší.

usuzovat na specifika překládání umělecké prózy. Lucie Chlumská uvádí, že překladové texty obecně obsahují více sloves (CHLUMSKÁ 2017: 133), což je ve shodě s výsledkem v dimenzi 1. Při zkoumání simplifikace v překladech dospěli Cvrček a Chlumská k závěru, že nepřekladové texty jsou lexikálně bohatší (CVRČEK — CHLUMSKÁ 2015), což podle výsledků v dimenzi 4 u samotné beletrie rovněž můžeme potvrdit.

5.5 Analýza pozice úryvku

V této části prověřujeme, zda v našich datech najdeme systematické rozdíly mezi úryvkem, který byl vybrán ze začátku, prostředku či konce textu. Ačkoli byl algoritmus pro vybírání úseků do značné míry náhodný, lze předpokládat, že tato kategorie bude odrážet vývojové fáze příběhu (např. expozice bude častější ve vzorcích ze začátku, zatímco katastrofa či katarze v koncových). Rozložení hodnot faktorových skóre, které udávají pozici textu na dané dimenzi, shrnuje pro jednotlivé druhy úryvků podle jejich pozice v originálním textu obr. 8.

Ačkoli rozdíly v grafu v obr. 8 nevypadají výrazné, jsou v mnoha případech statisticky signifikantní.

Z výsledků testů plyne, že ve většině dimenzí nacházíme rozdíl mezi počátečními úseky a úseky neinicíálními (tedy středovými a koncovými). S výjimkou 8. dimenze (míra postojovosti či faktuality) to platí pro všechny.

Celkově lze shrnout, že iniciální části jsou ve srovnání s prostředkem a koncem textů statictější (dimenze 1), méně spontánní ve vyjadřování (dimenze 2), vykazují nižší míru koheze (dimenze 3), bývají polytematictější či lexikálně bohatší (dimenze 4), z hlediska adresnosti (oslovování mezi postavami) méně explicitní (dimenze 5) a s větším podílem retrospektivního vyjadřování (dimenze 7).

Naproti tomu statisticky signifikantní rozdíl mezi střední a závěrečnou částí nacházíme pouze v 6. (obecný × konkrétní) a 8. dimenzi (postojovost × faktualnost). V obou případech sledujeme v obr. 8 u těchto dimenzí postupný pokles, což značí, že prozaická díla mají tendenci postupně snižovat svou obecnost a postojovost ve prospěch konkrétních pojmenování a faktuality modality.

Tyto výsledky by si rovněž zasloužily detailnější analýzu, která jde nad rámec stávající studie, nicméně pouhé konstatování rozdílů mezi počátečními, středovými a koncovými pasážemi nás opravňuje k uvažování o horizontálním členění prozaického díla, které je pozorovatelné na registrové charakteristice (srov. s charakteristikou v tab. 3).

6. Závěr

Předložená studie představuje první sondu do vývoje česky psané prózy pomocí empirické kvantitativní korpusové metodologie (MDA). Klíčem přitom pro nás nejsou literárněvědné charakteristiky, na jejichž uchopení použitý

aparát cílit nemůže, ani introspektivní analýzy stylistické, jež mnohdy postrádají potřebnou metodologickou rigoróznost a potenciál ke zobecnění. Soustředíme se na analýzu registru, tedy redefinici toho, co bylo tradičně chápáno pod termínem *umělecký styl*, a to s využitím informace o souvýskytech širokého spektra jazykových prostředků v textech.

Možnost aplikace multidimenzionálního modelu na vybranou prózu nám poskytuje ucelenější pohled na charakterizaci tak široké skupiny textů, jako jsou texty umělecké. Vzhledem k tomu, že kvalitativní stylistikou bývají literární texty nejčastěji nahlíženy jako texty s převahou estetické funkce a při konkrétních popisech je akcentována „mluvenost“ těchto textů (ať už ji badatelé uchopovali jakkoli), naskytá se zde příležitost rozšířit jejich popis o další prostředky než jen ty napodobující běžný mluvený jazyk. Zároveň se otevírá možnost nově přemýšlet o aktualizaci jako o konstituujícím prvku estetické funkce, která je pro beletrii základní. Můžeme tak uvažovat o jejím doplnění o jazykové prostředky ne nutně aktualizované, jež se na této funkci nicméně také podílejí.

Sonda, kterou jsme výše představili, opravňuje k opatrnému formulování toho, jak vypadá současná česky psaná próza. Z výsledků můžeme dovozovat, že v průběhu posledního čtvrtstoletí došlo k posunu v tom, že texty jsou méně kohezni; svou strohou výčtovostí působí akčnějším dojmem rychlého střídání obrazů. Zároveň v nich převládá polytematicnost, opřená buď o větší vyjadřovací variabilitu, nebo motivovaná sklonem kombinovat větší počet témat v rámci jednoho textu, zdánlivá roztržitost nebo naopak otevřenost, zobrazení nejen výseku skutečnosti, ale celého univerza (např. fantasy ságy a budování celých imaginárních světů).

