



UNIVERSITÄT
LEIPZIG

Philologische Fakultät
Institut für Slavistik



UNIVERZITA
KARLOVA

Filozofická fakulta
Ústav translatologie

BAKALÁŘSKÁ PRÁCE

Mezikulturní komunikace: čeština a němčina pro překlad a tlumočení /
Interkulturelle Kommunikation: Übersetzen und Dolmetschen Tschechisch-Deutsch

Vlasta Řenčová

Komentovaný překlad: Terminologie in der neuronalen maschinellen Übersetzung. Tom Winter, Daniel Zielinski. In: Maschinelle Übersetzung für Übersetzungsprofis, ed. Jörg Porsiel, 2020, BDÜ, s. 210-233.

Annotated translation: Terminologie in der neuronalen maschinellen Übersetzung. Tom Winter, Daniel Zielinski. In: Maschinelle Übersetzung für Übersetzungsprofis, ed. Jörg Porsiel, 2020, BDÜ, p. 210-233.

Vedoucí práce: Prof. Dr. Danuta Rytel-Schwarz, Mgr. Věra Kloudová, PhD.

Leipzig / Praha 2022

Zadání

Zadaný text přeložte do češtiny a svůj překlad doplňte překladatelským komentářem v rozsahu min. 20 normostran. V komentáři nejprve celkově charakterizujte výchozí text: uveďte, s jakým cílem byl text napsán a jaké stylistické postupy autor/ka volí k dosažení svého záměru. Dále popište, na jaké problémy jste v překladu narazila, a zdůvodněte použité překladatelské postupy a nezbytné posuny, které jste v překladu provedla na úrovni lexika, syntaxe a především v rovině stylistické. Postupujte přitom od celkové koncepce svého překladu k dílčím řešením. Komentář opatřete bibliografickým soupisem použitých primárních i sekundárních zdrojů, včetně internetových. Kromě dodržení formálních náležitostí stanovených Pravidly pro organizaci studia FF UK (Čl. 19) připojte ke každému vázanému exempláři práce vždy dvě kopie výchozího textu: jednu napevno svázanou s ostatními listy a druhou volně vloženou.

Poděkování

Ráda bych poděkovala Mgr. Věře Kloudové, PhD., za pomoc s výběrem vhodného textu, vedení práce, její cenné rady a připomínky k originálnímu textu i mému překladu a za trpělivost a čas věnovaný konzultacím. Děkuji také Prof. Dr. Danutě Rytel-Schwarz za podporu a vedení teoretické části mé práce. Můj dík dále patří PhDr. Bc. Tomáši Svobodovi, Ph.D., za pomoc s terminologií, Mgr. Lukáši Klimešovi za konzultaci vybraných pravopisných problémů a mým nejbližším za morální podporu a kritické připomínky. V neposlední řadě chci poděkovat všem vyučujícím a spolužákům, kteří mi v průběhu studia předávali své poznatky, pomáhali mi, motivovali mě a inspirovali.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že má práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 9.5.2022

Řeňčová
.....

Vlasta Řeňčová

Abstrakt

Tato bakalářská práce se skládá ze dvou částí. První, praktická část, je překlad článku *Terminologie in der neuronalen maschinellen Übersetzung* autorů Toma Wintera a Daniela Zielinského, který vyšel jako součást publikace *Maschinelle Übersetzung für Übersetzungsprofis*. V článku je tematizována problematika terminologie ve strojových překladech, jednotlivé druhy chyb vyskytující se ve strojových překladech a trénink neuronových strojových systémů, jehož cílem je zlepšení kvality strojových překladů. Druhou částí práce je komentář, který zahrnuje překladatelskou analýzu, metodu překladu a typologii překladatelských problémů a posunů. Cílem této práce není pouze překlad výchozího článku, ale také okomentování zvolených překladatelských řešení.

Klíčová slova: překlad, překladatelská analýza, překladatelský problém, překladatelský posun, terminologie, strojový překlad, neuronový strojový překlad, trénink překladačů

Abstract

This bachelor thesis consists of two parts. The first part is a Czech translation of the German article *Terminologie in der neuronalen maschinellen Übersetzung* by Tom Winter and Daniel Zielinski, which was published as part of the publication *Maschinelle Übersetzung für Übersetzungsprofis*. The article addresses the issue of terminology in machine translations, various types of errors occurring in machine translations and the training of neural machine systems, the aim of which is to improve the quality of NMT. The second part of the thesis is a commentary on the translation, which includes translation analysis, translation method and typology of translation problems and shifts. The aim of this thesis is not only the translation of the initial article, but also the commentary of the selected translation solutions.

Key words: translation, translation analysis, translation problem, translation shift, terminology, machine translation, neural machine translation, training of NMT-engines

Obsah

1. Úvod.....	7
2. Překlad.....	8
3. Komentář.....	32
3.1 Překladatelská analýza originálu.....	32
3.1.1 Vnětextové faktory.....	32
3.1.2 Vnitrotextové faktory.....	36
3.2 Překladatelská metoda a postupy.....	40
3.3 Typologie a řešení překladatelských problémů.....	41
3.3.1 Lexikální rovina.....	41
3.3.2 Morfologická a syntaktická rovina.....	47
3.3.3 Stylistická rovina.....	50
3.3.4 Pragmatická rovina.....	51
3.4 Typologie posunů.....	52
3.4.1 Intelktualizace.....	52
3.4.2 Vynechávky.....	52
3.4.3 Substituce.....	53
4. Závěr.....	54
Použitá literatura.....	55
Přílohy.....	57
Selbständigkeitserklärung.....	58

1. Úvod

Tato bakalářská práce se skládá ze dvou základních částí – z překladu článku *Terminologie in der neuronalen maschinellen Übersetzung* od autorů Toma Wintera a Daniela Zielinského a z komentáře k tomuto překladu. Komentář obsahuje analýzu textu podle modelu translatoložky Christiane Nordové (2009), popis metody překladu, typologii překladatelských problémů a strategii zvolenou u konkrétních řešení.

