

HYPOTÉZA *UNIQUE ITEMS* V PŘEKLADU Z ANGLIČTINY

ADÉLA ŠPÍNOVÁ

ABSTRACT

This article aims to test the so-called unique items hypothesis on Czech language data. The hypothesis formulated by Sonja Tirkkonen-Condit presumes the underrepresentation of unique items (target-language elements that have no direct counterparts in the source language) in translated texts compared to non-translations in the same language. A monolingual comparable sample corpus consisting of Czech translations and non-translations served as language data, both subcorpora containing approximately 17 million tokens. The tested Czech unique items were chosen from lexical units, word-formation phenomena, syntactic structures and language use phenomena. The frequency of these items was subjected to statistical testing with R software. The results reveal a general tendency of translated Czech language to contain less unique items. However, some of the individual items do not correspond to this tendency.

Key words: unique items hypothesis, corpus-based translation studies, comparable monolingual corpus, translated Czech

1. Úvod: hypotéza *unique items* jako možná univerzálie

Hypotézu „jedinečných prvků“ (*unique items hypothesis*) jako možnou překladovou univerzálii formulovala a pojmenovala Sonja Tirkkonen-Conditová v článku „Translatiōnese – a myth or an empirical fact?“ (Tirkkonen-Condit 2002), kde popisuje experiment, v němž měli čtenáři označovat texty buď za originální, nebo za přeložené. Studie se týkala přeložené a nepřeložené finštiny a respondenti v ní svá rozhodnutí zdůvodňovali i slovním komentářem.

Tirkkonen-Conditová uvádí nepřítomnost jedinečných prvků jako hlavní znak, podle kterého byly texty označovány za překlady (ať už správně či nesprávně). Proto usoudila, že tyto tzv. *unique items* představují pro výzkum překladového jazyka relevantní jev a následně provedla i korpusovou studii zaměřenou na ověření hypotézy jedinečných prvků (Tirkkonen-Condit 2004).

Tendenci překladatelů nevyužívat plně potenciál cílového jazyka v případě, že pro určité prvky neexistuje přímá opora ve zdrojovém jazyce, formuloval ovšem již Jiří Levý (1963: 73): „Nebude-li překladatel užívat specifických prostředků svého jazyka, pro které

nemůže být dán podklad v originále, bude výrazová škála překladu chudší, než bývá v původní literatuře, která je sepsána v jazyce překladu.“ Levý zde zmiňuje Güttingerův text *Zielsprache* (Güttinger 1963), z něž nepřímo, přes práci Reissové (1971), vyšla i Tirkkonen-Condítová.

Güttinger (1963: 219) navrhuje testovat kompetenci překladatelů tak, že se ověří frekvence nejběžnějších slov cílového jazyka, která ve zdrojovém jazyce neexistují. Tato „chybějící slova“ mají ukázat, nakolik překladatel ovládá cílový jazyk. Podle Reissové (1971: 19) se pak tento jev netýká pouze slov, ale veškerých jazykových prostředků, které ve zdrojovém jazyce nefungují stejným způsobem jako v jazyce cílovém.

Hypotéza jedinečných prvků tedy sice porovnává překlady a nepřeklady v tomtéž jazyce, je ovšem závislá na specifické kombinaci zdrojového a cílového jazyka; funguje vlastně jako jakási „negativní interference“. Zároveň také odporuje hypotéze „konvencionalizace“ v tom smyslu, že překlady namísto postulované „normality“ vykazují oproti nepřekladům nižší, tedy nenormální frekvence určitých specifických jevů.

Z výše uvedeného je patrné, že hypotéza jedinečných prvků jako taková není nijak nová, její rozsáhlejší kvantitativní výzkum ovšem umožňují teprve jednojazyčné srovnatelné korpusy, které začaly vznikat až v posledních desetiletích a pro řadu jazyků nejsou dosud k dispozici vůbec. Korpusové studie zaměřené na hypotézu jedinečných prvků vznikají zatím spíše ojediněle a nemají jednotné výsledky.

Hypotéza samotná byla také kritizována z teoretického hlediska (Chesterman 2007, Kolehmainen 2013). Chesterman hypotézu jedinečných prvků podrobil detailní kritice v článku „What is a unique item?“ Upozorňuje především na nedostatečně specifické vymezení „jedinečnosti“ v podání Tirkkonen-Condítové. Podle Chestermana (2007: 6) navíc některé prvky jsou „jedinečnější“ než jiné: mezi prvky zdrojového a cílového jazyka může být různá míra podobnosti a tato podobnost může mít mnoho různých podob. Jedinečnost je tak pouze relativní pojem a znamená vlastně určitou škálu „relativní nepodobnosti“ prvků v obou jazycích.

Chesterman (2007: 7) proto navrhuje definovat „jedinečný prvek“ jako takový, který nelze zpátky do zdrojového jazyka jednoduše přeložit bez toho, co Catford (1965: 79) nazývá *unit shift*, tedy bez posunu mezi strukturními jednotkami jazyka.

Kolehmainenová (2013) kritizuje předpokládaný status překladové univerzálie přisuzovaný hypotéze jedinečných prvků. Argumentuje tím, že nižší výskyt takzvaných jedinečných prvků není univerzálním jevem, který by se vyskytoval ve všech překladech, a zároveň ani není unikátní pro překlady. Objevuje se i v jiných situacích jazykového kontaktu, např. v řeči nerodilých mluvčích. Tato skutečnost ovšem nijak nesnižuje zajímavost samotné *unique items hypothesis* pro translátologický výzkum.

2. Metodika výzkumu

2.1 Výběr českých unique items v kontrastu s angličtinou

V tomto výzkumu byl zvolen postup podle Tirkkonen-Condítové, tedy přístup *corpus-based* (na korpusu založený, nikoli korpusem řízený) s výběrem potencionálních jedinečných prvků založeným na intuici a kontrastivní analýze.