Dále můžeme spekulovat, že pozorovaný odklon od adresnosti v textech souvisí s tím, že děj neposouvá dopředu interakce postav, která může mít zži-votňující vliv na čtenáře, ale i sevřenější pohled postav, respektive vypravěčů, kteří mohou nabídnout pohled vnější i vnitřní.

Výsledky (zejm. v obr. 8) rovněž indikují přítomnost horizontálního členění textu, vývoj příběhu od úvodu, expozice a nastínění problematiky (dokládá statičnost, menší spontaneita, nižší míra textové soudržnosti atd.) po vyvrcholení a katarzi (větší konkrétnost, faktualnost).

Uvědomujeme si, že tyto postřehy zaslouží podrobnější analýzu na dalších vzorcích. V tomto smyslu tak naše studie slouží spíše k postulování nových otázek než k formulování definitivních odpovědí.

Prameny

- ASIMOV, Isaac — MACBRIDE, Allen Roger
 1997 *Kalibán*; přel. Veronika Volhejnová (Praha: Knižní klub)
- BERKOVÁ, Alexandra
 1993 *Utrpení oddaného Všiváka* (Brno: Petrov)
- FUCHS, Oskar
 2019 *Frost* (Praha: Epoque)
- PETR, Daniel
 2015 *Straka na šibenici* (Brno: Host)
- ŠÁMAL, Petr
 2016 *Záchranáři* (Praha: Epoque)
- VERNER, Pavel
 1994 *Pražské hyeny* (Praha: Mladá fronta)

Bibliografie

- BENEŠOVÁ, Lucie — KŘEN, Michal — WACLAWIČOVÁ, Martina
 2013 *ORAL2013: Reprezentační korpus neformální mluvené češtiny* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022
- BENKO, Vladimír
 2014 „Aranea: Yet another family of (comparable) web corpora“; in Petr Sojka, Ivan Kopeček, Aleš Horák, Karel Pala (edd.): *International Conference on Text, Speech, and Dialogue* (Berlin: Springer), s. 257–264
- BERMEL, Neil
 2000 *Register Variation and Language Standards in Czech* (München: Lincom Europa)
 2001 „Střídání kódů či míšení jazykových prostředků“; *Naše řeč* LXXXIV, č. 1, s. 16–30
- BIBER, Douglas
 1988 *Variation Across Speech and Writing* (Cambridge: Cambridge University Press)
 1995 *Dimensions of Register Variation: A Cross-Linguistic Comparison* (Cambridge: Cambridge University Press)
- BIBER, Douglas — CONRAD, Susan
 2009 *Register, Genre, and Style* (Cambridge: Cambridge University Press)
- CVRČEK, Václav — CHLUMSKÁ, Lucie
 2015 „Simplification in translated Czech: A new approach to type-token ratio“; *Russian Linguistics* XXXIX, č. 3, s. 309–325
- CVRČEK, Václav et al.
 2010 *Mluvnice současné češtiny: Jak se píše a jak se mluví. I* (Praha: Karolinum).
 2018 „Variabilita češtiny: Multidimenzionální analýza“; *Slovo a slovesnost* LXXIX, č. 4, s. 293–321

- 2021 „From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA“; *Corpus Linguistics and Linguistic Theory* XVII, č. 2, s. 351–382
- 2020 *Registry v češtině* (Praha: Nakladatelství Lidové noviny)
- 2015 *Korpus prezidentských projevů Speeches* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022
- ČECH, Radek
- 2016 *Tematická koncentrace textu v češtině* (Praha: ÚFAL)
- ČECHOVÁ, Marie — KRČMOVÁ, Marie — MINÁŘOVÁ, Eva
- 2008 *Současná stylistika* (Praha: Nakladatelství Lidové noviny)
- ČERMÁK, František — ADAMOVIČOVÁ, Anna — PEŠIČKA, Jiří
- 2001 *PMK: Pražský mluvený korpus* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022
- ČMEJRKOVÁ, Světa — HOFFMANNOVÁ, Jana (edd.)
- 2011 *Mluvená čeština: Hledání funkčního rozpětí* (Praha: Academia)
- GAMMELGAARD, Karen
- 1997 *Spoken Czech in Literature: The Case of Bondy, Hrabal, Placák and Topol* (Oslo: Scandinavian University Press)
- HLADKÁ, Zdeňka
- 2002 *BMK: Brněnský mluvený korpus* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>
- 2006 *KSK-dopisy (Korpus soukromé korespondence): Přepisy ručně psaných dopisů z let 1990–2004* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022
- HOFFMANNOVÁ, Jana et al.
- 2016 *Stylistika mluvené a psané češtiny* (Praha: Academia)
- CHLUMSKÁ, Lucie
- 2017 *Překladová čeština a její charakteristiky* (Praha: Nakladatelství Lidové noviny)
- JELÍNEK, Milan
- 1995 „Stylistika“; in Petr Karlík, Marek Nekula, Zdenka Rusínová (edd.): *Příruční mluvnice češtiny* (Praha: Nakladatelství Lidové noviny), s. 699–780
- JELÍNEK, Tomáš
- 2008 „Nové značkování v Českém národním korpusu“; *Naše řeč* XCI, č. 1, s. 13–20
- KADERKA, Pter — PETEREK, Nino
- 2012 *Korpus DIALOG, ver. 1.1* (Praha: Ústav pro jazyk český AV ČR) <http://ujc.dialogy.cz>, přístup 3. 6. 2022
- KARLÍK, Petr — NEKULA, Marek — RUSÍNOVÁ, Zdenka (edd.)
- 1995 *Příruční mluvnice češtiny* (Praha: Nakladatelství Lidové noviny)
- KODÝTEK, Vilém
- 2007 *On the Replicability of the Biber Model: The Case of Czech* <https://www.korpus.cz/biblio/2722>, přístup 3. 6. 2022