Překládaný článek se věnuje problematice tréninku systémů neuronového strojového překladu, konkrétně tomu, jak tyto systémy překládají terminologii a jak je možné tento proces optimalizovat, aby se snížily náklady na posteditaci. Článek vyšel ve sborníku odborných článků *Maschinelle Übersetzung für Übersetzungsprofis*, která byla vydaná v roce 2020 a do češtiny dosud nebyla přeložena. Při výběru článku jsem vycházela z toho, že oblast strojových překladů je relevantní také pro zainteresované publikum v České republice (studenty překladatelství, překladatele). Motivací pro výběr pro mě byl i osobní zájem o dané téma.

V komentáři k překladu se na originální text odkazují pomocí písmene „O“ a čísla stránky tak, jak je uvedeno v publikaci, na překlad se odkazují pomocí písmene „P“ a čísla řádku.

2. Překlad

11. Terminologie v neuronovém strojovém překladu

Tom Winter, Daniel Zielinski

11.1 Úvod

- 5 Globální technologické trendy jako strojové učení a umělá inteligence v poslední době silně ovlivňují vývoj produktů a služeb. Předpokladem pro to byla dostupnost výkonných grafických procesorů (*Graphics Processing Unit* – GPU), nových algoritmů na poli strojového učení a dostatečně velké množství kvalitních dat.
- 10 Od počátku internacionalizace a globalizace, tedy od konce osmdesátých a začátku devadesátých let, nashromáždilo mnoho mezinárodních společností velké množství jedno- nebo vícejazyčných dat ve formě textů a audiozáznamů; jedná se především o velké internetové koncerny, jako jsou Amazon, Baidu, Facebook, Google, Microsoft, Yandex a další. Důsledkem toho se právě jazyky záhy staly hlavní používanou platformou při vývoji systémů strojového učení a jejich
- 15 „inteligentního“ využívání. Těžiště tvoří analýza a klasifikace textů, rozpoznávání jazyka a strojový překlad.

- Dnes je na trhu k dispozici celá řada služeb, které strojový překlad poskytují pro takřka všechny pro trh relevantní jazykové kombinace. Tyto služby mohou být do samotných procesů a technologií
- 20 integrovány snadno přes aplikační rozhraní (API).

- U systémů strojového překladu se rozlišují systémy s generickými překladovými modely (anglicky *generic* nebo *stock models*) a systémy s překladovými modely, které zohledňují potřeby klienta či možnosti domény (anglicky *custom models*). Generické překladové modely jsou založené
- 25 na obecných jazykových datech, zatímco modely přizpůsobené pro potřebu klienta či domény využívají k rozvoji odborná jazyková data tak, aby při překladu byly upřednostněny odpovídající jazyk, styl, idiomatika a terminologie.

- Strojový překlad se stal středem zájmu mnoha společností a organizací, a to nejen kvůli mediálnímu
- 30 pokrytí. Týká se to jak oblastí, které se jazykem a komunikací zabývají odjakživa (např. jazykové

a překladatelské služby), tak i oborů, jako je IT, marketing, vývoj produktů, zákaznické služby apod.

35 Strojový překlad se dnes uplatňuje v mnoha kontextech. Kromě toho, že podporuje práci profesionálních překladatelů, se používá k překládání a získávání informací (*gisting*) nebo k překládání v reálném čase v dialogových systémech (chaty, chatboti, jednotné zasílání zpráv). Požadavky na kvalitu strojových překladů se přitom liší podle oblasti a cíle využití, druhu textu a jazykového páru.

40 Neuronový strojový překlad (NMT – *Neural Machine Translation*) vedl v porovnání se statistickým strojovým překladem (SMT – *Statistical Machine Translation*) nezávisle na jazykové kombinaci k celkovému zlepšení kvality překladů, především se to týká plynulosti (*fluency*) a přesnosti (*accuracy*) (Koehn 2010, str. 217–246). Přestože už jsou strojové překlady pro účel předání informace dostatečné, musejí být před případnou publikací ještě posteditovány. Rozlišuje se 45 základní posteditace (*light post-editing*) a úplná posteditace (*full post-editing*). U základní posteditace stojí v popředí obsahová správnost a srozumitelnost, úplná posteditace k tomu zohledňuje jazyk a styl (ISO 18587:2017).

Jazykové a stylistické chyby může relativně snadno opravit překladatel nebo osoba, která ovládá 50 cílovou řeč, složitější ale bývá korektura chyb v terminologii. Vyžaduje totiž dodatečné znalosti problematiky daného oboru a zákaznických a firemních kontextů a preferencí. Čím dál častěji je posteditorem odborník, který ovládá více cizích jazyků (Nitzke a kol. 2019).