Pro výběr kandidátních jazykových jevů byly předem stanoveny následující orientační principy: nemělo by jít o prvky příliš závislé na obsahové stránce textů zahrnutých do korpusu; nemělo by jít o prvky příliš řídké; vybrané prvky by měly být co možná nejméně ovlivněné idiolektem autorů a překladatelů; měly by být zastoupeny různé jazykové roviny (lexikální jednotky, jevy z oblasti slovtvorby, syntaxe, uzuální jevy); pokud je to možné, mělo by spíše než o jednotlivé prvky jít o skupiny prvků, které lze sdružit podle nějakého rysu (tento požadavek souvisí s prvními třemi výše uvedenými principy); všechny vybrané prvky musí být nějakým smysluplným způsobem vyhledatelné v korpusu.

Většinu zkoumaných *unique items* tvoří slovtvorné jevy. Do výzkumu byla zahrnuta předponová slovesa, zesilující předpony *pra-* a *vele-*, zesilující adjektivní přípony *-ičký*, *-oučký*, *-inký*, *-ounký*, zájmeno *tentam*, zájmena *kdopak*, *copak*, *kterýpak*, *jakýpak* a *čipak*, číslovka *nejeden* a adverbium *nejednou*, řadové číslovky *kolikátý*, *tolikátý* a *několikátý* a druhové číslovky.

Z lexikální roviny byly vybrány lexémy *prý*, *totiž*, *čili* a *nikoli*; žádný z nich by neměl být výrazně závislý na tématu textů. Ze syntaktických jevů potom byly do výzkumu zahrnuty dva: slovosled věty s předmětem na první pozici a bezpodmětová syntaktická konstrukce – věta jednočlenná slovesná. Posledním typem jedinečných prvků jsou jevy uzuální – konstrukce typu *pan učitel*, *paní doktorka*, *slečna magistra*, tedy oslovení či označování osob s určitou profesí, postavením nebo titulem bez uvedení vlastního jména.

2.2 Jazykový materiál

Použitý jazykový materiál pochází z jednojazyčného srovnatelného korpusu Jerome, který vznikl právě za účelem výzkumu současné překladové češtiny. Tento korpus je lemmatizovaný, morfologicky značkový a obsahuje také strukturní atributy včetně např. pohlaví autora a překladatele textu.

Pro účely analýzy četnosti jedinečných prvků byly v rámci korpusu Jerome vytvořeny dva srovnatelné subkorpusy. Oba obsahují pouze beletristické texty, a to romány (NOV) a kratší prózy (COL). Literatura faktu (FAC) nebyla do použitých subkorpusů zařazena na doporučení autorky korpusu Jerome Lucie Chlumské,¹ neboť se svými vlastnostmi ostatním textům zařazeným do kategorie beletrie vymykají a mohly by zkreslovat výsledky. Subkorpus nepřekladových textů z korpusu Jerome má výslednou velikost 16 809 069 pozic, subkorpus českých překladů z angličtiny pak 17 323 495 pozic.

Kvůli zahrnutí jevu, který lze vyhledávat jedinečně v syntakticky značkováném korpusu, bylo nutné vytvořit ještě subkorpusy syntakticky anotované. Jediným takovým dostupným korpusem současné češtiny je SYN 2015, v němž se vyskytují jak české původní texty, tak překlady z různých jazyků. Při použití stejných kritérií jako u korpusu Jerome by však velikost překladového a nepřekladového subkorpusu byla značně rozdílná. Tento rozdíl je poněkud zmenšen vyřazením kratších próz ze subkorpusu překladů, kterých je v SYN 2015 více, nicméně výsledné subkorpusy ani tak velikostně srovnatelné nejsou. Subkorpus nepřekladů v rámci SYN 2015 má celkovou velikost 8 189 903 pozic, zatímco subkorpus překladů z angličtiny 15 836 313 pozic. Zároveň v těchto subkorpusech nejsou dodrženy zásady srovnatelnosti jako u korpusu Jerome, neboť SYN 2015 nebyl vytvořen

¹ Osobní sdělení, 13. 6. 2017.

pro studium překladového jazyka. Je proto nutné považovat výsledky získané z těchto subkorpusů vzhledem k výzkumu překladových univerzálií pouze za orientační; pro skutečně relevantní výsledky by bylo nutné vytvořit také skutečně srovnatelný subkorpus.

2.3 Testovaná hypotéza a statistické zpracování dat

Nulová hypotéza, kterou výzkum testuje, předpokládá stejnou frekvenci vybraných „jedinečných prvků“ v českých beletristických nepřekladech a v překladech, jejichž zdrojovým jazykem byla angličtina. Alternativní hypotéza potom předpokládá, že frekvence daných prvků v nepřekladech a překladech se liší. Pokud platí hypotéza jedinečných prvků, pak by frekvence těchto prvků v překladech měla být statisticky významně nižší než v nepřekladech.

Statistické zpracování dat vychází z doporučení uvedených v knize *Corpus Linguistics with BNCweb – a Practical Guide* (Hoffmann et al. 2008), která jsou přímo určená pro testování hypotéz o četnostech výskytu jazykových jevů v korpusech. Vzhledem k možnosti zjistit přesnou p-hodnotu bylo zvoleno testování v prostředí softwaru R² pomocí tzv. testu proporcí (*proportions test*).

P-hodnota zjištěná testem proporcí se porovnává s předem stanovenou hladinou významnosti (pokud ji nepřesáhne, nulovou hypotézu lze zamítnout). V souladu se všeobecně přijímanými mezními hodnotami, tedy $p < 0,05$ pro minimální úroveň signifikance výsledku, $p < 0,01$ pro již značně důvěryhodný výsledek a $p < 0,001$ pro téměř jistý výsledek, bylo testování provedeno na hladině významnosti 0,05. U výsledků je potom uvedena i případná nižší hladina významnosti, na které jsou signifikantní.

Zamítnutí nulové hypotézy na základě statistického testování nicméně ještě nevyovídá nic o příčinách pozorovaného jevu – pouze o vysoké pravděpodobnosti existence rozdílu v hodnotách četností daného prvku. Nulová hypotéza se tedy přijímá nebo zamítá pouze na základě pravděpodobnosti, nikoli jistoty (Vasishth & Nicenboim 2016:12).