- 2008 „Variace v mluvené češtině v Čechách: Sonda do ORAL2006“; in Marie Kopřivová, Martina Waclawičová (edd.): *Čeština v mluveném korpusu* (Praha: Nakladatelství Lidové noviny), s. 132–140
- KŘEN, Michal et al.
- 2016 „Representative Corpus of Contemporary Written Czech“; in Nicoletta Calzolari, Khalid Choukri, Thierry Declerck (edd.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (European Language Resources Association), s. 2522–2528
- 2021 SYN (Verze 9) (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022
- MAGLIONE, Concetta
- 2001 „Nové uplatnění obecné češtiny v literatuře“; *Naše řeč* LXXXIV, č. 2, s. 74–80
- MAREŠ, Petr
- 2012 *Nejen jazykem českým. Studie o vícejazyčnosti v literatuře* (Praha: Univerzita Karlova, Filozofická fakulta)
- 2017 „Umělecký styl“; in Petr Karlík, Marek Nekula, Jana Pleskalová (edd.): *Nový encyklopedický slovník češtiny* <https://www.czechency.org/slovník/UMĚLECKÝ STYL>, přístup 3. 6. 2022
- MINÁŘOVÁ, Eva
- 2017 „Stylový faktor“; in Petr Karlík, Marek Nekula, Jana Pleskalová (edd.): *Nový encyklopedický slovník češtiny* <https://www.czechency.org/slovník/UMĚLECKÝ STYL>, přístup 8. 12. 2022
- MRÁZKOVÁ, Kamila
- 2016 „Rejstřík“; in Petr Karlík, Marek Nekula, Jana Pleskalová (edd.): *Nový encyklopedický slovník češtiny* <https://www.czechency.org/slovník/REJSTŘÍK>, přístup 3. 6. 2022
- OAKES, Michael
- 1998 *Statistics for Corpus Linguistics* (Edinburgh: Edinburgh University Press)
- SPOUSTOVÁ, Drahomíra et al.
- 2007 „The Best of Two Worlds: Cooperation of Statistical and Rule-Based Taggers for Czech“; in Jakub Piskorski, Hristo Tanev (edd.): *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*, Association for Computational Linguistics <https://www.aclweb.org/anthology/W07-I709>, s. 67–74, přístup 3. 6. 2022
- STRAKOVÁ, Jana — STRAKA, Milan — HAJIČ, Jan
- 2013 „A New State-of-The-Art Czech Named Entity Recognizer“; in Ivan Habernal, Václav Matoušek (edd.): *Text, Speech, and Dialogue* (Berlin: Springer), s. 68–75 <https://doi.org/10.1007/978-3-642-40585-3_10>, přístup 3. 6. 2022
- 2014 „Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition“; in Kalina Bontcheva, Jingbo Zhu (edd.): *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (ACL), s. 13–18 <https://doi.org/10.3115/v1/P14-5003>, přístup 3. 6. 2022

ŠTÍCHA, František

- 2017 „Historický prézens“; in Petr Karlík, Marek Nekula, Jana Pleskalová (edd.): *Nozý encyklopedický slovník češtiny* (Praha: Nakladatelství Lidové noviny) <https://www.czechency.org/slovník/HISTORICKÝ_PŘEZENS>, přístup 3. 6. 2022

VOLÍN, Jan

- 2007 *Statistické metody ve fonetickém výzkumu* (Praha: Epoque)

ZASINA, Adrian Jan — KOMRSKOVÁ, Zuzana

- 2019 „Koditex — korpus diverzifikovaných textů“; *Studie z aplikované lingvistiky X*, č. I, s. 127–132

ZASINA, Adrian Jan et al.

- 2018 *Koditex: Korpus diverzifikovaných textů* (Praha: Ústav Českého národního korpusu FF UK) <www.korpus.cz>, přístup 3. 6. 2022

Creative Commons

Toto dílo podléhá licenci Creative Commons — Uveďte původ — Neužívejte komerčně — Nezpracovávejte 4.0 Mezinárodní licence. Licenční podmínky jsou dostupné zde: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.cs>.