Jednou z možností, jak optimalizovat náklady na posteditaci a jak snížit náklady na opakované 55 opravování stejných terminologických chyb, je přizpůsobení systémů strojového překladu tak, aby se byly schopny naučit odbornou nebo klientovu terminologii a při generování překladu ji dokázaly upřednostnit před jinými možnými pojmenováními v cílovém jazyce. Výzkum na poli strojového překladu hovoří o přizpůsobení modelů strojového překladu vybraným oborům jako o *Domain Adaptation* (Chu a kol. 2018). Doménová adaptace fungovala už za časů statistického strojového 60 překladu. U doménové adaptace se rozlišuje mezi procesy vztahujícími se na data (*data centric*) a procesy vztahujícími se na modely (*model centric*). Procesy, které se vztahují na data, se soustředí na výběr, přípravu nebo generování odborných trénovacích dat pro trénování přizpůsobeného modelu strojového překladu (Moore a kol. 2010, Wang a kol. 2014, Chen a kol. 2016). Procesy, které se vztahují na model, naproti tomu pracují s přizpůsobováním cvičných algoritmů nebo

65 s laděním stávajících překladových modelů (Sennrich a kol. 2013, Durrani a kol. 2015, Zhou a kol. 2015). S nástupem neuronového strojového překladu začali vědci přebírat principy doménové adaptace také při vývoji neuronových překladových modelů.

70 Na trh se první možnosti doménové adaptace v komerčně dostupných systémech využívajících neuronový strojový překlad dostaly v roce 2017 (Globalease, Lilt). V roce 2018 se objevily i další systémy jako Google AutoML (Fei-Fei Li 2018), IBM Language Translator, Microsoft Custom Translator (Microsoft 2018) a SDL ETS 8.0 (Savenkov 2018). V první linii jsou využívány procesy vztahující se na data, které umožňují uživateli nahrávat data přizpůsobená doméně či klientovi. Tato data jsou pak použita ke zhotovení nebo přizpůsobení modelu strojového překladu.

75 Tento příspěvek se věnuje možnostem optimalizace výstupu komerčních systémů strojového překladu ve vztahu k terminologii. Kapitola 11.2 se nejprve zaměří na význam terminologie při překládání. V návaznosti na to jsou v kapitole 11.3 popsány příklady různých typických terminologických chyb při strojovém překládání, jakožto i jejich závažnost. Hlavní část tohoto článku tvoří kapitola 11.4, která se zabývá různými možnostmi přizpůsobení modelů strojových překladů tak, aby používaly správnou upřednostňovanou terminologii. V kapitole 11.5 jsou na závěr příspěvku shrnuty nejdůležitější poznatky.

11.2 Terminologie v překladu

85 Terminologie hraje při překládání textů, především těch odborných, centrální roli, protože je z velké části nositelkou významu. Má tedy přímý vliv na porozumění textu, ale také na právní jistotu a odpovědnost za výrobek. Kromě jednoslovných a víceslovných pojmenování patří k terminologii také odborné fráze (srov. Arntz 1999). Množství odborných pojmenování a hustota použití terminologie v textech se liší podle toho, do jaké míry je daný text odborný.

90 Používání správného, zadavatelem preferovaného nebo poskytnutého názvosloví v cílovém jazyce je jedním z klíčových požadavků na kvalitu překladu.

11.2.1 Porozumění textu: Terminologie ve výchozím textu

95 V praxi se překladatelé často v souvislosti s terminologií potýkají s problémy. Význam pojmenování ve výchozím textu nemusí být jednoznačný, nebo může být názvosloví ve výchozím

textu používáno bez sjednoceného významu, respektive nekonzistentně. Vyjasnění pojmů ve výchozím jazyce jakožto předpoklad pro porozumění výchozímu textu je tedy prvním krokem při práci na odborném překladu.

100

Překladaelé se v tuto chvíli opírají o různé druhy znalostí, jako jsou jazykové znalosti, informace o textu, odborné znalosti nebo informace o zadavateli. Mezi informační zdroje, které se používají při práci s terminologií v rámci překladu, patří kromě výchozích textů také odborná literatura z příslušného oboru, odborné slovníky, referenční texty zadavatele, glosáře, terminologické a informační databanky, terminologické příručky a samozřejmě internet. V případech pochybností ze strany zadavatele je nezbytné oslovit odborníka. Všechny tyto informace pomáhají utvářet kontext výchozího textu a kontext překladu. Při vyjasňování významu pojmů, na které je ve výchozím textu uvedena reference, postupuje překladatel podle sémantické analýzy (srov. Arntz a kol. 2014, str. 56 a dál). Podle té jsou zpracovány takové diskriminační charakteristiky, které oddělují jeden pojem od jiných pojmů.

110

11.2.2 Překlad terminologie: Určení ekvivalentů v cílovém jazyce

Jestliže překladatel porozuměl významu použitých pojmů a názvosloví ve výchozím textu, může začít s přenosem tohoto významu do cílového jazyka. Při vyhledávání ekvivalentů v cílovém jazyce používá překladatel kromě výše zmíněných informačních zdrojů také překladové paměti (*translation memory*) a paralelní texty.

