

Univerzita Karlova v Praze
Filozofická fakulta
Ústav Českého národního korpusu
Filologie – Matematická lingvistika

TEZE DISERTAČNÍ PRÁCE

Překladová čeština a její charakteristiky

Translated Czech and Its Characteristics

Mgr. Lucie Chlumská

Vedoucí disertační práce: doc. Mgr. Václav Cvrček, Ph.D.

Praha 2015

1 Úvod

Překladová literatura u nás

Překlady z cizích jazyků tvoří snad ve všech kulturách s psanou tradicí určitou část vydávané literatury, o to významnější pak v malých zemích, kde se domácí produkce nemůže objemem rovnat produkci zahraniční. To je i případ českého prostředí. Ačkoli se jazyky a kultury, z nichž se nejvíce překládalo, v dějinách naší země proměňovaly, překladová literatura měla v naší kultuře vždy nezastupitelné místo. S rozvojem domácí kultury se překlad postupně emancipoval a stal se její relativně samostatnou součástí. Tím se otevřely možnosti zkoumání překladu jako takového – od jeho dějin, tradice či norem, až po výzkum hypotézy zastarávání překladu či samotného jazyka překladu.

Úměrně se zvyšováním domácí literární produkce se dařilo a daří i překladům. Podíváme-li se na situaci v nedávných pěti letech (od roku 2008 do roku 2012), počet překladů neperiodických publikací – kam patří beletrie, populární i odborná literatura – stále narůstá. V roce 2012 bylo vydáno téměř o 98 % překladových knih více než v roce 2008. V roce 2012 vyšlo celkem 5 871 titulů překladové literatury, což je více než 34 % z celkového objemu knižní produkce (viz statistiky NKP¹).

Je všeobecně přijímaným faktem, že recepce textů má vliv na percepci i na produkci jazyka příjemců. Uvážíme-li tedy, že více než třetinu produkce dnes tvoří překladová literatura, musí nás nutně zajímat, zda se překladový jazyk nějak neliší, zdali není svébytným kódem, který má své vlastní zákonitosti a svá pravidla. Je překladová čeština jiná než čeština původních, česky psaných děl? Vykazuje jazyk překladů nějaké specifické rysy, jež jsou pro něj typické bez ohledu na jazyk, z něhož byl překlad pořízen? Je možné tyto rysy odhalit a popsat pomocí korpusových metod? Nejen na tyto otázky se snaží odpovědět tato disertační práce.

¹Dostupné ve formátu PDF zde: http://www.nipos-mk.cz/wp-content/uploads/2013/05/Statistika_kultury_2012_III.KNIHOVNY_web.pdf nebo zde: http://text.nkp.cz/soubory/ostatni/vykaz_dd2012.pdf.

Téma práce

Tématem této disertace je tedy překladový jazyk, angl. *language of translation*², v tomto případě jazyk textů přeložených z cizího jazyka do češtiny. Cílem výzkumu je s pomocí kvantitativních metod zjistit, zda se tento jazyk nějak odlišuje od jazyka nepřekladového, a pokud ano, formulovat a popsat jeho charakteristické rysy. Za nepřekladový jazyk jsou zde považovány takové texty, které byly původně napsány i publikovány v českém jazyce pro české publikum. Mezi zkoumané texty je zařazena jak beletrie, tak i odborná literatura. Je třeba předem zdůraznit, že cílem práce není hodnotit kvalitu překladu textů, ani by to vzhledem k objemu a povaze zkoumaných dat nebylo možné. Hlavním předmětem zájmu jsou jakékoli rysy typické pro překladový jazyk, jejich interpretace a kategorizace s přihlédnutím k dosavadnímu výzkumu v této oblasti (viz kapitola 2).

Kvantitativní pohled je v této oblasti české translatologie výjimkou (např. Kubáčková 2008), většina studií a vědeckých prací, které se zabývají překladovým jazykem a překladovými univerzáliemi, se omezuje na kvalitativní srovnání jedné nebo více verzí konkrétního překladu s původním dílem (např. Polišínská 2010, Kočová 2009, Středová 2009). Tato práce naopak využívá rozsáhlý textový korpus překladového a nepřekladového jazyka v řádu desítek milionů slov, který byl pro tyto účely sestaven na Ústavu Českého národního korpusu FF UK (viz kapitola 3.2). Výhodou tohoto přístupu je tak jeho interdisciplinarita, která plyne ze spojení metod korpusové lingvistiky s poznatky moderní translatologie.

Struktura práce

Tato práce je rozčleněna do pěti kapitol. Po krátké úvodní kapitole následuje kapitola 2, která shrnuje dosavadní výzkum v oblasti moderní translatologie se zaměřením na deskriptivní translatologii.

Třetí kapitola s názvem Data a metodologie nejprve stručně představuje druhy korpusů používané v korpusové translatologii a vymezuje použitou terminologii. Jejím těžištěm však jsou informace o korpusu Jerome. Část věnovaná metodologii představuje hlavní důvody pro využití kvantitativní analýzy a předkládá formulaci základní výzkumné hypotézy.

Kapitola 4 nazvaná Rysy překladové češtiny tvoří jádro práce. Je rozdělena do několika podkapitol, které jsou věnovány tradičně vymezovaným překladovým univerzáliím a dalším jazykovým jevům, které jsou zde testovány.

Závěrečná kapitola 5 pak shrnuje výsledky celé práce a poukazuje na možnosti výzkumu v oblasti překladového jazyka a překladových univerzálií.

²Doslovným ekvivalentem v češtině je „jazyk překladu“, který se v tomto smyslu také používá. Termín „překladový jazyk“ se však jeví jako vhodnější, především kvůli analogii s „překladovou literaturou“ a kvůli možnosti vytvořit antonymum „nepřekladový jazyk“, tedy jazyk nepřeložených textů (v kontrastu k původním textům či originálům, které zpravidla označují zdrojové texty překladu).

2 Přehled dosavadního výzkumu

Cílem přehledu, který tvoří druhou kapitulu práce, je stručně představit vývoj translatologie, který na počátku jedenadvacátého století vyústil v systematické zkoumání jazyka překladu a jeho charakteristických rysů. Na výzkum tzv. překladových univerzálií měl nepochybně největší vliv příklon k deskriptivní translatologii na konci sedmdesátých let dvacátého století a později nástup rozsáhlých jazykových korpusů v lingvistice, proto je těmto tématům věnována největší pozornost.

Korpusový výzkum v translatologii

Podíváme-li se na vývoj translatologie v posledních desetiletích, dá se říci, že propojení deskriptivní translatologie a korpusové lingvistiky bylo prakticky nevyhnutelné; obě disciplíny totiž zastávají obdobný pohled na jazyk. Jak korpusová lingvistika, tak deskriptivní translatologie vychází z empirické perspektivy a zkoumá jazyk na základě reálných dat, nikoli na základě intuitivních předpokladů. Výběr textů ke zkoumání se v obou případech neřídí nějakou předem či obecně danou představou „vhodných“ textů, nýbrž cílem popsat jazyk (resp. překladový jazyk) takový, jaký ve skutečnosti je.

Skutečnou průkopnicí korpusových nástrojů v translatologii a zakladatelkou disciplíny, kterou zde budeme nazývat „korpusová translatologie“ (*corpus-based translation studies*, příp. *corpus translation studies*), je Mona Bakerová, která na počátku devadesátých let svým článkem (Baker 1993) zásadním způsobem ovlivnila posledních dvacet let translatologie a odstartovala honbu za překladovými univerzáliemi.

Za východiska korpusové translatologie můžeme považovat deskriptivní pohled na jazyk a překlad, orientaci na cílový text a cílovou kulturu a využití metodologie korpusové lingvistiky a empirických dat. Ústředním bodem zkoumání se stala myšlenka, že přeložené texty vykazují určité společné rysy, jež je odlišují od textů nepřeložených, napsaných v původním jazyce. Ještě před Bakerovou, která pro rysy tohoto typu zavedla termín „univerzálie“, se v pracích translatologů objevovaly pod názvem „třetí kód“, „překladaletština“ či „interference“. Tyto termíny jsou však dnes již vnímány jako příznakové a často pejorativní. Samotný termín univerzálie také prošel určitým vývojem, dnes se používá spíše v referenci na původní vymezení Bakerové a přednost je dáována neutrálním rysům či vlastnostem překladového jazyka.

Překladové univerzálie

Tradičně se uvádějí čtyři původní překladové univerzálie, které vešly do širšího povědomí (Baker 1996: 176–7):

1. **simplification** [...] the idea that translators subconsciously simplify the language or message or both
2. **explicitation** [...] the tendency to spell things out in translation, including, in its simplest form, the practice of adding background information
3. **normalisation or conservatism** [...] the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggerating them
4. **levelling-out** [...] the tendency of translated text to gravitate around the centre of any continuum rather than to move towards the fringes

Mnoho kritiků Bakerové oprávněně považuje její formulace za vágní a příliš obecné, což s sebou nese nutnost operacionalizace hypotéz (viz dále). Překladové univerzálie přesto představují bohatý zdroj pro výzkum jazyka překladu a inspirovali badatele k identifikaci dalších možných rysů, jako je např. novější *unique items hypothesis* (Tirkonnen-Condit 2002, 2004), *shining-through* (Teich 2003), *gravitational pull* (Halverson 2003) a další.

Na překladové univerzálie lze nahlížet ze dvou základních úhlů, které souvisejí s daty, která k výzkumu badatel potřebuje. Praktické rozlišení, které je uplatňováno i v této práci, přinesl Andrew Chesterman (2004a: 8, 2004b: 39). Vychází přitom ze vztahu mezi třemi typy textů: překlady v jazyce A, jejich zdrojovými texty v jazyce B (originály ve smyslu původní texty, jež byly podkladem pro překlad) a nepřeklady v jazyce A (originály ve smyslu původně psaná, nepřekladová literatura).