3. Výsledky výzkumu

3.1 Lexikální jednotky

Soubor vybraných lexikálních jednotek, které nemají přímé protějšky v angličtině, tvoří příslovce *prý*, spojky *čili* a *totiž* a záporná částice *nikoli*. V korpusu Jerome byla vyhledána frekvence výskytů celé této skupiny lexikálních jednotek pomocí CQL³ příkazu: [lemma="prý|čili|totiž|nikoli"].

V subkorpusu nepřekladů bylo zjištěno 12 106 výskytů (720,21 i.p.m.⁴) těchto lemmat, v subkorpusu překladů 6 025 výskytů (347,79 i.p.m.), p-hodnota $< 0,001$ značí signifikantní rozdíl.

Následně byla lemmata vyhledána jednotlivě. Frekvence jsou následující: *Prý*: nepřeklady 5 116 výskytů (i.p.m. 304,36), překlady 1 923 výskytů (i.p.m. 111,01); *čili*: nepřekla-

² Volně dostupný software pro statistické výpočty a grafy. Stažen z [www: https://cran.r-project.org](https://cran.r-project.org)

³ CQL = *corpus query language*, dotazovací jazyk korpusu.

⁴ i.p.m. = *instances per million*, relativní četnost na milion pozic.

dy 334 výskytů (i.p.m. 20,47), překlady 175 výskytů (i.p.m. 10,1); *totiž*: nepřeklady 5 139 výskytů (i.p.m. 305,73), překlady 2 932 výskytů (i.p.m. 169,25); *nikoli*: nepřeklady 1 507 výskytů (i.p.m. 89,65), překlady 995 výskytů (i.p.m. 57,44). Ve všech případech zjištěná p-hodnota < 0,001 potvrzuje signifikantní rozdíl.

Tyto výsledky naznačují, že žádné ze slov se nevymyká zjištěné celkové distribuci, která je v překladech z angličtiny výrazně řidší než v původních českých textech: frekvence daných slov v překladech je zhruba poloviční oproti frekvenci v nepřekladech.

3.2 Předponová slovesa

Odvozování nových sloves pomocí prefixace je v češtině velmi rozšířený a produktivní způsob slovtvorby. V angličtině existuje také, ovšem v mnohem omezenější podobě a s omezenější škálou funkcí. Předponová slovesa tedy představují oblast bohatou na jedinečné prvky. Je ovšem nutné vybrat taková slovesa, která v angličtině nemají nejen slovtvorný, ale ani přímý významový jednoslovný ekvivalent. Proto byl zvolen postup od vytipování určitých funkčních skupin předpon a předponových sloves k vytrídění konkrétních lemmat vyhledávaných v korpusu.

Specifickým případem odvozování sloves předponami je vícenásobná prefixace. Existence produktivního vícenásobného předponového odvozování se uvádí i v angličtině, jedná se nicméně o případy jako *ex-vice-president*, *pseudo-anti-intellectual* a podobné (Lehrer 2009: 142), které se netýkají sloves. Z hlediska hypotézy jedinečných prvků jsou relevantní slovesa již jednou prefigovaná, u nichž se připojením prefixu *po-* vyjadřuje malá míra: *poodejít*, *poodhřnout*, *poohlédnout se*, *pootevřít*, *popohnat*, *popojít*, *popostrčit*, *povyříst* a další (Karlík et al. 1995: 209). Tato slovesa lze vytrdit z frekvenčního seznamu lemmat vyhledávaných dotazem: [tag="V.*"&lemma="poo.*|popo.*|povy.*"].

Vybereme-li z frekvenčního seznamu korpusu Jerome 35 nejčastějších lemmat odpovídajících požadovanému typu předponových sloves, dostaneme pak výčet pro dotaz: [lemma="popovídat|pootevřít|popojít|pootočít|povytáhnout|poodstoupit|poohlédnout|poodejít|popotáhnout|popojet|poopravit|popohnat|povyříst|popohánět|poposednout|popotahovat|popojíždět|poodhřnout|povyskočit|popoběhnout|poodhalit|popocházet|popostrčit|povyprávět|poohlížet|povykládat|poposedávat|povyrazit|popouštět|povytaovat|povylézt|popolézt|pootáčet|poodhalovat|povystřít"].

Frekvence těchto lemmat v nepřekladech je 2 019 (120,11 i.p.m.), v překladech pak 1 998 výskytů (115,3 i.p.m.). Mezi těmito frekvencemi nebyl zjištěn signifikantní rozdíl (p-hodnota = 0,2037)

Dalším jedinečným prvkem v oblasti předponových sloves by mohla být ta, která Uher (1987: 74) řadí do „třídy předponových sloves členitého děje“, konkrétně „skupina prefixativ přerušovaně moderativního děje“. Karlík et al. (1995: 209) tento typ uvádějí jako „malou míru děje přerušovaného“. Jde o slovesa jako *pobolívat*, *podřimovat*, *pokřikovat*, *pokukovat*, *polehávat* apod.

U těchto sloves je ovšem problematické vyhledávání pomocí CQL dotazů, protože existuje velké množství frekventovaných sloves začínajících *po-* a končících *-at*, která do dané kategorie nespádají (např. *podívat*, *potřebovat*, *používat*, *poznat...*). Proto byly k vytvoření seznamu lemmat využity především příklady z citovaných publikací. Výsledkem je dotaz: [lemma="postávat|posedávat|polehávat|pobolívat|podřimovat|po-

kašlávat|posmrkávat|pokřikovat|pořvávat|pokukovat|pošilhávat|poprchávat|poštuchovat|potloukat|potulovat|povalovat|pochrchlávat|pozpěvovat|pobroukávat|podupávat|poskakovat|pobíhat|poletovat|pokulhávat|podřimovat").

Celkový počet výskytů v nepřekladech je 2 682 (159,56 i.p.m.), v překladech 2 492 (143,85 i.p.m.). Mezi těmito frekvencemi se již prokázal signifikantní rozdíl (p-hodnota = 0,0002054).