115

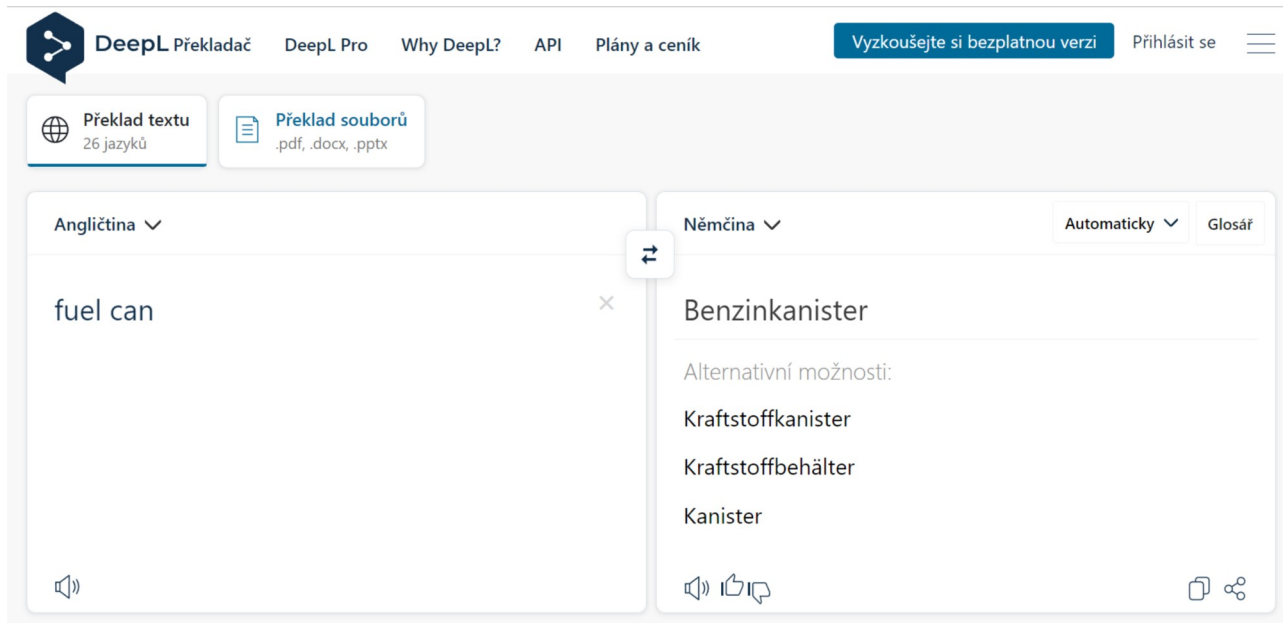
Možné varianty překladu pro pojmenování ve výchozím jazyce se nejprve identifikují, poté se mezi sebou navzájem porovnávají, dokud není nalezeno vhodné překladové řešení. I zde nachází sémantická analýza využití. Při ní se doporučuje podrobit každého kandidáta na překlad kontrole pomocí seznamu definovaných hodnotících kritérií pro jednotlivé pojmy. K těmto kritériím patří jednoznačnost, soulad se zákony/normami, transparence, jednotnost, míra používání, vhodnost v závislosti na cílové skupině, délka, výslovnost, jazykově-logická správnost, konotační svoboda, odvoditelnost a mezinárodnost (srov. Deutscher Terminologie Tag e.V. 2014, M3-16-24).

125

V cílovém jazyce neexistuje vždy přesně odpovídající ekvivalent pro odborné nebo klientem specifikované označení (srov. Arntz a kol. 2014, str. 158). Proto musí překladatel daný pojem nebo předmět diskuze buď popsat jinými slovy, použít přejímku nebo vytvořit kalk, případně prosadit nové pojmenování v cílovém jazyce (srov. Arntz a kol. 2014, str. 148–149).

130 11.2.3 Terminologie v generických modelech strojového překladu

Generické modely strojového překladu jsou trénovány prostřednictvím milionů větných párů. Věty se netýkají pouze konverzačních témat, zohledněna jsou také odborná témata z různých odvětví. To vede k tomu, že v překladovém modelu může pro jeden odborný výraz ve výchozím jazyce existovat v závislosti na kontextu více překladů. Platí tedy, že generické překladové modely neznají žádné globálně upřednostňované názvosloví.



Překladač DeepL: <https://deepl.com/translator>, naposledy načteno 20. 3. 2022

Jaké pojmenování bude nakonec zvoleno jako vhodný překlad, záleží na kontextu. U většiny systémů strojového překladu se na tomto místě mluví o „větě určené k překladu“, respektive o „segmentu určeném k překladu“. U profesionálního překladu, při němž se využívají překladové nástroje, jsou do systému strojového překladu většinou zaslány pouze ty segmenty, pro které nebyl nalezen ekvivalent s dostatečně vysokým počtem výskytů v překladové paměti. To znamená, že systém strojového překladu nemá přístup ke kontextu – nezná předchozí ani následující větu, natož pak celý výchozí text.

Rozhodnutí, které pojmenování bude zvoleno v cílovém jazyce, tedy závisí v těchto případech pouze na bezprostředním textovém kontextu. V praxi to vede k nekonzistentním překladům, což dobře ukazuje následující příklad:

Původní věta (EN)	Strojový překlad (EN/IT)	Doslovný překlad (IT/CZ)¹
I bought a red <i>fuel can</i> .	Ho comprato una <i>tanica di benzina</i> rossa.	Koupil jsem červený kanystř na benzín.
I bought a green <i>fuel can</i> in the online shop.	Ho comprato una <i>tanica di carburante</i> verde nel negozio online.	Koupil jsem zelený kanystř na palivo v internetovém obchodě.
I bought a golden <i>fuel can</i> .	Ho comprato un <i>bidone</i> d'oro.	Koupil jsem plechovku zlata.

Překladač Google: <https://translate.google.com>, naposledy načteno 15. 1. 2020

150 Další příklad potvrzuje, že systémy strojového překladu při volbě vhodné terminologie v cílovém jazyce nedisponují znalostí širšího kontextu.

Původní věta (DE)	Strojový překlad (DE/IT)	Doslovný překlad (IT/CZ)
Ich habe bei OBI im Angebot eine Bank für 45,00 EUR gekauft.	Ho comprato una banca da OBI per 45,00 EUR.	Koupil jsem banku z OBI za 45 eur.