Pokud chceme srovnávat překlad a originál, zajímá nás pravděpodobně vztah ekvivalence (*equivalence*) v širokém slova smyslu: nakolik si texty odpovídají po obsahové i stylistické stránce. Pokud však porovnáváme překlady s nepřeklady, o ekvivalenci nelze hovořit; kritériem srovnání je zde přirozenost či přijatelnost překladového textu. Na základě tohoto rozdělení pak rozeznává univerzálie dvojího typu. **S-univerzálie** odrážejí vztah mezi překlady a jejich zdrojovými texty (vztah ekvivalence) a týkají se způsobu, jakým překladatelé nakládají se zdrojovým textem (S = *source*). Oproti tomu **T-univerzálie** lze hledat ve srovnání překladových textů s nepřekladovými (vztah podobnosti). T-univerzálie jsou výsledkem toho, jak překladatelé zacházejí s cílovým jazykem (T = *target*). K oběma skupinám uvádí Chesterman (2004: 8) příklady potenciálních projevů. V případě S-univerzálií může jít například o simplifikaci (zjednodušování z hlediska originálu), o explicitaci a s ní související prodlužování překladového textu, snahu o neopakování týchž výrazů, normalizaci dialektu nebo o užívání konvenčnějších kolokací oproti originálu. U T-univerzálií zmiňuje také simplifikaci (ovšem z hlediska srovnání s nepřekladovými texty), užití netypických lexikálních clusterů (Mauranen 2000) nebo naopak nedostatečné zastoupení takových jevů, jež jsou pro cílový jazyk jinak specifické (*unique items hypothesis*, Tirkonnen-Condit 2002, 2004).

Mezi oběma těmito typy univerzálií (či překladových rysů) je tedy třeba pečlivě rozlišovat, neboť oba vyžadují pro výzkum různé korpusy: buď paralelní, jež obsahují překlady a jejich zdrojové texty, nebo srovnatelné, které zahrnují obdobné texty v překladovém a nepřekladovém jazyce (přehled a využití různých korpusů shrnuje kapitola 3.1). Vzhledem k datové základně, jež je využita v této práci (viz kapitola 3.2), jsou zde zkoumány T-univerzálie, konkrétně pak simplifikace, konvergence a typické slovní kombinace. Tyto analýzy doplňuje i porovnání distribuce slovních druhů a POS-gramů.

Operacionalizace hypotéz

Jak již bylo řečeno, původní vymezení překladových univerzálií je velmi obecné a v praxi netestovatelné. Federico Zanettin (2013) na základě dvaceti vybraných studií z korpusové translologie, jež se věnují univerzáliím, sestavil pomocný interpretační rámec, skládající se ze čtyř rovin, od nejabstraktnější až po nejkonkrétnější. Rozlišení, které nabízí, je užitečné a přehledné a umožňuje lepší operacionalizaci hypotéz a jejich testování v praxi na reálných datech, proto z něj vycházím i v této práci. Zanettin rozlišuje čtyři roviny abstrakce:

1. rovinu **teorie**, která v tomto případě zahrnuje výchozí hypotézu, že všechny překladové texty vlivem procesu překladu sdílejí určité vlastnosti, které je odlišují od podobných nepřekladových textů;
2. rovinu **deskriptivních rysů** podporujících danou teorii, zde konkrétní překladové univerzálie: simplifikace, konvergence a další;
3. rovinu **jazykových indikátorů** (*linguistic indicators*), jimiž se realizuje ten který deskriptivní rys na různých jazykových rovinách, např. lexikální hustota (*lexical density*) nebo netypické kolokace;
4. rovinu **formálních operátorů** (*formal operators*), s jejíž pomocí se zmíněné abstraktní lingvistické rysy zkoumají v textech (konkrétní metody výpočtu a testy).

(Zanettin 2013: 21)

U výchozí hypotézy a dílčích hypotéz lze vycházet z dosavadních výzkumů, ačkoli i tam je třeba přesně specifikovat vybrané překladové rysy (viz konkrétní části kapitoly 4), neboť jednotliví badatelé si mnohdy odporují nebo označují obdobný jev různě. Nejdůležitější a zároveň nejobtížnější krok ve výzkumu překladových rysů však představuje samotná identifikace relevantních jazykových indikátorů a s tím související volba vhodných prostředků k jejich zkoumání. Je zjevné, že operacionalizací abstraktních ukazatelů, jakými může být například lexikální kreativita nebo naopak repetitivnost, vždy nutně dochází k redukci či generalizaci; s tím však musí badatel u kvantitativního výzkumu na datech tohoto rozsahu počítat.

3 Data a metodologie

Těžiště této kapitoly spočívá v popisu korpusu Jerome, který byl sestaven speciálně pro účely zkoumání překladové češtiny a tvoří hlavní materiálovou základnu této práce. Část 3.3 pak shrnuje základní metodologické požadavky pro korpusovětranslatologický výzkum a uvádí výchozí výzkumnou hypotézu práce.

Korpus Jerome

Korpus Jerome (Chlumská 2013) je jednojazyčný srovnatelný korpus, který je určen ke zkoumání překladové češtiny v porovnání s češtinou nepřekladovou. Před zveřejněním korpusu byl výzkum překladové češtiny obtížněji uskutečnitelný; korpusy synchronního psaného jazyka (řada SYN), kterými čeština disponuje, sice překladové texty zahrnují, ale nebylo pro uživatele snadné si vytvořit vlastní vyvážený subkorpus. Ačkoli byl korpus Jerome zamýšlen především jako materiálová základna pro výzkum představený v této práci, je zdarma přístupný všem registrovaným uživatelům Českého národního korpusu³ a může sloužit k výzkumu překladové češtiny všem zájemcům z řad translatologů i bohemistů.

Výběr dat je (nejen) v korpusovém výzkumu vždy zásadním krokem, proto je složení korpusu Jerome v práci věnována velká pozornost. Při vytváření jednojazyčného srovnatelného korpusu je nezbytně nutné si nejprve určit **hlavní kritéria**, která budou dodržena u obou částí korpusu, překladové i nepřekladové. Prioritou při sestavování korpusu Jerome byla především jeho výsledná **velikost**, která by umožnila alespoň částečnou generalizaci výzkumných zjištění o překladové češtině. Srovnatelnost obou částí korpusu byla zachována v následujících kategoriích: počet tokenů u překladů i nepřekladů, typ textu, časové rozpětí vydání textu a heterogenita autorů/překladatelů. Požadavek vyvážit korpus z hlediska pohlaví autora či překladatele se ukázal být v přímém rozporu s prioritou velikosti korpusu – pokud bychom trvali na stejném počtu autorů a autorek (resp. překladatelů a překladatelek), výsledný korpus by dosahoval pouze velikosti v řádu stovek tisíc tokenů.

Jako zdroj dat pro korpus Jerome posloužila **databáze textů ČNK**, konkrétně pak texty zahrnuté do korpusu SYN⁴, jež byly vybrány **ručně** a doplněny o translatologicky relevantní anotaci. Počty zahrnutých překladových děl přibližně odráží skutečnou situaci překladové literatury u nás.

³<http://www.korpus.cz>

⁴<http://wiki.korpus.cz/doku.php/cnk:syn>

Složení korpusu Jerome

Korpus pro výzkum překladových rysů by měl v ideálním případě obsahovat v překladové části obdobný počet textů (resp. tokenů) z co možná nejvíce typologicky odlišných jazyků, aby se vyloučila možná interference z převažujícího zdrojového jazyka. Tento požadavek předpokládá, že tvůrce korpusu má k dispozici obdobné množství textů ve vybraných jazycích. To však – zvláště u malých jazyků – ve skutečnosti neplatí. Zpravidla je zde jeden zdrojový jazyk, jehož překlady výrazně převažují nad ostatními – v posledních letech je tím jazykem nejen v českém prostředí **angličtina**. Podíváme-li se podrobněji na statistiky překladové literatury (neperiodických publikací – tedy beletrie a odborné literatury)⁵, zjistíme, že překlady z angličtiny tvoří více než polovinu celkového počtu překladů

Dá se tedy říci, že český čtenář přichází do kontaktu převážně s překlady z angličtiny a ty se pak významně podílejí na jeho představách o překladovém jazyku. Z tohoto předpokladu tak vychází celá **koncepce korpusu Jerome**. Jeho cílem je zobrazit překladovou češtinu tak, jak se s ní pravděpodobně setkává běžný uživatel jazyka, tedy i s převahou textů přeložených z jediného jazyka. I v korpusu je proto ponechán jako převažující zdrojový jazyk angličtina (přesné údaje o počtu jazyků a děl viz kapitola 3.2.2). Tím se podařilo zachovat i hlavní prioritu korpusu – jeho velikost. Ovšem aby bylo možné v dalším výzkumu rozeznat překladové rysy/univerzálie od možné interference angličtiny, byl v rámci korpusu Jerome sestaven také jazykově vyvážený subkorpus (viz 3.2.3).

Důležitým faktorem při sestavování korpusu je jeho **heterogenita**. Zvláště u menších korpusů může převaha jednoho autora významně zkreslit výsledky zkoumání, neboť nelze rozlišit vliv idiolektu od rysů překladového jazyka. Při výběru textů tak bylo stanoveno **pravidlo tři** – žádný autor ani překladatel není v korpusu zastoupen více než třikrát. Počet tři byl stanoven arbitrárně jako kompromis mezi snahou o heterogenitu a úsilím o zařazení co možná nejvíce textů do korpusu.

Dalším podstatným kritériem při tvorbě srovnatelného korpusu je stáří textů. Cílem korpusu Jerome je zobrazit **současný jazyk**, ten lze ovšem definovat různým způsobem. Tradičně se současný jazyk chápe jako období tří generací (Cvrček et al. 2010: 34). Toto období je však z hlediska vývoje moderní češtiny příliš dlouhým časovým úsekem – důležitým předělem je zde rok 1989 a začátek devadesátých let, kdy došlo k proměně jazyka. Do korpusu Jerome proto byly zařazeny texty, které byly vydány mezi lety **1992–2009** (rokem 2009 končí nejnovější tituly zveřejněné v korpusu SYN2010).

Korpus Jerome je stejně jako všechny korpusy synchronní psané češtiny vytvořené v rámci ČNK lemmatizovaný a morfologicky značkový. Kromě anotace jednotlivých pozic v korpusu disponuje korpus Jerome doplněnou anotací, ve které nechybí uvedení prvního vydání, informace o tom, zda se jedná o překlad, a údaje o pohlaví autora a překladatele.

Nakonec se při dodržení ostatních relevantních kritérií podařilo dosáhnout velikosti **85 milionů pozic**. Velikost korpusů bývá tradičně udávána ve slovech, tedy řetězcích alfabetských znaků (bez číslic a interpunkce). Vzhledem k tomu, že i in-

⁵<http://www.nkp.cz/sluzby/sluzby-pro/sluzby-pro-vydavatele/vykazy>

terpunkce a její použití může být relevantním faktorem při zkoumání překladového jazyka, uvádím údaje o velikosti v pozicích (neboli tokenech), tedy všech jednotkách oddělených v textu mezerami (vč. interpunkce a číslic). Následující tabulka přehledně shrnuje velikost srovnatelných částí korpusu (v pozicích).