Jako další relevantní předpona u sloves se jeví kombinovaný prefix *vyna-*, u nějž *Příruční mluvnice češtiny* (Karlík 1995: 210 a 213) uvádí, že vyjadřuje velkou míru děje a u většiny sloves zároveň také hodnotící význam „uspokojení z plnosti děje“. Mezi tato slovesa se řadí např. *vynadívat se*, *vynachválit se*, *vynapovídat se*, *vynasmát se...*; pro vyhledání dalších možných lemmat tohoto typu lze použít dotaz na všechna slovesa začínající řetězcem „vyna“: [tag="V.*"&lemma="vyna.*"]. Vytríděním nalezených lemmat a jejich doplněním k příkladům z příručky vznikne následující dotaz: [lemma="vynadívat|vynadivit|vynachválit|vynakoukat|vynapovídat|vynasmát|vynasnažit|vynadávat|vynařikat"].

Zjištěné frekvence těchto sloves jsou v nepřekladech 94 výskytů (5,59 i.p.m.) a v překladech 115 výskytů (6,64 i.p.m.). V překladech je tedy těchto sloves více, rozdíl ve frekvencích ovšem není statisticky signifikantní (p-hodnota = 0,2437). Jev je navíc značně řídký.

Poslední zkoumanou skupinou předponových sloves jsou lemmata *napomoci*, *napomáhat*, *vypomoci* a *vypomáhat*, u nichž prefixace posouvá význam základového slovesa jen nepatrně. Vyhledají se jednoduchým výčtem: [lemma="napomoci|napomáhat|vypomoci|vypomáhat"].

Zjištěné frekvence jsou 156 (9,28 i.p.m.) v nepřekladech a 127 (7,33 i.p.m.) v překladech. Tento rozdíl není statisticky signifikantní, ačkoli jde téměř o hraniční výsledek (p-hodnota = 0,05509). I tato skupina sloves se vyskytuje poměrně řídko.

Sloučí-li se všechny uvedené skupiny předponových sloves do jedné, frekvence je pak 4 951 výskytů v nepřekladech a 4 732 v překladech. Tento výsledný rozdíl by byl statisticky signifikantní (p-hodnota = 0,0002161), nicméně pouze díky druhé skupině sloves; součet jednotlivých skupin v tomto případě nelze považovat za relevantní výsledek. Jedinými předponovými slovesy, která vykazují frekvence odpovídající hypotéze jedinečných prvků, jsou tedy „prefixativa přerušovaně moderativního děje“.

3.3 Předpona *pra(ch)*- u adjektiv a adverbí

Podobně jako u sloves jsou možnosti odvozování předponami omezené i u jiných slovních druhů. Angličtina tak nedisponuje například protějšky českých intenzifikujících předpon *pra-* (*prach-*) a *vele-*.

Adjektiva s předponami *pra-* a *prach-* lze v korpusu vyhledat pomocí dotazu: [tag="A.*"&lemma="pra.*"]. Výstupem je ovšem i řada lemmat, která nespádají do požadované skupiny adjektiv, u nichž předpona *pra-* plní intenzifikační funkci a zároveň nemají přímý anglický protějšek (např. u lemmatu *pravěký* jde o tutéž předponu, nicméně slovo má přímý protějšek *prehistoric*), případně u nich řetězec *pra* ani není předponou (*pravý*, *pracovní*, *pražský*, *pramenící* apod.).

Z frekvenčního seznamu lemmat získaných takto z obou subkorpusů je proto nutné ručně vytrdit ta, která kritériím odpovídají, a zadat nový dotaz v podobě výčtu:

[lemma="prastarý|prapodivný|pradávný|prazvláštní|prapůvodní|pramalý|prachobyčejný|pranepatrný|prachsprostý|prazákladní|prachmizerný|prachpitomý|praobyčejný|prazárodečný|prašpatný|prachbídný|prabídný|prachšpatný|prajednoduchý|pramizerný|praprádavný"].

Subkorpus nepřekladů obsahuje 18 lemmat s celkovým počtem 606 výskytů (36,23 i.p.m.), subkorpus překladů 19 lemmat s celkovým počtem 719 výskytů (41,5 i.p.m.). V tomto případě tedy předpokládaný jedinečný prvek vykazuje neočekávanou distribuci a převažuje v překladech. Rozdíl ve frekvenci je dokonce statisticky významný, p-hodnota = 0,01145. Vzhledem k tomu, že jev není extrémně řídký a korpus Jerome by neměl být ovlivněn idiolektem překladatelů, je nutné hledat vysvětlení spíše v paralelním korpusu, kde lze zjistit, z jakých anglických formulací jsou tato adjektiva překládána.

Výskyt jednotlivých lemmat v obou subkorpusech vykazuje podobné rysy. Sedm nejfrekventovanějších lemmat a zároveň jediných sedm s více než 10 výskyty se shoduje; jsou to lemmata *prastarý*, *prapodivný*, *pradávný*, *prazvláštní*, *pramalý*, *prapůvodní* a *prachobyčejný*.

Analogický postup lze použít i u adverbii s předponou *pra-* (*prach-*). V tomto případě se nejprve použije dotaz, který vyhledá všechna adverbia začínající daným řetězcem znaků: [tag="D.*"&lemma="pra.*"]. Následným ručním vyfiltrováním frekvenčního seznamu výsledků z obou subkorpusů pak získáme lemmata pro další dotaz: [tag="D.*"&lemma="pramálo|prapodivně|pradávně|prachmizerně|prachbídně|prapůvodně|prachsprostě|prašpatně|prastaře|pranepatrně|prabídně|prazvláštně|prachobyčejně"].

Výsledkem je v nepřekladech 11 lemmat s celkem 116 výskyty (6,9 i.p.m.) a v překladech 8 lemmat s celkem 92 výskyty (5,31 i.p.m.). V tomto případě je jev řídký než u adjektiv, a přestože frekvence u nepřekladů je mírně vyšší, není rozdíl signifikantní (p-hodnota = 0,06993). Opět se shodují dvě nejčastější lemmata s více než deseti výskyty: *pramálo* a *prapodivně*.

Vzhledem k tomu, že oba jevy (adjektiva a adverbia se zesilovací předponou *pra-*) jsou poměrně řídké a zároveň se jedná o tutéž předponu s tímtež významem, lze je pro účely statistického testování sloučit. Dostaneme tak 722, resp. 811 výskyty a výslednou p-hodnotu = 0,09738. Z ní plyne, že rozdíl v celkových frekvencích není statisticky signifikantní.