Překladač Google: <https://translate.google.com>, naposledy načteno 9. 1. 2020

155 Význam slova *die Bank* (lavice, lavička, místo určené k sezení) nebyl v tomto případě správně převeden do italského jazyka. Stalo se tak navzdory upozorňujícím, v originální větě přítomným slovům OBI (jméno obchodu se stavebními potřebami), ve slevě (velmi zřídka se koupí se slevou nabízejí finanční instituce) a 45 eur (kupní cena, která s větší pravděpodobností odpovídá ceně lavičky). I když mohou být terminologické chyby v některých případech závažné, představuje tento problém v oblasti profesionálního odborného překladu pouze zanedbatelné riziko (srov. Ottmann /Canfora 2020). Obecně však tyto chyby vedou k mnohem větší nutnosti posteditačních zásahů u překladů odborných textů.

160 Chyby v terminologii v generických strojových překladech se pohybují od málo závažných, jako jsou různé varianty pravopisu, které neovlivňují porozumění textu, až k závažným významovým

¹ Doslovné překlady ve všech tabulkách: Vlasta Řeňčová, březen 2022

posunům. V následující kapitole je uveden přehled různých typických chyb, které se ve strojových překladech vyskytují.

165 11.3 Terminologické chyby ve strojových překladech

Terminologické chyby ve strojových překladech se pohybují od málo závažných, jako jsou chyby v pravopisu či ohýbání slov, až k závažným chybám, které ovlivňují význam a konzistenci textu. Následující faktory ovlivňují pravděpodobnost a četnost výskytu chyb.

170 **Přirozené množství synonym a/nebo pravopisných variant:** Čím víc má nějaký pojem v dané řeči přirozených synonym nebo čím víc existuje variant, jak pojem korektně zapsat (například psaní *s* či *z* ve slovech cizího původu – filosofie/filozofie, president/prezident), tím méně pravděpodobné je, že data, která byla pro generický překladový model použita, budou konzistentní. U strojového překladu se dá následně očekávat několik variant překladu.

175

Přechod z odborné řeči k hovorové: Blízkost odborné řeči k hovorové zvyšuje pravděpodobnost, že v tréninkovém materiálu generického překladového modelu bylo zahrnuto odpovídající názvosloví. Jejich správný strojový překlad je v těchto případech tedy pravděpodobnější. Například pojmy z oborů informačních technologií nebo mechaniky jako „server“ nebo „náprava“ jsou zastoupeny v tréninkovém materiálu generického překladového modelu mnohem četněji, než je tomu u velice specifické terminologie z železničního provozu, jako je například německé slovo *Gleisfreimeldeanlage* (kolejový obvod).

180

Složitost názvosloví: Je logické, že čím je pojmenování složitější (například kompozita, sousloví či víceslovné termíny), automaticky tím u strojového překladu roste také náchylnost k chybám. Pokud se na tuto situaci podíváme matematicky, zjistíme, že za předpokladu průměrného počtu pěti variant na jednu součást pojmenování je pro jednoslovná pojmenování nalezeno pět možných variant. Jedná-li se však o kompozitum nebo o dvouslovný termín, vzrůstá počet variant na 25.

185

190 Jazykové experimenty tyto úvahy potvrzují: Při překladu tří vět, které obsahovaly specifický termín *Gleisfreimeldeanlage*, do českého jazyka pomocí vždy čtyř překladačů² se vytvořily čtyři varianty:

2 Věty, na nichž byl testován strojový překlad: Die Software für alle Einrichtungen der Gleisfreimeldeanlage musste aktualisiert werden. / Das einwandfreie Funktionieren der Gleisfreimeldeanlage ist von wesentlicher Bedeutung. / Die Reparatur der Gleisfreimeldeanlage hat viel Zeit in Anspruch genommen. Použité enginy: Překladač Google: <https://translate.google.com>; Překladač DeepL: <https://deepl.com>, IBM Watson: <https://language-translator->

systém detekce volných míst kolejí, systém detekce volných míst na kolejích, systém detekce volných míst na trati, systém detekce volných míst. Pro totožný text s větami, které místo kompozita *Gleisfreimeldeanlage* obsahovaly jednoslovný termín *Weiche* (výhybka), bylo nalezeno šest variant překladu do češtiny: spínač, vypínač, spínací zařízení, rozjezdové zařízení, rozjezdová skříň, výhybka.³

11.3.1 Významové chyby

K významovým chybám se řadí takové chyby, které posouvají nebo přímo mění význam slova. Tyto chyby se často vyskytují v následujících případech.

Homografy: Homograf je slovo, které má alespoň dva významy. Z toho přirozeně vyplývá, že riziko sémanticky odlišného chybného překladu je zde vyšší než u pojmů, které mají pouze jeden význam:

Originál (EN)	Strojový překlad (EN/DE)	Doslovný překlad (DE/CZ)
Fuel can	Kraftstoff kann	Palivo může
Fuel cans	Kraftstoffdosen	Kanystry s palivem
A fuel can	Eine Kraftstoffdose	Kanystr s palivem

Překladač Microsoft, <https://bing.com/translator>, naposledy načteno 9. 1. 2020

Riziko významové chyby se zvyšuje úměrně k množství poskytnutého kontextu pro segment určený k překladu. Částečně stačí velmi málo kontextu, například neurčitý člen, který, tak jako ve výše uvedeném příkladu, pomůže rozlišit mezi slovesem a podstatným jménem a tedy vygenerovat správný překlad.

Na následujícím příkladu se dá dobře poznat, jak kontext ovlivňuje výstup strojového překladu: Spojení *concrete steps* přeložil Překladač DeepL do němčiny jako *konkrete Schritte* (konkrétní kroky). Pokud ale rozšíříme větu o *concrete steps are painted*, je slovo *steps* přeloženo jako *Betonstufen* (betonové schody):

demo.ng.bluemix.net/, Překladač Microsoft: <https://www.bing.com/translator/>, stránky naposledy načteny 12. 3. 2022.