<i>JEROME</i>	nepřeklady	překlady	celkem
<i>beletrie</i>	26 551 540	26 617 523	53 169 063
<i>odborná</i>	15 949 930	15 946 319	31 896 249
<i>celkem</i>	42 501 470	42 563 842	85 065 312

Velikost korpusu Jerome (v pozicích)

Co se týče velikosti jednotlivých textů, cílem bylo, aby škála zahrnovala nejrůznější typy, od krátkých textů až po rozsáhlejší. Do korpusu byly proto vybrány texty v rozmezí délky 2 000 – 190 000 pozic. Po uplatnění pravidla tří a vyfiltrování relevantních textů obsahuje korpus Jerome texty 1 244 různých autorů nebo autorských kolektivů. V překladové části korpusu pak najdeme texty 607 překladatelů nebo překladatelských kolektivů.

Při tvorbě srovnatelného korpusu je nutné volit texty obdobného textového typu nebo žánru. Tento požadavek srovnatelnosti byl u korpusu Jerome naplněn na nejvyšší úrovni textového typu, tedy u skupin **beletrie** a **odborná literatura**, které jsou v překladové i nepřekladové části zastoupeny takřka totožným počtem pozic.

Problém kvality překladů zahrnutých v korpusu úzce souvisí s otázkou reprezentativnosti korpusů obecně. Fakt, že při sestavování korpusu Jerome nebyla kvalita překladu (ani textu jako takového u nepřekladové části) zohledněna, nemá v žádném případě sugerovat, že mezi překlady nejsou rozdíly, ba naopak. Tím, že do korpusu jsou zahrnuta nejrůznější díla z databáze ČNK, se otevírá prostor pro to, aby byly zastoupeny jak velmi kvalitní, průměrné, tak i ty méně kvalitní překlady, tedy **široké spektrum textů**, s nímž se běžný čtenář překladů reálně setkává. Základním předpokladem zde však zůstává, že samotná skutečnost, že překlady byly publikovány knižně pod hlavičkou fungujícího nakladatelství a musely tedy projít alespoň minimální redakční kontrolou, zaručuje alespoň určitou úroveň jejich přijatelnosti pro čtenáře. Velikost korpusu a heterogenita děl (zaručená pravidlem tří) pak představují další záruku, že případné nekvalitní překlady nebudou v tomto objemu dat zásadním způsobem zkreslovat výsledek.

Vyvážený subkorpus

Jak bylo zdůrazněno výše, hlavními přednostmi korpusu Jerome jsou jeho velikost a heterogenita textů, autorů i překladatelů. Vzhledem k těmto prioritám však nebylo možné korpus vyvážit z hlediska dalších kritérií, jako je srovnatelný počet textů ze všech zahrnutých zdrojových jazyků (u překladů) nebo srovnatelný počet žánrů v odborné literatuře. Z tohoto důvodu byl v rámci korpusu Jerome (ručním výběrem z jeho textů) vytvořen vyvážený srovnatelný subkorpus, který splňuje výše uvedené

požadavky. Subkorpus měl sloužit především k přesnější interpretaci výsledků, které byly zjištěny na korpusu Jerome, avšak v průběhu výzkumu, zvláště při statistickém testování, se ukázaly jeho limity, především s ohledem na jeho velikost.

<i>SUBKORPUS</i>	nepřeklady	překlady	celkem
<i>beletrie</i>	1 768 079	1 765 433	3 533 512
<i>odborná</i>	779 288	774 610	1 553 898
<i>celkem</i>	2 547 367	2 540 043	5 087 410

Velikost vyváženého subkorpusu (v pozicích)

Výchozí hypotéza

Zatímco při zkoumání S-univerzálií na paralelním korpusu se badatel může soustředit jak na shody překladu s originálem, tak na rozdíly mezi nimi, v případě T-univerzálií (na datech z *jednojazyčného* srovnatelného korpusu) jsou v centru pozornosti odlišnosti. Podobnost překladových textů s nepřekladovými v rámci jednoho jazyka totiž může značit nejen přiměřenost překladu, ale i pouhou přirozenost využitého jazyka, nic však už nevypovídá o společných rysech překladů coby samostatné skupiny textů. Z toho důvodu je výzkum překladové češtiny zaměřen právě na to, čím se tato odlišuje od textů nepřekladových.

Budeme-li tedy předpokládat, že překladové texty vykazují jiné vlastnosti než nepřekladové, nulová hypotéza na nejvyšší rovině abstrakce bude znít takto:

H_0 : Překladová a nepřekladová čeština se neliší.

Tato hypotéza pak bude konkrétním způsobem testována na vybraných rysech, a pakliže se nepotvrdí (testy a sondy odhalí rozdíly, které budou společné pouze překladovým textům), bude platit alternativní hypotéza:

H_1 : Překladová čeština se od nepřekladové odlišuje.

Tato výchozí hypotéza (v podobě nulové a alternativní hypotézy), jež je vzhledem ke své obecnosti v této podobě netestovatelná, tvoří **základ dílčích hypotéz**, které se vztahují k vybraným deskriptivním rysům překladových textů. Těm se podrobně věnuje kapitola 4.

4 Rysy překladové češtiny

Tato kapitola představuje těžiště výzkumu: přináší analýzy vybraných rysů překladového jazyka na základě korpusu Jerome. Všechny části této kapitoly jsou proto strukturovány tak, že postupují od nejvyšší roviny abstrakce až po nejnižší ve shodě se Zanettinovým interpretačním rámcem. Každá část (s výjimkou 4.1) je věnována konkrétnímu rysu překladového jazyka, přičemž **výběr zkoumaných rysů** se opírá o dosavadní translatické studie o univerzáliích provedených na jiných jazycích, které jsou v příslušné části vždy odůvodněny a okomentovány. Jednotlivé rysy založené na T-univerzáliích zahrnují simplifikaci (4.2), levelling-out/konvergenci (4.3) a (ne)typické slovní kombinace (4.4) vycházející z výzkumu n-gramů (častých sekvencí o délce n-slov).

Úvodní část 4.1 má obecnější charakter; není zaměřena na jeden konkrétní překladový rys, ale snaží se popsat překladovou češtinu z hlediska frekvence a distribuce slovních druhů a jejich kombinací a zároveň zohlednit rozdílnost textových typů (beletrie, odborná literatura). Tento přehled pak také slouží jako možný zdroj dat pro další analýzu v rámci jednotlivých překladových rysů. Cílem této kapitoly je také poukázat na limity často používaných statistických testů.

Obecné frekvenční charakteristiky

Frekvence neboli četnost výskytu jevu je v korpusové lingvistice jedním z klíčových konceptů. Frekvenční distribuce jevů v korpusu představuje zpravidla odrazový můstek pro další analýzu a může upozornit na rozdíly mezi zkoumanými soubory, v tomto případě mezi překladovými a nepřekladovými texty.

Frekvenční distribuce slovních druhů

Slovní druhy, klasifikující slova do skupin podle jejich vlastností a významů, představují v jazyce vyšší rovinu obecnosti – ačkoli bývají tradičně popisovány v rámci morfologie, jejich vymezení souvisí především s lexikonem a má důsledky i pro syntax. S kategorií slovních druhů pracují i některé výzkumy o překladových univerzáliích, především ty, které zkoumají bohatost slovní zásoby pomocí statistických měr (viz 4.2.3).

Výsledky srovnání v následujících dvou tabulkách ukazují, že v rámci obou textových typů, beletrie i odborné literatury, se distribuce slovních druhů v překladech i nepřekladech liší, avšak nikterak výrazně; rozdíly jsou sice statisticky signifikantní,

ale pohybují se maximálně v řádu jednotek procent. Pro výpočet věcné významnosti byla použita míra DIN neboli *difference index* (Cvrček & Fidler, v tisku), která ukázala, že naprostá většina slovních druhů v beletrii i odborné literatuře se pohybuje v rozmezí -10 až +10, což znamená, že jejich výskyt je v obou souborech relativně srovnatelný a nejde tedy o zcela odlišné tendence.

<i>BELETRIE</i>	nepřeklady		překlady		DIN
	tokeny	%	tokeny	%	
<i>substantiva</i>	5 502 041	20,72	4 832 060	18,15	-6,61
<i>adjektiva</i>	1 988 453	7,49	1 700 914	6,39	-7,92
<i>pronomina</i>	2 981 630	11,23	3 364 521	12,64	5,91
<i>numeralia</i>	446 766	1,68	305 719	1,15	-18,86
<i>verba</i>	4 201 372	15,82	4 684 899	17,60	5,32
<i>adverbia</i>	1 803 657	6,79	1 806 615	6,79	-0,04
<i>prepozice</i>	2 176 670	8,20	2 006 794	7,54	-4,18
<i>konjunkce</i>	1 889 785	7,12	1 885 087	7,08	-0,25
<i>partikule</i>	364 230	1,37	340 423	1,28	-3,50
<i>interjekce</i>	27 059	0,10	31 220	0,12	7,02
<i>neurčeno</i>	301 556	1,14	276 834	1,04	-4,40
<i>interpunkce</i>	4 868 321	18,34	5 382 437	18,15	4,89
<i>celkem</i>	26 551 540	100,00	26 617 523	100,00	0,00

Srovnání frekvenční distribuce slovních druhů – beletrie

<i>ODBORNÁ</i>	nepřeklady		překlady		DIN
	tokeny	%	tokeny	%	
<i>substantiva</i>	4 439 389	27,83	4 045 297	25,37	-4,63
<i>adjektiva</i>	1 962 341	12,30	1 719 111	10,78	-6,60
<i>pronomina</i>	1 114 150	6,99	1 379 899	8,65	10,67
<i>numeralia</i>	442 884	2,78	401 701	2,52	-4,86
<i>verba</i>	1 860 967	11,67	2 187 532	13,72	8,08
<i>adverbia</i>	799 276	5,01	852 786	5,35	3,25
<i>prepozice</i>	1 424 911	8,93	1 348 440	8,46	-2,75
<i>konjunkce</i>	1 018 342	6,38	1 136 385	7,13	5,50
<i>partikule</i>	176 158	1,10	170 193	1,07	-1,71
<i>interjekce</i>	3 098	0,02	2 310	0,01	-14,56
<i>neurčeno</i>	233 009	1,46	198 145	1,24	-8,07
<i>interpunkce</i>	2 475 405	15,52	2 504 220	15,70	0,59
<i>celkem</i>	15 949 930	100,00	15 946 319	100,00	0,00

Srovnání frekvenční distribuce slovních druhů – odborná literatura

Fakt, že na nejvyšší úrovni slovních druhů nedochází mezi překlady a nepřeklady k zásadním rozdílům, však nemusí znamenat, že se překladové a nepřekladové texty neliší na nižší rovině obecnosti, v rámci jednotlivých slovních druhů. Proto byla provedena analýza dvou velkých slovních druhů (substantiv a verb), které mohou coby autosémantika poukázat na určité lexikální tendence ve zkoumaných souborech. Bližší pohled byl věnován také zájmenům, u nichž se může projevat interference, a interpunkci, jejíž výskyt může souviset s hypotézou o simplifikaci. Cílem analýzy

bylo především zjistit, zda a jak se frekvenční seznamy těchto slovních druhů liší a především kolik a případně jakých slov se vyskytuje typicky jen v jednom ze souborů (překladech či nepřekladech). Výsledky přinesla kapitola 4.1.1.