3.4 Předpona *vele-* u substantiv a adjektiv

Obdobný případ je předpona *vele-*, která má rovněž zesilovací funkci. Adjektiva s předponou *vele-* vyhledá příkaz: [tag="A.*"&lemma="vele.*"]. Opět je nutné vyřadit lemmata, která neodpovídají požadovaným podmínkám (např. *velebný*, *velehradský*, *veletržní*) a následně zadat nový dotaz pomocí výčtu: [lemma="veledůležitý|velevážený|velehorský|velectěný|velejemný|veleúspěšný|velevzácný|veleslavný|velevýznamný|veleznámý|velemocný|velejasný|veleoblíbený|velezajímavý|veledůstojný|veleučený|velevítaný|velemoudrý|velemužný|velecenný"].

Ten vyhledá v nepřekladech 17 lemmat s 81 výskyty (4,82 i.p.m.) a v překladech 11 lemmat s 42 výskyty (2,42 i.p.m.). Rozdíl frekvencí je zřetelný, protože se ovšem jedná o značně řídký jev, bylo statistické testování provedeno až v součtu s obdobně tvořenými substantivy. Ta vyhledá dotaz: [tag="N.*"&lemma="vele.*"]. Po vyřazení lemmat s přímými

anglickými protějšky (např. *velezrada*, *veletucet*) získáme výčet: [lemma="velehory|veletok|velehora|veledílo|velerada|veleříše|velechrám|veleduch|veleproud|velepiseň|vele-moudrost|veleměsto|velevýroba|veledílko|veledůležitost|velekoráb|veledůstojnost|veleostuda|veleváženost"].

Tento dotaz vyhledá v nepřekladech celkem 15 lemmat a 120 výskytů (7,14 i.p.m.), v překladech 10 lemmat a 25 výskytů (1,44 i.p.m.). Opět jde o velmi řídký jev. Po sečtení výskytů obou skupin zjistíme, že rozdíl je statisticky signifikantní (p-hodnota < 0,001), nicméně tomuto výsledku nelze přikládat příliš velkou váhu, neboť jev je nejen řídký, ale také značně závislý na konkrétním obsahu jednotlivých textů zařazených do korpusu.

3.5 Zesilující sufixy u adjektiv

Podobně jako u výše uvedených zesilujících prefixů angličtina oproti češtině postrádá i zesilující sufixy. Karlík et al. (1995: 177–178) pro adjektiva uvádí následující: *-ičký* (*mladický*, *celičký*), *-oučký* (*mladoučký*, *růžovoučký*), dále *-inký* (*mladinký*, *prostinký*) a *-ounký* (*mladounký*, *hezounký*) a jejich rozšířené podoby (*-ičičký*, *-oulinký*, *-inkatý* a další). Takto prefigovaná adjektiva vyhledá dotaz: [tag="A.*"&lemma="*inký|*ounký|*ičký|*ouč-
ký|*inkatý|*ičkatý"].

V nepřekladech je 2 448 výskytů (145,64 i.p.m.), v překladech 2 266 výskytů (130,81 i.p.m.), rozdíl je statisticky signifikantní (p-hodnota < 0,001).

3.6 Zájmena *tentam*, *tatam*, *totam*

Specificky utvořené samostatně stojící zájmeno *tentam* nemá přímý anglický protějšek. Pravopisně je možná i varianta *ten tam*, tu nicméně při vyhledávání v korpusu není možné odlišit od pouhého souvýskytu slov *ten* a *tam*, který se jeví jako častější, a proto byla použita pouze varianta psaná dohromady.

V tomto případě postačuje vyhledat jednoduše lemma *tentam* a získáme následující výsledky: 68 výskytů (4,05 i.p.m.) v nepřekladech a 77 výskytů (4,44 i.p.m.) v překladech. Kromě toho, že je počet výskytů v překladech v rozporu s očekáváním vyšší, je na první pohled patrná velmi nízká frekvence těchto zájmen. Rozdíl četností nicméně není statisticky signifikantní (p-hodnota = 0,6291).

Vzhledem k řídkosti a povaze jevu je možné předpokládat i vliv idiolektu jednotlivých autorů či překladatelů. Z frekvenčního seznamu podle dokumentů lze zjistit, že celá polovina výskytů v nepřekladech pochází z pouhých sedmi děl. U překladů je situace obdobná, z prvních sedmi děl pochází dokonce o něco více než polovina výskytů. Subkorpusy přitom obsahují po několika stovkách děl.

3.7 Zájmena *kdopak*, *copak*, *kterýpak*, *jakýpak*, *čipak*

Dalším typem českých zájmen, která v kontrastu s angličtinou představují jedinečný prvek, jsou zájmena *kdopak*, *copak*, *kterýpak*, *jakýpak* a *čipak*. *Příruční mluvnice češtiny* (Karlík et al. 1995:217) k nim uvádí, že „jinak prázdný sufix *-pak* má hovorovou povahu“. Tato zájmena lze v korpusu Jerome vyhledat následujícím dotazem: [lemma="copak|kdo-
pak|kterýpak|jakýpak|čipak"].

Výsledkem je 2 896 výskytů (172,29 i.p.m.) v nepřekladech a 2 493 výskytů (143,91 i.p.m.) v překladech. Rozdíl četností je statisticky signifikantní (p-hodnota < 0,001).

3.8 Číslovky *nejeden, nejedna, nejedno*, adverbium *nejednou*

Způsob tvoření číslovek *nejeden, nejedna, nejedno* i adverbia *nejednou* v angličtině sice existuje v podobě výrazů *not one, not once*, tyto výrazy ovšem mají zcela jiný význam: zatímco *nejednou* znamená „více než jednou“, *not once* znamená „ani jednou“, stejně tak je tomu u uvedených číslovek. Uvedená lemmata lze vyhledat jednoduchým příkazem: [lemma="nejeden|nejednou"].

V nepřekladech je 252 výskytů (14,99 i.p.m.), v překladech 191 výskytů (11,03 i.p.m.). Rozdíl mezi subkorpusy je statisticky signifikantní (p-hodnota = 0,001533).