³ Je zajímavé, že zatímco při překladu do francouzského jazyka vytvořily překladače více variant pro *Gleisfreimeldeanlage* (10) než pro *Weiche* (4), v českém jazyce je tento poměr 4:6. To dokazuje, jak významnou roli při strojovém překladu hraje také jazykový pár, v rámci kterého se překládá. (Poznámka překladatelky)

Původní věta (EN)	Strojový překlad (EN/DE)	Doslovný překlad (DE/CZ)
Concrete steps	Konkrete Schritte	Konkrétní kroky
Concrete steps are painted	Betonstufen werden gestrichen	Betonové schody se natírají

Překladač DeepL: <https://deepl.com/translator>, naposledy načteno 9. 1. 2020

- 220 **Akronymy** přebírá překladač do cílového textu zpravidla beze změny, nebo je nahrazuje možným odpovídajícím ekvivalentem v cílovém jazyce. Může se jednat například o běžnou zkratku nebo o plný tvar slova. Riziko chybného překladu je v obou případech velmi vysoké, protože akronym zůstává jen zřídka stejný napříč jazyky a systém strojového překladu je z důvodu vysokého a rostoucího množství variant, respektive víceznačných akronym, jen zřídka schopný je s jistotou
- 225 a správností rozluštit.

Původní věta (FR)	Strojový překlad (FR/DE)	Doslovný překlad (DE/CZ)
L'ADC demande la fermeture de la voie contigue pour aller á la visite de son train.	Der ADC fordert die Schließung der angrenzenden Gleise an, um seinen Zug zu besuchen.	ADC požaduje uzavření přilehlé koleje, aby mohl jít na návštěvu svého vlaku.

Překladač DeepL: <https://deepl.com/translator>, naposledy načteno 9. 1. 2020

- Namísto použití německého ekvivalentu *Triebfahrzeugführer* (strojvedoucí), případně německého
- 230 akronymu *Tf*, je ve výše uvedeném příkladu převzatý akronym beze změny z výchozího textu. I ti, kdo rozumí německému železničnímu žargonu, nerozklíčují význam této věty, pokud nehovoří francouzsky. V tomto případě se tedy jedná o závažnou chybu. Podobně je tomu i v následujícím příkladu:

Původní věta (FR)	Strojový překlad (FR/DE)	Doslovný překlad (DE/CZ)
L'ADC demande la fermeture de la voie contigue pour aller á la visite de son train.	Die CDA beantragt die Schließung des angrenzenden Gleises, um den Besuch ihres	CDA požaduje uzavření přilehlé koleje, aby umožnila

	Zuges zu ermöglichen.	návštěvu svého vlaku.
--	-----------------------	-----------------------

235 Chybné převedení akronymu ADC na CDA, Překladač DeepL: <https://www.deepl.com/translator>, naposledy načteno 9. 1. 2020.

11.3.2 Vynechávání

240 Strojové překlady někdy vynechají části pojmů nebo celá slova, což může vést k sémantickému posunutí nebo ke ztrátě původního významu.

Původní věta (DE)	Strojový překlad (DE/FR)	Doslovný překlad (FR/CZ)
Die Bediener der Stellische sind nicht an ihrem Platz.	Les opérateurs des tables ne sont pas a leur place.	Obsluhující stolů nejsou na svém místě.

Překladač DeepL: <https://www.deepl.com/translator>, naposledy načteno 9. 1. 2020

245 Redukce slova *Stelltisch* (pult reléového zabezpečovacího zařízení) na pouhé slovo *Tisch* (stůl) v tomto příkladu značně rozšiřuje možnosti, jak lze větu interpretovat. V případě slova *Bediener* (obsluhující, operátor), se může čistě teoreticky jednat o číšníka.

Původní věta (FR)	Strojový překlad (FR/DE)	Doslovný překlad (DE/CZ)
Ici, ADC 4712: le train circulant sur la voie contigue a un frein bloqué et il y a un jet d'étincelles.	Gestern ADC 4712: der auf dem Nachbargleis fahrende Zug hat eine blockierte Bremse und es gibt einen Funken.	Včera ADC 4712: vlak jedoucí po sousední koleji má zablokovanou brzdu a je tam jiskra.

Překladač Google, <https://translate.google.com/>, naposledy načteno 9. 1. 2020

250 Při redukci slova *Funkenstrahl* (proud jisker) na pouhou jiskru (*Funken*) ve druhém příkladu rozhoduje o závažnosti takové chyby kontext – pokud bychom mluvili o stodole v letním období, může mít jedna jiskra vážně dopady. Naproti tomu občasné jiskření v silně mechanickém prostředí železniční dopravy žádné nebezpečí nepředstavuje.

Ještě závažnější problém ve strojových překladech představuje vynechávání celých slov. Bez znalosti výchozí věty je v těchto případech porozumění překladu často nemožné, nebo
255 přinejmenším značně ztížené.

Původní věta (EN)	Strojový překlad (EN/DE)	Doslovný překlad (DE/CZ)
A 5 liter fuel can can be purchased online.	Ein 5-Liter-Kraftstoff kann online erworben werden.	Pětilitrové palivo lze zakoupit online.

Vynechání informace „kanystr“ ze spojení *fuel can* (kanystr na benzín). Překladač Microsoft, <https://www.bing.com/translator>, naposledy načteno 9. 1. 2020.