Srovnání nejčastějších POS-gramů

Pro výzkum POS-gramů v korpusu Jerome byla zvolena délka čtyř po sobě jdoucích jednotek, která byla určena na základě výzkumu desambiguace kontextu jako struktura, která má v korpusu průměrně největší počet různých realizací (Cvrček & Václavík, v tisku). Pro každý textový typ zvlášť byly v korpusu Jerome vyhledány nejčastější po sobě jdoucí sekvence slovních druhů o délce čtyř pozic (s vyloučením interpunkce).

Z porovnání POS-gramů, které se v obou souborech nejvíce odlišují, vyplynulo, že v překladech se více prosazují verbální konstrukce, kdežto v nepřekladových textech jsou to konstrukce nominální se substantivy a adjektivy. Potvrzuje to trend, který vyplývá z analýzy frekvence slovních druhů: překlady vykazují vyšší výskyt sloves, kdežto původně česky psané texty obsahují přibližně o 2 % více substantiv. Zajímavý je také fakt, že 4-gramy s nejrozdílnějším výskytem se v obou textových typech vůbec nepřekrývají, mezi dvanácti nejodlišnějšími POS-gramy nenajdeme jedinou společnou konstrukci. To jen potvrzuje odlišnost beletrie a odborné literatury a poukazuje na to, že je nezbytné oba textové typy analyzovat zvlášť.

Simplifikace

Simplifikací podle Bakerové označujeme tendenci podvědomě zjednodušovat jazyk nebo obsah sdělení, příp. obojí (1996: 176). Sara Laviosová, která se simplifikací zabývala dlouhodobě (1996, 1998a, 1998b), na základě svých studií formulovala několik hypotéz, které se staly inspiračním zdrojem i pro výzkum simplifikace v překladové češtině.

Jazykové indikátory a formální operátory

Jak vyplývá z přehledu dosavadního výzkumu o simplifikaci v překladu (4.2.1), předmětem analýz bývá především **slovní zásoba překladových textů** a jejich skladba, především s ohledem na **srozumitelnost** a snadnost čtení **textu**. To má několik důvodů – předně lze předpokládat, že tendence zjednodušovat text se bude projevat právě na těchto dvou rovinách, neboť připouštějí největší variabilitu (např. na úrovni morfolgie nelze většinu prostředků nahrazovat jinými, jednoduššími). Dalším důvodem, proč badatelé zkoumají simplifikaci právě na těchto jazykových indikátorech, je bezesporu fakt, že je lze poměrně snadno operacionalizovat, kvantifikovat a měřit. Na druhou stranu je třeba brát v potaz, že nejčastěji používané testy (jako je TTR nebo *lexical richness*) jsou často náchylné na délku textu, nezohledňují syntagmatiku jazyka a jejich využití i vypovídací hodnota jsou tak omezené. Výběrem různých testů pro **výzkum simplifikace v překladové**

češtině jsem se pokusila tato rizika minimalizovat a zohlednit je při interpretaci výsledků. Shrnutí vybraných rysů a metod výpočtu uvádí následující tabulka:

<i>SIMPLIFIKACE</i>	jazykové indikátory	formální operátory
<i>lexikon</i>	bohatost slovní zásoby	poměr mezi typy a tokeny (zTTR) Yuleův koeficient K
	hutnost vyjadřování	lexikální hustota (<i>lexical density</i>) frekvenční špička (<i>list head</i>)
<i>syntax</i>	srozumitelnost textu	délka vět index srozumitelnosti (<i>readability index, RI</i>)

Operacionalizace hypotéz – simplifikace

Shrnutí výsledků

Co se týče bohatosti lexikonu a hutnosti vyjadřování, které byly testovány pomocí nové metody výpočtu zTTR, Yuleova koeficientu, LD a frekvenční špičky, lze konstatovat, že na základě výsledků třech operátorů ze čtyř (kromě Yuleova koeficientu) je možné zamítnout nulovou hypotézu a dojít k následujícímu závěru:

Poměr mezi typy a tokeny (měřený pomocí zTTR), výskyt autosémantik (měřený pomocí *lexical density*) je v překladové češtině nižší a objem slov z frekvenční špičky v korpusu je v překladové češtině vyšší, což nasvědčuje simplifikačním tendencím v překladové češtině. Yuleův koeficient hypotézu o simplifikaci nepotvrdil.

Z hlediska srozumitelnosti textu, která byla operacionalizována pomocí průměrné délky věty a indexu srozumitelnosti ARI, je možné zamítnout nulovou hypotézu ve všech bodech, platí tedy, že:

Průměrná délka věty a hodnota indexu srozumitelnosti (měřená testem ARI) je v překladové češtině nižší než v češtině nepřekladové.

Ve všech případech, kdy se simplifikační tendence potvrdily, byly rozdíly mezi překladovými a nepřekladovými texty statisticky signifikantní, ovšem s ohledem na velikost účinku nejde o příliš výrazné rozdíly. Trend je z hlediska všech prokázaných testů obdobný – překladová čeština se ve srovnání s nepřekladovou jeví jako simplifikovanější, ovšem rozdíly na těchto makrorovínách, jako je slovní zásoba jako celek nebo srozumitelnost textu, nejsou pro běžného čtenáře pravděpodobně vůbec postřehnutelné.

Pokud bychom měli tyto rozdíly nějak kvantifikovat a lépe ilustrovat, jako referenční hodnota se nabízí například **rozdíl mezi textovými typy**, tedy beletrií a odbornou literaturou. Ačkoli všechna měření probíhala odděleně pro každý textový typ zvlášť, následné porovnání výsledků jasně ukazuje, že větší rozdíly nalezneme mezi beletrií a odbornou literaturou než mezi překlady a nepřeklady. Příkladem

mohou být například naměřené hodnoty u testu *lexical density*, které se mezi textovými typy liší rozdílem až 6 procentních bodů (49,28, resp. 51,09 % v beletrii a 55,18, resp. 57,18 % v odborné literatuře), kdežto mezi překlady a nepřeklady jde o rozdíl jednoho až dvou procent.

Konvergence

Původně byla tato univerzálie označována jako *levelling-out*. Ve snaze vystihnout lépe její podstatu, zavedla Laviosová (2002: 71) nové označení konvergence a definovala ji jako tendenci překladových textů být si navzájem podobnější (a jako skupina homogennější) z hlediska určitých lingvistických rysů. Dobrým ukazatelem homogenosti skupiny může být např. rozptyl, na němž je z velké části založena analýza hlavních komponentů (PCA), což je metoda, která byla použita jako hlavní formální operátor.

Konvergence nepatří mezi nejoblíbenější témata korpusově-translatologických výzkumů. Olohanová (2004: 100) to přičítá tomu, že ji nelze tak snadno operationalizovat a měřit. Badatelé, kteří se jí ve svých pracích věnovali, se proto většinou zaměřili na podobné rysy, které využívají ve výzkumu jiné univerzálie (zpravidla simplifikace). Na podobnost či odlišnost překladových textů však mohou mít vliv i jiné rysy než ty úzce související s bohatostí lexikonu či srozumitelností textu. Z toho důvodu jsem rysy – vstupní data pro PCA – vybírala i na základě jiných kritérií, včetně informací z odborné literatury zabývající se srovnáním češtiny a angličtiny (jakožto nejvíce zastoupeným zdrojovým jazykem) a na základě zjištění, jež vyplynula z analýzy POS-gramů a n-gramů.

Jazykové indikátory a formální operátory

Jazykové indikátory, které tvoří vstupní data pro metodu PCA, jsou rozděleny do dvou souborů – soubor A zahrnoval vybrané rysy naměřené v rámci výzkumu simplifikace (např. délka věty, zTTR), kdežto soubor B se skládal z nově naměřených hodnot (např. počet sloves v minulém čase, počet částic, počet ukazovacích zájmen, viz 4.3.3). Cílem měření na souboru A bylo potvrdit či vyvrátit zjištění Lapshinové-Koltunski (2015), že pro identifikaci konvergence se nejlépe osvědčují měření provedená v rámci výzkumu simplifikace. Soubor B pak sloužil jako jakýsi ověřovací soubor dat, jenž měl ukázat, zda se konvergence projevuje i v případě měření, která přímo nesouvisejí se simplifikací. Cílem metody PCA je jednak snížení počtu původních proměnných, ale také identifikace těch proměnných, které nejlépe vystihují chování souboru a mohou nejvíce vypovídat o případných rozdílech mezi srovnávanými soubory.

Shrnutí výsledků

Na základě analýzy PCA lze postulovat několik závěrů. První se týká pozorovaných parametrů, které vstupovaly do PCA jako **původní proměnné**. V souboru A, který

zahrnoval čtyři hodnoty naměřené v rámci výzkumu simplifikace, se jako klíčový ukazatel variability v datech projevila především průměrná délka věty a dále hodnota zTTR. V souboru B, který obsahoval celkem 16 různých hodnot z oblasti morfologie, lexikologie i syntaxe, se jako nejdůležitější parametry (tvořící základ hlavních komponentů) ukázaly slovesné charakteristiky: počet sloves v minulém a přítomném čase a počet sloves v infinitivu.