3.9 Číslovky *kolikátý, tolikátý, několikátý*

Řadové číslovky *kolikátý, tolikátý* a *několikátý* nemají v angličtině jednoslovné protějšky; lze tedy očekávat, že budou fungovat jako jedinečné prvky. Uvedené číslovky vyhledá dotaz: [lemma="kolikátý|tolikátý|několikátý"].

V nepřekladech bylo zjištěno 135 výskytů (8,03 i.p.m.), v překladech 44 výskytů (2,54 i.p.m.). Tento výsledek je statisticky signifikantní (p-hodnota < 0,001), jev je nicméně velmi řídký, a proto mu nelze připisovat příliš velkou váhu.

3.10 Číslovky *dvojí, trojí, čtvero, patero...*

Tento typ číslovek lze sice v korpusu vyhledávat i pomocí tagů, takové vyhledávání nicméně není tak spolehlivé jako výčet jednotlivých lemmat: [lemma="dvoj.|troj.|čtver.|pater.|šester.|sedmer.|osmer.|devater.|desaterý|jedenácter.|dvanácter.|třinácter.|čtrnácter.|patnácter.|šestnácter.|sedmnácter.|osmnácter.|devatenácter.|dvacater.|ster.|tisícer.|několiker.|toliker.|koliker."].⁵

Výsledkem je celkem 852 výskytů (50,69 i.p.m.) v nepřekladech a 384 výskytů (22,17 i.p.m.) v překladech. Rozdíl mezi četnostmi je statisticky signifikantní (p-hodnota < 0,001). Pohled do frekvenčního seznamu lemmat navíc ukazuje, že proporce výskytu jednotlivých číslovek jsou v obou subkorpusech různé: v překladech se jen velmi zřídka vyskytují jiná lemmata než prvních pět nejčastějších, kterými jsou *dvojí, trojí, tisícerý, několikérý* a *několikery*. V nepřekladech je postupný úbytek četností výrazně pozvolnější.

3.11 Věta začínající předmětem

Coby zkoumané prvky byly využity předměty substantivní, bezpředložkové, které obsazují první pozici ve větě. Komplikací pro využití tohoto jevu jakožto jedinečného prvku představuje skutečnost, že korpus Jerome není syntakticky značkováný. Proto byl využit korpus SYN 2015, který obsahuje v beletristické části jak překlady z angličtiny, tak nepřekladové texty. Je ovšem třeba přitom mít na paměti, že tento korpus nebude

⁵ Lemma *desatero* bylo záměrně vynecháno kvůli dvojznačnosti.

srovnatelný (subkorporusy překladů a nepřekladů nejsou stejně rozsáhlé, není zohledněno množství textu od stejného autora či překladatele atd.) a výsledky z něj získané je proto nutné považovat spíše za orientační.

Pro vyhledání bezpředložkových substantivních předmětů na začátku věty v subkorpusech SYN 2015 byl zadán CQL příkaz: <s>[afun="Obj"&>tag="N.*"].

Nepřekladové texty obsahovaly 8 110 výskytů (i.p.m. 990,24), překlady 17 540 výskytů (1107,58 i.p.m.). Tyto výsledky vypovídají o statisticky signifikantním rozdílu (p-hodnota < 0,001), a to s vyšším výskytem jevu v překladech, nikoli v původních textech.

Nicméně již zběžný pohled do konkordance odhalí, že chybovost morfologického i syntaktického značkování je velmi vysoká. V konkordanci se vyskytují substantiva v nominativu (*Osud to zařídil tak, že...*) a vokativu (*Drahoušku, jak to chceš dělat, ...*), zájmena (*Tomu říkám, pane, jídelniček.*) i adjektiva (*Ustrašení měšťané Vsetína nezamhouřili v noci oka.*)

Manuální kontrola prvních 100 položek v obou konkordancích ukázala chybovost 36, resp. 39 %. Vyhledání pozic značených jako předmět v dativu nebo instrumentálu pomocí příkazu: <s>[afun="Obj"&tag="N...[37].*"] ukazuje v nepřekladech chybovost značkování 17 %, zatímco v překladech 33 %. Míra chybovosti je tedy zjevně ovlivněna konkrétní podobou dotazu a značně zkrsluje výsledky vyhledávání.

Tento rozdíl zřejmě přinejmenším zčásti způsobuje to, že některá původně cizojazyčná vlastní jména (*Rose, Luce, Heloise* apod.), která se vyskytují především v překladech, jsou v češtině nesklonná a při jejich automatickém morfologickém značkování často dochází k chybnému vyhodnocení pádu.

Z výše uvedených důvodů byla věta začínající předmětem coby jedinečný prvek vyřazena ze statistického testování, neboť nebylo možné spolehlivě zjistit frekvenci jevu. Pro jakýkoli další výzkum tak bude nutné použít jinou metodu vyhledávání, případně jiný či jinak značkový materiál.

3.12 Věta jednočlenná

Jednočlenná věta slovesná představuje syntaktickou strukturu, v níž nefiguruje žádný podmět. Angličtina bezpodmětové věty na rozdíl od češtiny netvoří (Dušková 1988: 368). Pro účely využití jednočlenných slovesných vět jakožto jedinečného prvku je nutné s ohledem na možnosti korpusu Jerome stanovit způsob jejich vyhledávání. Jako nejspokladnější se přitom jeví vytipovat podle popisu a příkladů uváděných v jazykových příručkách seznam konkrétních lemmat, která se vyskytují pouze ve větách jednočlenných a nemohou tedy mít přímou oporu v originálu při překladu z angličtiny.