Tyto chyby se dají pozorovat napříč různými jazykovými páry. Strojový překlad vyhotovený
260 Překladačem DeepL anglického *fuel can* (kanystr na palivo) do italštiny postrádá význam slova *fuel* (palivo). Místo toho byl v překladu použit pojem *lattina*, který se používá spíše ve smyslu plechovka či konzerva v souvislosti s potravinami. Preferovanými variantami překladu by v kontextu umělohmotných produktů byla italská slova *serbatoio* (nádrž) nebo *tanica* (kanystr).

Původní věta (EN)	Strojový překlad (EN/IT)	Doslovný překlad (IT/CZ)
Fuel cans	Barattoli di carburante	Kanystry na pohonné hmoty
5 liter fuel cans	Lattine da 5 litri	Plechovky s pěti litry

Překladač DeepL: <https://deepl.com/translator>, naposledy načteno 9. 1. 2020.

265 11.3.3 Chyby ve flexi

Obecně se dá říct, že chyby ve flexi se v neuronových strojových překladech vyskytují o poznání
270 řídkěji, než je tomu u statistických strojových překladů. Častěji na ně narazíme v krátkých segmentech, které mají málo kontextu. V následujícím příkladu nebylo adjektivum *grün* (zelená) vyskloňováno. Změníme-li však kontext ve výchozí větě a místo *green* použijeme *blue*, výsledný překlad bude v tomto případě už vyskloňovaný správně:

Původní věta (EN)	Strojový překlad (EN/DE)	Doslovný překlad (DE/CZ)
Green fuel can	Grüne Benzinkanister	Zelená kanystr na benzín
Blue fuel can	Blauer Benzinkanister	Modrý kanystr na benzín

Překladač DeepL: <https://deepl.com/translator>, naposledy načteno: 9. 1. 2020.

11.3.4 Inkonzistence

Jak již bylo popsáno výše, generické překladové modely jsou založeny na množství textových dat z různých zdrojů. Mezi těmito textovými daty jsou různé druhy textů, kde se používají jiné styly a jiná terminologie. Jejich kvalita se často různí. Této inkonzistenci odpovídají také strojové překlady. Seznam dostupných pojmenování, která nabízejí systémy strojového překladu jako alternativní překlady, jen podtrhuje celkové množství variant (viz příklad anglického *fuel can* v kapitole Terminologie v generických modelech strojového překladu). Překladač Google obohacuje navržené alternativy také definicemi slov nebo údajem o frekvenci. Přesto se stává, že je v určitých situacích zvoleno méně používané a/nebo hovorové pojmenování, což lze vidět na následujícím příkladu.

Ačkoliv jsou v němčině používanějšími alternativami k anglickému *fuel* slova *Kraftstoff* či *Brennstoff* (palivo), je v překladu zvoleno slovo *Sprit* (š'áva, hovorově benzín).

Původní věta (EN)	Strojový překlad (EN/DE)	Doslovný překlad (DE/CZ)
A red 5 liter fuel can costs 25,99 EUR.	Ein roter 5 Liter Sprit kostet 25,99 EUR.	Červený pětilitrový kanystroběna stojí 25,99 eur.

Překladač Google, <https://translate.google.com/>, naposledy načteno: 9. 1. 2020

K chybám v konzistenci se počítá i nekonzistentní pravopis, například varianty se spojovníkem nebo bez něj, v angličtině například kolísavé užívání *s* (britská angličtina) a *z* (americká angličtina), v němčině obdobné *ß* (Německo) a *ss* (Švýcarsko), v češtině rozdílný pravopis slov cizího původu (*s* a *z*, krátké a dlouhé samohlásky).

V praxi to znamená, že pojmenování zvolená generickým systémem strojového překladu nejsou v rámci jedné zakázky nutně přeložena jednotně. U odborných překladů tedy musí posteditor zvlášť dbát na terminologickou správnost.

11.3.5 Závažnost

Uvážením kategorie, do níž chyba spadá (například *minor defects* (drobné závady), *major defects* (velké závady), *critical defects* (závažné závady) (viz DIN 40080), se dá vyvodit závažnost chyby, kterou pak strojový výstup dlouhodobě využívá k porovnávání možností, identifikování

300 a definování chyby na základě požadavků projektu. Toto uvážení by mělo proběhnout jednotlivě pro každý projekt, protože požadavky se mohou lišit. Významové posuny nebo chyby se zpravidla hodnotí vyšší závažností. Používání odlišných pravopisných pravidel může být ale v některých kontextech přípustné.

Uvedené příklady ukazují, jak často generický strojový překlad na poli terminologie chybuje. Aby se snížilo množství chyb a tím pádem i případné náklady na jejich opravu, mohou být překladové modely trénovány na určité stylistické, idiomatické a terminologické použití jazyka.

11.4 Trénink a customizace překladačů

310 K přizpůsobení překladových modelů z hlediska odbornosti či pro požadavky projektu či klienta se nabízejí následující možnosti:

Trénink znamená, že nový překladový model bude vyhotoven „od nuly“. Tato možnost však předpokládá bohatý tréninkový materiál (tedy alespoň milion dvojjazyčných segmentových párů), který pro většinu případů, respektive pro většinu oborů a jazykových kombinací, není dostupný v dostatečném množství a kvalitě.

315 Pro dvojjazyčný trénink jsou potřeba překladová data ve výměnném formátu TMX (*Translation Memory Exchange Format*). Dalšími podporovanými datovými formáty jsou dvojjazyčná pracovní data ze systémů překladové paměti, například MQXLIFF, MQXLZ, MXLIFF, SDLXLIFF, TXLF, XLIFF, XLF nebo XLZ. Někteří poskytovatelé, mezi které patří například Microsoft, podporují využívání paralelních textů, které jsou automaticky zarovnané. Zde patří k podporovaným formátům TXT, HTML, PDF a DOCX.