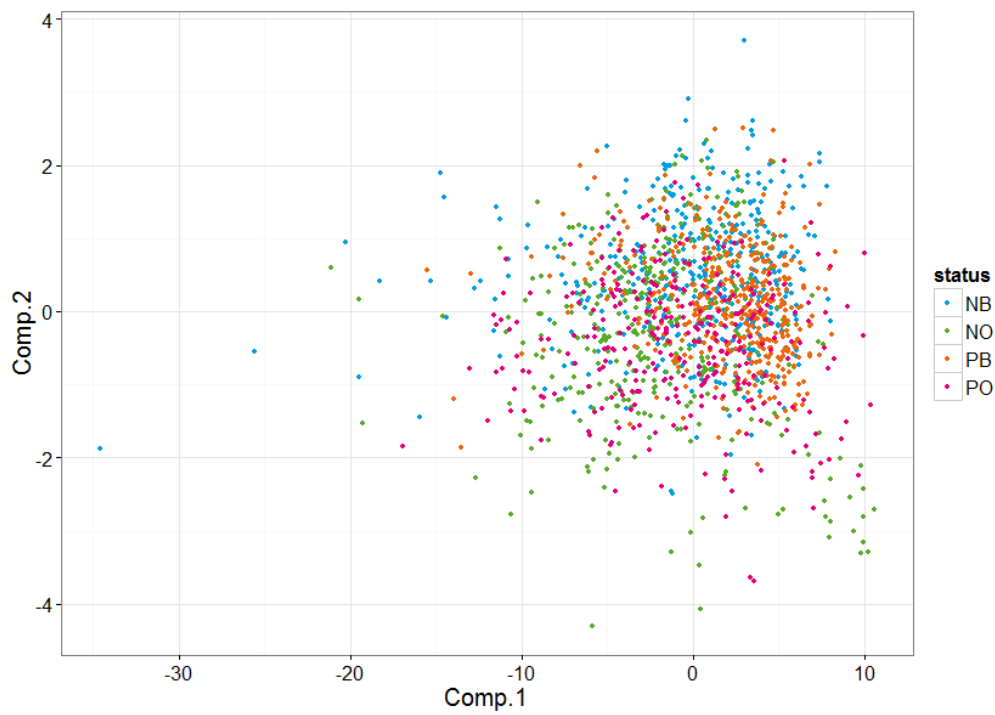
Druhý závěr se vztahuje k samotné výzkumné otázce, zda u překladových textů dochází ke konvergenci či nikoli. Samotné **testování hypotézy** proběhlo na základě PCA, přičemž odlišnost či podobnost překladů a nepřekladů z hlediska variability následně potvrdil provedený F-test, který slouží k porovnání rozptylů dvou vzorků. S ohledem na dílčí hypotézu tak platí následující výrok:

Beletristické překladové texty jsou si z hlediska metody PCA na základě uvedených rysů (souboru A a souboru B) podobnější než texty nepřekladové. V odborné literatuře na základě téže analýzy ke konvergenci u překladových textů nedochází.

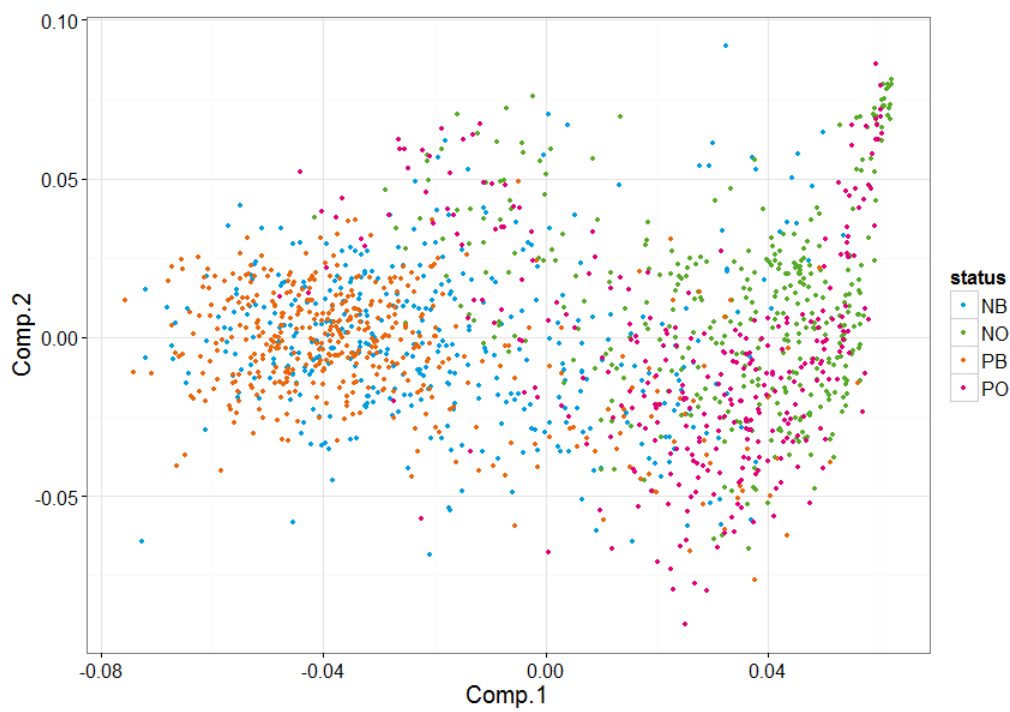
Dále se analýzou potvrdilo, že parametry získané v rámci výzkumu simplifikace, soubor A, je možné využít jako vstupní data pro PCA a že mohou být jedním z indikátorů konvergence. Totéž ovšem platí i pro rozsáhlejší, a tedy spolehlivější soubor B, který měl ukázat, zda se podobnost překladových textů z hlediska rozptylu projevuje i u jiných proměnných než u těch, jež se bezprostředně vztahují k simplifikaci. I tento předpoklad se tímto potvrdil.

A konečně třetí a poslední závěr lze formulovat s ohledem na **odlišnost výsledků v obou textových typech**. Metoda PCA opět poukázala na fakt, že texty beletristické mají jiné vlastnosti než texty odborné. Aby byl rozdíl co nejlépe patrný, provedla jsem pro doplnění PCA na celém korpusu Jerome bez žánrového rozlišení a výsledky graficky zobrazila v grafech na následující straně. V případě souboru A tvoří všechny texty dohromady jeden velký shluk, v němž lze jen stěží pozorovat určitou tendenci odborných textů, jež se vyskytují spíše ve spodní části grafu, zatímco beletristické texty nacházíme spíše nad hodnotou 0 osy y. Osa y však značí druhý hlavní komponent (PC2 tvořený hodnotou zTTR), který nemá takovou výpovědní hodnotu. Výsledek je i přesto konzistentní se závěry výzkumu simplifikace na základě zTTR, kde se rovněž projevuje rozdíl mezi beletrií a odbornou literaturou.

U souboru B je odlišný trend obou textových typů vidět mnohem zřetelněji. Zatímco beletristické texty (modrá, oranžová) se bez ohledu na status (překlad/nepřeklad) shlukují v levé části grafu, odborné (zelená, růžová) nacházíme vpravo, přičemž jakousi dělicí linií může být hodnota 0 na ose x. Hlavní komponenty tvoří jako v případě oddělených analýz opět především parametry MIN, INF a PRIT. Lze tak znovu konstatovat, že z pohledu této analýzy je rozdíl mezi textovými typy pro uživatele jazyka mnohem snáze zachytitelný než rozdíl mezi překladovými a nepřekladovými texty.



Výsledky analýzy PCA pro soubor A – celý korpus Jerome



Výsledky analýzy PCA pro soubor B – celý korpus Jerome

(Ne)typické slovní kombinace v překladech

Na rozdíl od předchozích dvou částí, které se zabývaly vždy jednou konkrétní překladovou univerzálií (simplifikací a konvergencí), je poslední část čtvrté kapitoly věnována obecnějšímu tématu slovních kombinací v překladovém jazyce. S ohledem na kapacitu této kapitoly je tato část spíše sondou než zevrubným popisem této problematiky. Translatologické studie týkající se specifických víceslovných jednotek v překladu se zpravidla vztahují k jedné ze dvou univerzálií – normalizaci a hypotéze o jedinečných jednotkách (*unique items hypothesis*), proto tato část zahrnuje i stručné představení posledního výzkumu v této oblasti (viz 4.4.1).

Cílem analýzy na základě n-gramů je identifikovat a blíže popsat takové slovní kombinace, které jsou pro překladový jazyk charakteristické – ať už se jedná o jevy v češtině typické (tedy časté, široce používané) nebo naopak o kombinace netypické, které v nepřekladové češtině nenacházíme tak často. Výraz **slovní kombinace** je zde přitom volen záměrně jako neutrální zastřešující termín (*word combination* podle Mauranen 2000: 131), jenž může zahrnovat kolokace, koligace, *lexical bundles* a další typy lexikálních struktur.

Bližší pozornost je věnována výběrově jen těm nejzajímavějším slovním kombinacím, které mají co nejvíce všeobecnou platnost (nejsou tolik podmíněny složením korpusu a konkrétním textem) a mohou poukazovat na odlišné trendy v překladovém jazyce. Jde především o víceslovné jednotky idiomatičké povahy⁶ (viz Tirkkonen-Conditt 2002), kombinace s textově strukturující funkcí (viz Mauranen 2000, Baker 2004) a kombinace poukazující na určitý specifický vzorec, který je vyděluje od ostatních 4-gramů (např. repetitivnost prvků, viz dále).

Shrnutí výsledků

Cílem této analýzy bylo odhalit případné odlišnosti mezi překlady a nepřeklady na základě srovnání 4-gramů. Ačkoli metoda analýzy n-gramů se snaží vycházet ze všech dat a poskytnout tak co možná nejobektivnější pohled, výběr konkrétních slovních kombinací k analýze je i přes předem stanovená kritéria nevyhnutelně subjektivní. S tímto vědomím je možné opatrně formulovat následující tendence, které lze v datech na základě zvoleného postupu vypořádat.