Příruční mluvnice češtiny (Karlík et al. 1995: 389–394) poskytuje poměrně rozsáhlý přehled typů predikátorů, které se ve větách jednočlenných slovesných mohou vyskytovat. Většinou se jedná o predikátory označující procesy nebo změny. Prvním typem jsou změny či procesy atmosférické (Karlík et al. 1995: 389): *Venku prší (leje, mžít, sněží...)*; *Hřmí; Už se šerí (rozednívá)*; *Svítlá; Přituhlo nám trochu* apod. Změny vyjadřují rovněž bezpodmětové věty typu *Udělal se chladno.*

Dalším typem jsou procesy a změny v tělesné a duševní sféře člověka (Karlík et al. 1995: 389): *Otci se stýská; Petrovi přeskočilo; Nemocnému se ulevilo (přítížilo)*; *Už mu*

vytrávil; Už mu otrnulo. Dále pak věty typu *Zachtělo se mu ředitelského platu; Králi se zželelo nebohé dívky.*

Další typy predikátorů představují slovesa, která se běžně vyskytují spíše v podmětových strukturách, a proto nejsou v tomto výzkumu využitelná. Jedná se o typ, který je v Příruční mluvnici češtiny označován jako „procesy a změny probíhající uvnitř neživých objektů“ (Karlík et al. 1995: 389): *V hodinách hrklo; V peci vyhaslo; Z hlediště zašumělo;* dále „procesy a změny probíhající v částech těla“ (Karlík et al. 1995: 390): *Hlady mu kručelo v břiše; Lechtá mě v uchu;* „procesy a změny s konkrétním původcem“, tedy věty typu: *Matce vyhaslo v peci; Teče mi z nosu;* dále „spontánní nastání děje nebo stavu“ se slovesy *dojít* (k něčemu), *schylovat se* (k něčemu): *Nakonec došlo k tomu, že se pohádali; Schyluje se k bouřce* (Karlík et al. 1995: 392–393) a větné struktury vyjadřující „předmět zájmu“ se slovesy *jít, běžet a jednat se* (o něco) a *záležet a nesejít* (na něčem): *Jde o všechno; Nezáleží (nesejde) mi na ničem* (Karlík et al. 1995: 393).

Z výsledného seznamu lemmat byla po předběžném vyhledání v korpusu vyřazena lemmata *hřmít* a *zahřmít*, u nichž se ukázal častý výskyt s podmětem (*Syčela tu tráva, hřměl vítr; Když zahřměly motory na plné obrátky...*). Naopak lemma *pršet* bylo do výzkumu zahrnuto přesto, že se může vyskytovat v konstrukcích typu *Prší vám štěstí, Hvězdy jen pršely*, protože bez ohledu na to, zda je namísto slova štěstí a *hvězdy* v těchto případech považovat spíše za podmět nebo za předmět, jsou tyto konstrukce vzhledem k častému výskytu lemmatu *pršet* jen velmi řídké. Lemmata byla vyhledána CQL příkazem: [lemma="pršet|mrholit|sněžit|mžít|šerit|sešeřilo|stmívat|svítat|rozednívat|přituhnout|přituhovat|vytrávit|otrnout|zachtít|zželet"].⁶

Subkorpus nepřekladů obsahoval 2 003 výskytů těchto lemmat (119,16 i.p.m.) a subkorpus překladů 1 273 výskytů (73,48 i.p.m.). Mezi frekvencemi je signifikantní rozdíl, p-hodnota < 0,001.

Zajímavá je i distribuce jednotlivých lemmat v subkorpusech. Z frekvenčního seznamu je patrné, že lemmata *zachtít* a *sněžit se* vymykají obecnému nižšímu výskytu jednočlenných vět slovesných v překladových textech. Nejedná se přitom o natolik řídká slova, aby byl pravděpodobný vliv faktorů, jako je idiolekt překladatele nebo obsah konkrétních textů. O příčinách tohoto jevu lze ovšem bez dalšího výzkumu pouze spekulovat.

3.13 Oslovování a označování osob

Posledním typem jedinečných prvků češtiny v kontrastu s angličtinou zkoumaných v této práci jsou uzuální jevy, konkrétně ustálené kolokace v oslovování či označování osob s určitým titulem, postavením nebo profesí.

V češtině je zvykem používat kombinaci slov „pan/paní/slečna“ a titulu/funkce, tedy např. „pan učitel“, „paní doktorka“, zatímco v angličtině je možná buď kombinace „Mr/Mrs“ a příjmení či (u učitelů, duchovních) „Professor“, „Reverend“ a příjmení, nebo oslovení pouhým „Sir/Madam“.⁷ Kombinace „Mr/Mrs“ a titulu/funkce se, podobně jako

⁶ Z neznámého důvodu se v korpusu Jerome nevyskytuje lemma *sešeřit*; u očekávaného tvaru „sešeřilo se“ bylo zjištěno lemma *sešeřilo*.

⁷ English Grammar Today © Cambridge University Press, dostupné z [www: http://dictionary.cambridge.org/grammar/british-grammar/people-and-places/names-and-titles-addressing-people](http://dictionary.cambridge.org/grammar/british-grammar/people-and-places/names-and-titles-addressing-people) [online 26. 6. 2017].

v češtině, používá pouze v případě funkcí typu *President, Secretary, Chairman* (prezident, ministr, předseda).

Pro účely této práce byla využita slova *učitel, magistr, doktor, docent, profesor, inženýr, farář* a jejich přechýlené protějšky ve spojení se slovy *pan, resp. paní/slečna*. Kombinace těchto slov ve všech tvarech byly vyhledány v korpusu Jerome pomocí CQL příkazu: [lemma="pan|paní|slečna"][lemma="učitel|učitelka|magistr|magistra|doktor|doktorka|docent|docentka|profesor|profesorka|inženýr|inženýrka|farář|farářka"]

V nepřekladech je 2 342 výskytů (139,33 i.p.m.) těchto kombinací, v překladech 870 (50,22 i.p.m.), p-hodnota < 0,001 značí signifikantní rozdíl.

4. Závěry

Tabulka 1: Shrnutí výsledků kvantitativního výzkumu

jedinečný prvek	nepřeklady	překlady	p-hodnota	sig. rozdíl
prý, čili, totiž, nikoli	12 106	6 025	< 2.2e-16	ano
pra-	722	811	0.09738	ne
vele-	201	67	< 2.2e-16	ano
-ický, -ounký	2 448	2 266	0.000241	ano
tentam	68	77	0.6291	ne
kdopak, čípak	2 896	2 493	4.587e-11	ano
kolikátý	135	44	4.227e-12	ano
dvoje, troje	657	316	< 2.2e-16	ano
nejeden, nejednou	252	191	0.001533	ano
předponová slovesa	nejednotné	výsledky	-	spíše ne
věta jednočlenná	2 003	1 273	< 2.2e-16	ano
pan učitel	2 342	870	< 2.2e-16	ano

Jak je patrné z tabulky, u většiny kandidátních jedinečných prvků se prokázal statisticky významný rozdíl mezi četností v překladech a nepřekladech, a to pokaždé s nižším výskytem v překladech – tak, jak předpokládá *unique items hypothesis*.