Nejprve je třeba konstatovat, že analýza n-gramů skutečně poukázala na dílčí rozdíly v použití konkrétních slovních kombinací, nelze však říci, že by se 4-gramy v překladovém jazyce na první pohled výrazně vymykaly. Na základě uvedených slovních kombinací, které tvoří poměrně různorodý celek, je poměrně obtížné formulovat nějaký obecný trend pro překladové texty. Zjevný je zde opět rozdíl mezi žánry (ve smyslu beletrie a odborné literatury), neboť v beletrii nalezneme na základě kritérií zejména slovní kombinace idiomatičké charakteru, kdežto v odborné literatuře jde převážně o metatextové formulace a pasivní konstrukce. Na základě analýzy nelze ani konstatovat, že by překladový jazyk vykazoval více či naopak méně idiomatičkých výrazů, spíše můžeme říci, že preferované frazémy se liší, což

⁶Slovní kombinace zde označuji za frazémy ve shodě se Slovníkem české frazeologie a idiomatiky (Čermák, Hronek & Machač 2009).

však může souviset i se složením korpusu a zaměřením textů. V překladu se prosazují spojení *pro všechno na světě, držet jazyk za zuby* či *ze všech sil*, zatímco v nepřekladové beletrii častěji narazíme na výrazy *do roka a do dne, od rána do noci* či *mezi nebem a zemí*.

Tendence vyhýbat se repetitivnosti se objevuje u struktur po sobě jdoucích adjektiv (např. *(stále) lepší a lepší, nové a nové*), ovšem v případě adverbii (*víc a víc, dál a dál* apod.) k odlišnému trendu nedochází, výskyt je zcela srovnatelný. Jistou tendenci rozlišovat mezi přibližnými synonymy (Mauranen 2000) můžeme pozorovat u výrazů *souviset s* a *mít co do činění s*, jež mají podobný význam, avšak každý ze zkoumaných souborů preferuje jiný z nich. Bez zdrojových textů můžeme opět pouze odhadovat, že vliv na preferenci *mít co do činění* v překladech může mít anglická konstrukce *have something to do with*, která překladatele patrně inspiruje k použití překladového ekvivalentu se stejným základovým slovesem.

Některé slovní kombinace také poukázaly na možné lexikální vzorce, které se v překladu odlišují: jako příklad může sloužit spojení se substantivem *doba*, jež je velmi frekventované v nepřekladové beletrii. Při dodatečném pohledu na jeho kolokace se ukázalo, že v překladu častěji narazíme na spojení *během doby*, zatímco v nepřekladových textech se objevují častěji *v průběhu doby* či *postupem doby*. Může jít o doslovný překlad anglického *during the time*, ovšem pro ověření této teorie bychom potřebovali analyzovat i zdrojové texty a zde použité fráze.

Tím se dostáváme k jednomu z klíčových závěrů. V okamžiku analýzy nižší jazykové roviny, jako jsou konkrétní lexikální vzorce či preference synonymických kombinací, se více než kdy jindy ukazuje zásadní nedostatek jednojazyčného srovnatelného korpusu, kterým je absence originálů. Bez zdrojových textů je velmi obtížné dostat se za hranici deskripce a spolehlivě vysvětlit, proč k výše zmíněným tendencím dochází, zda jde o vliv konkrétního jazykového jevu v konkrétním zdrojovém jazyce či o obecnou překladovou tendenci. Jedním z možných budoucích kroků by tak mělo být ověření zjištěných tendencí na reprezentativním paralelním korpusu, který by byl s to odhalit příčiny těchto jevů. Pozorované rozdíly tak prozatím slouží jako určité vodítko na dosud nepřilíh probádaném území víceslovných jednotek v překladu, jako jakési ukazatele na mapě, které naznačují, kde se z hlediska překladu odehrává něco potenciálně zajímavého.

5 Závěr

Cílem této disertace bylo zmapovat a charakterizovat češtinu v překladech na základě rozsáhlých korpusových dat a s využitím kvantitativních metod, tedy tak, jak v českém prostředí doposud systematicky zkoumána nebyla. K tématu překladového jazyka lze jistě přistupovat z mnoha různých úhlů a na základě četných teorií, jak bylo naznačeno v kapitole , jež shrnovala dosavadní výzkum v oblasti translatologie. Přístup využitý v této práci v sobě kombinuje možnosti a metody korpusové lingvistiky s nejnovějšími poznatky deskriptivní translatologie, především pak s ohledem na teorii překladových univerzálií.

Výhody kvantitativního přístupu jsou v případě zvoleného tématu nesporné: patří mezi ně především větší míra spolehlivosti a autentičnosti rozsáhlých dat, uplatnění statistických metod a snížení rizika subjektivního faktoru při výzkumu. Kombinace těchto rysů umožňuje badateli alespoň částečnou generalizaci výsledných zjištění, v tomto případě ve vztahu k překladové češtině jako celku, nikoli jen k několika málo konkrétním překladovým textům.

Ve všech výzkumech kvantitativní povahy hrají přitom klíčovou roli především dvě věci – **zvolená data**, především co do velikosti vzorku, a operacionalizace výzkumné otázky. Co se týče prvního bodu, jako datová základna pro tento výzkum posloužil korpus Jerome, rozsáhlý jednojazyčný srovnatelný korpus, který byl sestaven přímo za účelem srovnávání překladové a nepřekladové češtiny. Ačkoli jeho budování předcházela pečlivá příprava a volba kritérií, podle nichž byly texty do korpusu vybírány, v průběhu výzkumu se ukázaly i jeho limity, zvláště s ohledem na vliv určitých žánrů či textových typů (např. literatury faktu zařazené mezi beletrii). Snaha vykompenzovat převažující zdrojový jazyk textů (angličtinu) v podobě malého vyváženého subkorpusu se také ukázala být v určitých aspektech nedostatečná – malá velikost subkorpusu v řádu jednotek milionů totiž znemožňuje podrobnější statistické testování a srovnání výsledků s rozsáhlejšími daty. I přes tyto nedostatky však korpus Jerome poskytl dostatečný zdroj informací na to, abychom na základě testovaných hypotéz mohli konstatovat, jaká čeština v překladech je a co ji charakterizuje.

Druhým klíčovým bodem kvantitativního výzkumu je **operacionalizace výzkumných otázek**. Inspirací k výběru výzkumných hypotéz byly v případě této práce překladové univerzálie (Baker 1993) a jejich často velmi obecné definice. Bylo proto nutné je operacionalizovat, tedy zvolit konkrétní jazykové indikátory a formální metody jejich testování. Tento úkol s sebou kromě nevyhnutelného redukcionismu vždy nese jistý subjektivní faktor, neboť není prakticky možné předem určit, zda byla operacionalizace přiměřená či nikoli. Hlavním cílem při výběru ja-

zykových indikátorů jednotlivých testovaných rysů proto bylo zahrnout co možná nejvíce různých ukazatelů a testů a tím eliminovat riziko nevhodně zvolené metody. I přes tato opatření je však třeba brát v potaz, že zvolené operacionalizace jsou zcela jistě jen jedny z možných, a s tímto vědomím přistupovat k interpretaci výsledků.

Shrneme-li tedy nejpodstatnější závěry, které z výzkumu vyplynuly, dojdeme k následujícím charakteristickým rysům překladové češtiny, jak je zachycena v korpusu Jerome. Z pohledu **simplifikace**, za jejíž možné indikátory zde byly považovány menší bohatost a pestrost lexikonu, menší hutnost vyjadřování a větší srozumitelnost textu, se překladová čeština jeví na základě pěti ze šesti testů skutečně nepatrně jednodušší, např. obsahuje kratší věty, častěji používá velmi frekventovaná slova a zahrnuje méně autosémantik než čeština nepřekladová. Všechny rozdíly se ukázaly být statisticky signifikantní, ovšem z hlediska věcné významnosti nikterak dramatické. Obdobný výsledek pozorujeme i v případě porovnání frekvenční distribuce slovních druhů a jejich kombinací, u nichž lze rovněž vysledovat odlišné tendence (např. překlady obsahují v beletrii i odborné literatuře přibližně o 2 % více substantiv a o 2 % méně verb), ovšem tento rozdíl nebude pro čtenáře patrně vůbec postřehnutelný.

Ve shodě s hypotézou další zkoumané univerzálie **konvergence** lze říci, že beletristické překladové texty jsou si na základě provedené metody PCA podobnější než texty nepřekladové, jinými slovy mají tendenci dosahovat v testech obdobných hodnot. V odborné literatuře na základě téže analýzy ke konvergenci u překladových textů nedochází. Stejně jako v případě simplifikace lze výraznější rozdíl pozorovat mezi beletrii a odbornou literaturou než mezi překlady a nepřeklady. Závěrečná **analýza n-gramů** pak poukázala na dílčí odlišnosti v překladové češtině, především v preferenci idiomatických výrazů a přibližných synonym, a znovu upozornila na zásadní vliv složení korpusu.

Zbývá tedy celkově vyhodnotit **platnost výchozí hypotézy**. S vědomím všech výše zmíněných faktorů je možné konstatovat, že překladová čeština se od nepřekladové češtiny skutečně liší, ale hned vzápětí je třeba dodat, že odhalené rozdíly zdaleka nejsou tak výrazné a zásadní, jak by se na základě formulovaných hypotéz a předchozích translatických prací mohlo zdát. Vzhledem k tomu, že se u obou zkoumaných souborů (překladů i nepřekladů) jedná stále o tentýž jazyk, o češtinu, nelze ani očekávat, že rozdíly budou na první pohled zarážející, je tedy poměrně obtížné vyjádřit, jak výrazné ony odlišnosti v překladové češtině jsou.

Vhodným referenčním rámcem zde však může být **rozdíl mezi textovými typy**, beletrii a odbornou literaturou, které v rámci tohoto výzkumu prošly stejným testováním, a je tedy možné výsledky navzájem porovnávat. V takřka všech provedených testech se rozdíly mezi beletrii a odbornou literaturou ukázaly být jako výraznější a prominentnější než v případě srovnání překladů a nepřekladů v rámci žánru. Nejlepším názorným příkladem v tomto ohledu byla metoda PCA, která na základě zvolených rysů seskupila dohromady nikoli překlady a nepřeklady, ale spíše beletrii a odbornou literaturu. Obdobné výsledky pozorujeme u všech provedených testů. Je tedy pravděpodobné, že čtenář na základě zvolených jazykových indikátorů mnohem lépe postřehne rozdíl mezi beletrii a odbornou literaturou než mezi překlady a nepřeklady.

To samozřejmě neznamená, že by překladová čeština nedisponovala i takovými rysy, které ji mohou do jisté míry identifikovat – nikoli však na vyšších rovinách zobecnění (jaké zde byly především zkoumány), ale spíše na úrovni konkrétních slovních kombinací a lexikálních vzorců v rámci textu. Dá se přitom konstatovat, že čím níže badatel při výzkumu sestupuje, tím větší roli zde hraje složení korpusu (především z hlediska žánru, zdrojového jazyka apod.). Především při výzkumu typických slovních kombinací na základě n-gramů se tak ukázalo, že k lepší interpretaci zjištěných rozdílů by bylo zapotřebí analyzovat kromě referenčních nepřekladových textů také texty zdrojové, tedy originály zahrnutých překladů. Těmi však jednojazyčný srovnatelný korpus nedisponuje, proto tato oblast zůstává otevřena budoucím badatelům.

S posledním bodem úzce souvisí **možnosti budoucího výzkumu** v oblasti překladové češtiny a rysů překladového jazyka. Je nesporné, že využití reprezentativního paralelního či ještě lépe recipročního korpusu by přineslo nové poznatky (především v oblasti S-univerzálií) a zároveň umožnilo přesnější interpretaci zde zjištěných fakt. Jednojazyčný srovnatelný korpus Jerome má z hlediska svého složení přes veškerou snahu také své limity, což platí i pro jeho subkorpus vyvážený z hlediska zdrojového jazyka. Zvláště v případě subkorpusu by bylo vhodné shromáždit větší množství překladů z více jazyků, aby jeho výsledná velikost umožnila pokročilejší statistické testování a ověřování možného vlivu interference.

Z hlediska témat, jež se ukázala jako potenciální zdroj dalších informací o překladové češtině, ale vzhledem k rozsahu práce zde nemohla být dále rozvíjena, stojí za zmínku především otázka víceslovných jednotek v překladu. Zde provedená analýza naznačila, že na úrovni konkrétních lexikálních jednotek a jejich kombinací může docházet k rozdílům, jež by si zasloužily další výzkum. Jednou z možností, jak víceslovné jednotky v překladu zkoumat, přitom může být kvantitativní analýza variability kontextu (viz Cvrček 2013) nebo dílčí kvalitativní studie věnované konkrétním kolokacím v překladové a nepřekladové češtině.

Závěrem je možné konstatovat, že čeština v překladech představuje (nejen) pro korpusovětranslatologický výzkum bohatý inspirační zdroj, který touto disertační prací zdaleka nebyl vyčerpán. Ačkoli výsledky naznačily, že překladová čeština není zcela odlišná od češtiny v původně českých dílech, současně s tím ukázaly, že i přesto má lingvistům a translatologům z výzkumného hlediska mnoho co nabídnout.

Kompletní seznam literatury

- Aijmer, K., Altenberg, B. & Johansson, S. (Eds.) (1996). *Languages in Contrast: Papers from a Symposium on text-based Cross-linguistics Studies*. Lund: Lund University Press.
- Anderman, G. & Rogers, M. (2008). *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). Amsterdam-Philadelphia: John Benjamins.
- Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7(2), 223–243.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in language engineering, in honour of Juan C. Sager* (pp. 175–86). Amsterdam: John Benjamins.
- Baker, M. (2004). A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics*, 9(2), 167–193.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baroni, M. & Bernardini, S. (2006). A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text. *Literary and Linguistic Computing*, 21(3), 259–274.
- Becher, V. (2010). Abandoning the Notion of "Translation-Inherent" Explicitation. Against a Dogma of Translation Studies. *Across Languages and Cultures*, 11(1), 1–28.
- Bernardini, S. (2007). Collocations in Translated Language. Combining parallel, comparable and reference corpora. Příspěvek přednesený na konferenci *Corpus Linguistics 2007*, Lancaster, UK. Dostupný z: http://ucrel.lancs.ac.uk/publications/CL2007/paper/15_Paper.pdf
- Bernardini, S. & Zanettin, F. (2004). When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* (pp. 53–62), Amsterdam-Philadelphia: John Benjamins.
- Biber, D. & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

- Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17–35). Tübingen: Gunter Narr.
- Cantos Gómez, P. (2013). *Statistical Methods in Language and Linguistic Research*. Sheffield: Equinox.
- Catford, J. C. (1965). *A Linguistics Theory of Translation*. London: Oxford University Press.
- Corpas Pastor, G., Mitkov, R., Afzal, N. & Pekar, V. (2008). Translation universals: Do they exist? A corpus-based NLP study of convergence and simplification. *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas (AMTA-08)*. Waikiki Hawaii.
- Cvrček, V. (2013). *Kvantitativní analýza kontextu*. Praha: Nakladatelství Lidové noviny.
- Cvrček, V. et al. (2010). *Mluvnice současné češtiny*. Praha: Nakladatelství Karolinum.
- Cvrček V., Čermák, F. & Křen M. (2007). Statistické aspekty jazyka Karla Čapka, zvláště jeho lexikonu. In F. Čermák (Ed.), *Slovník Karla Čapka* (pp. 673–690). Praha: Nakladatelství Lidové noviny.
- Cvrček, V. & Fidler, M. (v tisku). A data-driven analysis of reader viewpoints: Reconstructing the historical reader using keyword analysis. *Journal of Slavic Linguistics*.
- Cvrček, V. & Chlumská, L. (v tisku). Simplification in translated Czech: a new approach to type-token ratio. *Russian Linguistics*, 39(3).
- Cvrček, V. & Kovářiková, D. (2011). Možnosti a meze korpusové lingvistiky. *Naše řeč*, 94(3), 113–133.
- Cvrček, V. & Václavík, J. (v tisku). Jednoznačnost a kontext. *Kvantitativní studie. Korpus – gramatika – axiologie*.
- Čermák, F., Hronek J. & Machač, J. (Eds.) (2007). *Slovník české frazeologie a idiomatiky 1–4 (SČFI)*. Praha: LEDA.
- Dayrell, C. (2007). A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics*, 12(3), 375–414.
- De Sutter, G., Cappelle, B. & Loock, R. (2013). Competing Motivations in Dutch/French Legal Translation: a Quantitative Corpus-Based Study of the Interaction between Interference and Normalisation. Příspěvek přednesený na konferenci *ICLC 7 – UCCTS 3*, 11–13. 7. 2013, Gent, Belgie.
- De Sutter, G., Goethals, P., Leuschner, T. & Vandepitte, S. (2012). Towards methodologically more rigorous corpus-based translation studies. *Across Languages and Cultures*, 13(2), 137–143.
- Duguid, A. (2010). Newspapers discourse informalisation: a diachronic comparison from keywords. *Corpora*, 5(2), 109–138.
- Dušková, L. (1988). *Mluvnice současné angličtiny na pozadí češtiny*. Praha: Academia.

- Fernandes, L. (2006). Corpora in Translation Studies: revisiting Baker's typology. *Fragmentos*, 30, 87–95.
- Frawley, W. (1984). *Prolegomenon to a Theory of Translation Translation: literary, linguistic and philosophical perspectives*. Newark: University of Delaware Press.
- Friedbichler, I. & Friedbichler, M. (1997). The potential of domain-specific target-language corpora for the translator's workbench. Příspěvek přednesený na konferenci *Conference on Corpus Use and Learning to Translate*, Bertinoro.
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin & H. Lindquist (Eds.), *Translation Studies in Scandinavia* (pp. 88–95). Lund: CWK Gleerup.
- Gentzler, E. (1993). *Contemporary Translation Theories*. London and New York: Routledge.
- Goethals, P. (2007). Corpus-driven Hypothesis Generation in Translation Studies, Contrastive Linguistics and Text Linguistics: A Case Study of Demonstratives in Spanish and Dutch Parallel Texts. *Belgian Journal of Linguistics*, 21, 87–103.
- Grabowski, Ł. (2012). On translation universals in selected contemporary Polish literary translations. *Studies in Polish Linguistics*, 7(1), 165–183.
- Granger, S. (1996). From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast: Papers from a Symposium on text-based Cross-linguistics Studies* (1994) (pp. 38–51). Lund: Lund University Press.
- Granger, S. (2013). Tracking the third code: A crosslinguistic corpus-driven approach to discourse markers. Příspěvek přednesený na konferenci *ICLC 7 - UCCTS 3*, 11–13. 7. 2013, Gent, Belgie.
- Halverson, S. (2003). The cognitive basis of translation universals. *Target*, 15(2), 197–241.
- Hareide, L. & Hofland, K. (2012). Compiling a Norwegian-Spanish Parallel Corpus: methods and challenges. In Michael P. Oakes & Meng Ji (Eds.), *Quantitative Methods in Corpus-Based Translation Studies* (pp. 75–113). Amsterdam: John Benjamins.
- Hendl, J. (2009). *Přehled statistických metod*. Praha: Portál.
- Hermans, T. (1999). *Translation in Systems*. Manchester: St. Jerome.
- Holmes, J. S. (1972). The name and nature of translation studies. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 172–185). London and New York: Routledge.
- House, J. (1997). *Translation Quality Assessment: A Model Revisited*. Tübingen: Gunter Narr.
- House, J. (2008). Beyond Intervention: Universals in Translation? *trans-kom: Zeitschrift für Translationswissenschaft und Fachkommunikation*, 1(1), 6–19.
- Hrala, M. et al. (2002). *Kapitoly z dějin českého překladu*. Praha: Karolinum.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Chesterman, A. (2004a). Hypotheses about translation universals. In G. Hanse, K. Malmkjær & D. Gile (Eds.), *Claims, Changes and Challenges in Translation*