Poněkud překvapivě se ovšem jeví skutečnost, že se tento trend nepotvrdil u natolik specifického a zároveň natolik frekventovaného prvku, jakým jsou předponová slovesa. K přinejmenším částečnému odhalení příčin by mohla přispět analýza zdrojových výrazů provedená na paralelním korpusu.

Výsledky kvantitativní analýzy sice ukazují na povětšinou statisticky signifikantní rozdíl, nevypovídají nicméně o velikosti tohoto rozdílu (k tomu by bylo nutné kromě p-hodnoty vyhodnocovat také konfidenční intervaly). Mimo to na základě těchto výsledků nemůžeme usuzovat, že *unique items hypothesis* zkrátka platí. K tomu je zapotřebí dalších rozsáhlých výzkumů korpusových dat a podrobnějších analýz možných příčin zjištěných frekvencí jedinečných prvků, včetně kognitivní analýzy procesu překlada.

Hlavním výstupem tohoto výzkumu tak zůstává skutečnost, že existují rozdíly v četnosti určitých jazykových prvků mezi překladovou a nepřekladovou češtinou, které je zapotřebí dále zkoumat, ověřovat a pokud možno také vysvětlovat. Samotné zjištění, že četnost konkrétních jedinečných prvků v překladech neodpovídá nepřekladům, může také v praxi pomoci překladatelům k uvědomělejší práci s češtinou jako cílovým jazykem a lepšímu využití její výrazové škály.

BIBLIOGRAFIE

- Catford, John (1975) *A Linguistic Theory of Translation*, London: Oxford University Press.
- Duškova, Libuše (1988) *Mluvnice současné angličtiny na pozadí češtiny*, Praha: Academia.
- Hoffmann, Sebastian, Evert, Stephan, Smith, Nicholas, Lee, David, Berglund Prytz, Ylva (2008) *Corpus Linguistics with BNCweb – a Practical Guide*, Frankfurt am Main: Peter Lang.
- Holmes, James (1972) 'The Name and Nature of Translation Studies', in James Holmes (ed.) *Translated! Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi, 67–80.
- Chesterman, Andrew (2007) 'What is a unique item?', in Yves Gambier, Miriam Shlesinger, Radegundis Stolze (eds.) *Doubts and Directions in Translation Studies*, Amsterdam-Philadelphia: John Benjamins, 3–13.
- Chlumská, Lucie (2013) *JEROME: jednojazyčný srovnatelný korpus pro výzkum překladové češtiny*, Praha: Ústav Českého národního korpusu FF UK: <http://www.korpus.cz> (last access: 21. 3. 2018).
- Karlík, Petr, Nekula, Marek, Rusinová, Zdeňka (eds.) (1995) *Příruční mluvnice češtiny*, Praha: Nakladatelství Lidové noviny.
- Kolehmainen, Leena (2013) 'Die Unikat-Hypothese der Translation', *trans-kom* 6(1), 92–114.
- Křen, Michal, Chlumská, Lucie, Jelínek, Tomáš, Kovářiková, Dominika, Petkevič, Vladimír, Procházka, Pavel, Skoumalová, Hana, Škrabal, Michal, Truneček, Petr, Vondříčka, Pavel, Zasina, Adrian (2015) *SYN2015: reprezentativní korpus psané češtiny*, Praha: Ústav Českého národního korpusu FF UK: <http://www.korpus.cz> (last access: 21. 3. 2018).
- Lehrer, Adrienne (2009) 'Prefixes in English Word Formation', *Folia linguistica* 29(1–2), 133–148.
- Levý, Jiří (1963) *Umění překladu*, Praha: Československý spisovatel.
- Reiss, Katharina (1971) *Möglichkeiten und Grenzen der Übersetzungskritik*, München: Max Hueber.
- Tirkkonen-Condit, Sonja (2002) 'Translationese – a myth or an empirical fact?' *Target* 14(2), 207–220.
- Tirkkonen-Condit, Sonja (2004) 'Unique items – over- or under- represented in translated language?', in Anna Mauranen, Pekka Kujamäki (eds.) *Translation Universals: Do They Exist?*, Amsterdam-Philadelphia: John Benjamins.
- Uher, František (1987) *Slovesné předpony*, Brno: UJEP.
- Vasishth, Shravan & Nicenboim, Bruno (2016) 'Statistical Methods for Linguistic Research: Foundational Ideas – Part I', *Language and Linguistics Compass* 10(8), 349–369.

RESÜMEE

Dieser Artikel hat zum Ziel, die Unikat-Hypothese (*unique items hypothesis*) an tschechischen Sprachdaten zu testen. Diese von Sonja Tirkkonen-Condit formulierte Hypothese setzt die Unterrepräsentation unikaler Elementen (solche Mittel der Zielsprache, die keine direkten Entsprechungen in der Ausgangssprache haben) in übersetzten Texten im Vergleich zu Nichtübersetzungen derselben Sprache voraus. Als Materialgrundlage dient ein einsprachiges, vergleichbares Referenzkorpus tschechischer Übersetzungen und Nichtübersetzungen. Beide Subkorpora haben den Umfang von ungefähr 17 Mio.

Token. Die getesteten tschechischen unikaln Elemente wurden aus Lexemen, Wortbildungsphänomenen, syntaktischen Strukturen und Sprachgebrauchsphänomenen ausgewählt. Die Frequenz dieser Elemente wurde statistischen Tests im Programm R unterzogen. Die Ergebnisse zeigen eine allgemeine Tendenz der übersetzten tschechischen Sprache weniger unikaln Elemente zu enthalten. Einige der einzelnen Elemente entsprechen jedoch dieser Tendenz nicht.

Mgr. Adéla Špínová
Filozofická fakulta Univerzity Karlovy, Praha
adela.spinova@ff.cuni.cz