- Studies. Selected Contributions from the EST Congress Copenhagen 2001.* (pp. 1–14). Amsterdam-Philadelphia: John Benjamins.
- Chesterman, A. (2004b). Beyond the Particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* (pp. 33–49). Amsterdam-Philadelphia: John Benjamins.
- Chlumská, L. (2013). JEROME: jednojazyčný srovnatelný korpus pro výzkum překladové češtiny. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z <http://www.korpus.cz>.
- Chlumská, L. (2014). Není *korpus* jako *korpus*. Korpusy v kontrastivní lingvistice a translatologii. *Časopis pro moderní filologii*, 96(2), 221–232.
- Chlumská, L. & Kovářiková, D. (2009). The reflection of linguistics tradition in translation. In F. Čermák, P. Corness & A. Klégr (Eds.), *InterCorp: Exploring a Multilingual Corpus* (pp. 146–158). Praha: Nakladatelství Lidové noviny.
- Chlumská, L. & Richterová, O. (2014). Jak zkoumat překladovou češtinu. Výzkum simplifikace na korpusu Jerome. *Korpus – gramatika – axiologie*, 9, 16–29.
- Ilisei, I., Inkpen, D., Corpas Pastor, G. & Mitkov, R. (2010). Identification of Translationese: A Machine Learning Approach. Příspěvek přednesený na konferenci Cicling 2010.
- Jakobson, R. (1959/2000). On linguistic aspects of translation. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 113–118). London and New York: Routledge.
- Johansson, S. & Oksenfjell, S. (1998). *Corpora and Cross-linguistic Research*. Amsterdam-Atlanta: Rodopi.
- Johansson, S. (1998). On the role of corpora in cross-linguistic research. In S. Johansson & S. Oksefjell (Eds.), *Corpora and Cross-linguistics Research* (pp. 3–24). Amsterdam-Atlanta: Rodopi.
- Johansson, S. (2004). Why change the subject? On changes in subject selection in translation from English into Norwegian. *Target*, 16(1), 29–52.
- Kenny, D. (1998). Creatures of Habit? What Translators Usually Do with Words. *Meta*, 43(4), 515–523.
- Kenny, D. (2001). *Lexis and Creativity in Translation*. A Corpus-based Study. Manchester: St. Jerome.
- Kočová, I. (2009). *Legis Speak and Creative Translation Solutions in EU Texts: Strictly Legislative Versus Related Texts* (Diplomová práce). Vedoucí DP: Mgr. Renata Kamenická, Ph.D. Brno: FF MU.
- Kruger, A. (2002). Corpus-based translation research: Its development and implications for general, literary and Bible translation. *Acta Theologica Supplementum*, 2, 70–106.
- Kubáčková, J. (2008). *Generalizace a specifikace lexikálního významu v soudobém uměleckém překladu* (Diplomová práce). Vedoucí DP: PhDr. Zuzana Jettmarová, M.Sc. Praha: FF UK.
- Lapshinova-Koltunski, E. (2015). Variation in translation: evidence from corpora. In C. Fantionuoli & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (pp. 93–114). Berlin: Language Science Press.

- Laviosa, S. (1998a). The English Comparable Corpus. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in Diversity?*. Manchester: St. Jerome Publishing.
- Laviosa, S. (1998b). Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta: Translator's Journal*, 43(4), 557–571.
- Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, findings, applications*. Amsterdam-New York: Rodopi.
- Laviosa-Braithwaite, S. (1996). *Investigating Simplification in English Comparable Corpus of Newspaper Articles*. Szombathely: Daniel Berzsenyi College Printing Press.
- Levý, J. (1971). Bude teorie překladu užitečná překladatelům? *Bude literární věda exaktní vědou?* (pp. 147–157). Praha: Československý spisovatel.
- Levý, J. (1983). *Umění překladu*. Praha: Panorama.
- Malmkjær, K. (1997). Punctuation in Hans Christian Andersen's Stories and in their Translations into English. In F. Poyatos (Ed.), *Nonverbal Communication and Translation: New Perspectives and Changes in Literature, Interpretation and the Media* (pp. 151–162). Amsterdam: John Benjamins.
- Malmkjær, K. (1998). Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators? *Meta*, 43(4), 534–541.
- Mathesius, V. (1947). Přívlastkové ten, ta, to v hovorové češtině. In *Čeština a obecný jazykozpyt* (pp. 185–189). Praha: Melantric.
- Mauranen, A. (2000). Strange Strings in Translated Language. A Study on Corpora. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome Publishing.
- Mauranen, A. & Kujamäki, P. (Eds.) (2004). *Translation Universals - Do They Exist?* Amsterdam-Philadelphia: John Benjamins.
- May, R. (1997). Sensible Elocution: How Translation Works in & upon Punctuation. *The Translator*, 3(1), 1–20.
- McEnergy, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- McEnergy, T., Xiao R. & Tono, Y. (2006). *Corpus-based Language Studie: An Advanced Resource Book*. London and New York: Routledge.
- McEnergy, T. & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Munday, J. (2008). *Introducing Translation Studies. Theories and applications*. London: Routledge.
- Neumann, S. (2014). Beyond translation properties: The contribution of corpus studies to empirical translation theory. Plenární přednáška přednesená na konferenci *UCCTS 4*, Lancaster, UK, 25.7.2014.
- Newmark, P. (1981). *Approaches to Translation*. Oxford and New York: Pergamon.
- Newmark, P. (1988). *A Textbook of Translation*. New York and London: Prentice-Hall.
- Nida, E. (1964). *Towards a science of translating*. Leiden: E. J. Brill.

- Nida, E. & Taber, C. R. (1969). *The Theory and Practise of Translation*. Leiden: E. J. Brill.
- Nord, Ch. (1988). *Textanalyse und Übersetzen: Theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. Heidelberg: J. Groos.
- Oakes, Michael P. & Ji, M. (2012). *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam-Philadelphia: John Benjamins.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London: Routledge.
- Øverås, L. (1998). In Search of the Third Code: An Investigation of Norms in Literary Translation. *Meta: Translator's Journal*, 43(4), 557–570.
- Paloposki, O. (2001). Enriching translations, simplified language? An alternative viewpoint to lexical simplification. *Target*, 13(2), 265–288.
- Partington, A. (2010). Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers – an overview of the project. *Corpora*, 5(2), 83–108.
- Polišenská, M. (2010). *Translation Universals in the English and Spanish Translations of Saturnin by Zdeněk Jirotka* (Diplomová práce). Vedoucí DP: Mgr. Renata Kamenická, Ph.D. Brno: FF MU.
- Popovič, A. (1975). *Teória umeleckého prekladu: aspekty textu a literárnej metakomunikácie*. Bratislava: Tatran.
- Puurtinen, T. (2003). Genre-specific Features of Translationese? Linguistic Differences between Translated and Non-translated Finnish Children's Literature. *Literary and Linguistic Computing*, 18(4), 389–406.
- Pym, A. (2008). On Toury's laws of how translators translate. In A. Pym, M. Shlesinger & D. Simeoni (Eds.), *Beyond Descriptive Translation Studies*. Amsterdam-Philadelphia: John Benjamins.
- Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. München: Max Hueber.
- Rodríguez-Castro, M. (2011). Translationese and punctuation. *Translation and Interpreting Studies*, 6(1), 40–61.
- Savický, P. & Hlaváčová, J. (2003). Measures of word commonness. *Journal of Quantitative Linguistics*, 9(3), 215–231.
- Shamaa, N. (1978). *A Linguistic Analysis of Some Problems of Arabic to English Translation* (Dizertační práce). Oxford: Oxford University.
- Shlesinger, M. (1989). *Simultaneous Interpretation as a Factor in Effecting Shifts in the Position of Texts on the Oral-Literate Continuum* (Diplomová práce). Tel Aviv: Tel Aviv University.
- Shlesinger, M. (1991). Interpreter Latitude vs. Due Process. Simultaneous and Consecutive Interpretation in Multilingual Trials. In S. Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies* (pp. 147–155). Tübingen: Gunter Narr.
- Smith, E. A. & Senter, R. J. (1967). *Automated Readability Index*. Ohio: Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Paterson Air Force Base.
- Soukup, P. (2013). Věcná významnost výsledků a její možnosti měření. *Data a výzkum – SDA Info 2013*, 7(2), 125–148.

- Středová, A. (2009). *Explicitation and Implication in Non-literary Translations* (Diplomová práce). Vedoucí DP: Mgr. Renata Kamenická, Ph.D. Brno: FF MU.
- Stubbs, M. (1986). *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.
- Teich, E. (2003). *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: Mouton de Gruyter.
- Thompson, P. & Sealey, A. (2007). Through children's eyes? Corpus evidence of the features of children's literature Vol. 12 No. 1 1–12
- Tirkkonen-Condit, S. (2000). In Search of Translation Universals: Non-equivalence or Unique Items in a Corpus test. Příspěvek přednesený na konferenci *UMIS-T/UCL Research Models in Translation Studies*, Manchester, 28.–30. 4. 2000.
- Tirkkonen-Condit, S. (2002). Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target*, 14(2), 207–220.
- Tirkkonen-Condit, S. (2004). Unique Items? Over- or Under-Represented in Translated Language? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* (pp. 177–185). Amsterdam-Philadelphia: John Benjamins.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam-Philadelphia: John Benjamins.
- Toury, G. (1980). *In Search of a Theory of Translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.
- Toury, G. (1995). *Descriptive Translation Studies ?- and Beyond*. Amsterdam-Philadelphia: John Benjamins.
- Toury, G. (2004a). Probabilistic Explanations in Translation Studies: Universals – or a Challenge to the Very Concept? In G. Hanse, K. Malmkj(æ)r & D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies. Selected contributions from the EST Congress Copenhagen 2001* (pp. 15–25). Amsterdam-Philadelphia: John Benjamins.
- Toury, G. (2004b). Probabilistic Explanations in Translation Studies: Welcome as they are, would they qualify as universal? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals – Do They Exist?* (pp. 15–32). Amsterdam-Philadelphia: John Benjamins.
- Tymoczko, M. (1998). Computerized Corpora and the Future of Translation Studies. *Meta: Translator's Journal*, 43(4), 652–660.
- Vanderauwera, R. (1985). *Dutch Novels Translated into English: The Transformation of a "Minority" Literature*. Amsterdam: Rodopi.
- Venuti, L. (Ed.) (2000). *The Translation Studies Reader*. London and New York: Routledge.
- Vermeer, H. & Reiss, K. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- Vinay, J.-P. & Darbelnet, J. (1958). *Stylistique comparée du français et de l'anglais: méthode de traduction*. Paris: Didier.
- Volín, J. (2007). *Statistické metody ve fonetickém výzkumu*. Praha: Epocha.

- Wang, K. & Qin, H. (2010). A Parallel Corpus-based Study of Translational Chinese. In R. Xiao (Ed.), *Using Corpora in Contrastive and Translation Studies* (pp. 164–181). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Wang, W. (2006). *A corpus-driven study on translation units in an English-Chinese parallel corpus* (Diplomová práce). University of Birmingham.
- Xiao, R. (2010). How different is translated Chinese from native Chinese? *International Journal of Corpus Linguistics*, 15(1), 5–35.
- Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
- Zanettin, F. (2011). Translation and corpus design. *SYNAPS – A Journal of professional Communication*, 26, 14–23.
- Zanettin, F. (2013). Corpus Methods for Descriptive Translation Studies. *Procedia – Social and Behavioral Sciences*, 95, 20–32.
- Zubatý, J. (1917). Ten. *Naše řeč*, 1(10), 289–293.
- Zubatý, J. (1920). Ten nejlepší člověk. *Naše řeč*, 4(3), 74–76.