Retraining a customizace (také *domain adaptation*) znamená, že jsou stávající generické překladové modely přizpůsobeny na základě specifických překladových dat pro daný projekt či klienta. Pro tuto možnost jsou dostatečné i menší objemy dat – záleží na poskytovateli, u některých lze hovořit o 5 000 až 15 000 segmentech. Dat může být přitom i více. Podle shodujících se 325 výpovědí různých poskytovatelů (Globalese, Microsoft) však platí, že se nárůst kvality s nárůstem dat proporcionálně snižuje, respektive může od určitého bodu i stagnovat.

Na tomto místě je vhodné upozornit na skutečnost, že trénink a customizace často nerozlišují aktuální jazykové použití, a běžně tedy dochází k používání synonym.

Někteří poskytovatelé umožňují při tréninku a customizaci **upřednostnění** (priorizaci) vybraných překladů (Globalese: *Master vs. Auxiliary Corpora*, Microsoft: *Tuning data*). Pokud například 2 000 z 15 000 segmentů odpovídají svým stylem, idiomatikou a/nebo terminologií přesně cíli projektu, mohou být pomocí dat pro ladění (*tuning data*) při tréninku ohodnoceny vyšším stupněm.

Díky **začlenění terminologie** (hovoříme o glosářích a slovnících) lze docílit kýženého použití stanovených slov a frází. Většina systémů strojového překladu umožňuje nahrát terminologická data ve formátech XLSX, CSV a TSV. Systémy jako Globalese navíc podporují také používání TBX datových souborů. Začlenění terminologie funguje ovšem u většiny poskytovatelů čistě na základě vyhledávání a nahrazování ustálených textových řetězců, což ale omezuje používání pouze na výrazy, u nichž nedochází k ohýbání (například slogany nebo názvy společností jako *Deutsche Bahn AG*). Navíc se dá očekávat potenciální zvýšení četnosti gramatických chyb. Microsoft tedy například otevřeně odrazuje od používání terminologie ke customizaci a v dokumentaci vlastního překladače (*custom translator*) ovlivňovat terminologii výhradně na základě korpusů.

Aby bylo možné ohodnotit kvalitu přizpůsobeného modelu strojového překladu, lze použít **testovací dataset**. Ten je jakýmsi „zlatým standartem“. Zahrnuje pevně nastavený a poté již neměnicí se dataset, v němž jsou dvojjazyčné segmentové páry vztahující se k projektu. Pomocí těchto dat se výstup strojového překladu neustále doladuje. Množství testovacích dat se liší poskytovatel od poskytovatele. Zpravidla stačí zhruba 2 000 dvojjazyčných segmentových párů pro jedno vyhodnocení. Pomocí porovnání strojových překladů s referenčními překlady v souboru testovacích dat je u většiny systémů strojového překladu automaticky vyhotoveno BLEU skóre. To udává, v jaké míře se strojový překlad blíží referenčním překladům. Skóre má sice menší vypovídající hodnotu ve vztahu ke kvalitě překladu (srov. Dougal 2018, s. 34), ale je dobrým indikátorem pro zjištění, zda trénink nebo retraining překladového modelu přiblížil výstup očekávaným výsledkům. Někteří poskytovatelé automaticky generují testovací dataset z náhodně vybraných segmentových párů z tréninkového materiálu, pokud uživatel neposkytne žádná vlastní testovací data. Zpravidla se ovšem doporučuje určit cílený testovací dataset, který souvisí s projektem.

Ve výzkumu strojového překladu existují i další přístupy k integraci terminologických a ontologických zásob. Jedním z takových přístupů je například tzv. *Term-Injection* (viz Dougal, 2018, s. 21–27). Při tomto procesu se prostřednictvím zásahu do neuronové sítě importuje terminologie a flexe dotyčných pojmů, jakožto i všech závislých slov, je morfosyntakticky přizpůsobena. Protože tyto přístupy nejsou doposud dostupné na trhu, soustředíme se v této kapitole

pouze na v současnosti nejpoužívanější metody: přizpůsobení generických překladových modelů pomocí retrainingu, respektive customizace.

11.4.1 Trénink a customizace jako projekt

365 Hlavní výzvu při plánování vlastního projektu představuje správný odhad časových a finančních nákladů. Samotný retraining lze sice relativně dobře naplánovat, ale zabírá pouze malou část celkového projektu. Mnohem náročnější částí je opatření a zpracování tréninkových dat a také evaluace strojově přeloženého výstupu.

370 Časové náklady na trénink a retraining jsou zpravidla v přímém vztahu k velikosti tréninkového datasetu a počtu jazykových párů. Většina poskytovatelů buď vypočítá cenu za 1 000 000 znaků nebo účtují celkový výpočetní čas. Záleží na poskytovatelích, jestli do nákladů zahrnou také náklady za hosting přizpůsobených překladových modelů či za získávání překladů na základě znaků ve výchozím textu. Tak jako se liší ceny podle poskytovatele (pohybují se od 5 do 50 eur za 375 1 000 000 přeložených znaků), odlišují se i doby trvání tréninku: zatímco retraining s 20 000 segmenty může být u některých poskytovatelů hotový za méně než hodinu, u jiných to trvá více hodin (viz také Savenkov 2018).

Náklady na sběr a zpracování tréninkových dat a na evaluaci výstupu strojového překladu se odvíjejí od cílů projektu. Tyto činnosti se musí často několikrát zopakovat. Čím vyšší jsou nároky na výslednou kvalitu, tím víc roste nutnost trénink několikrát zopakovat, dokud není dosaženo 380 cílené kvality.

Z hlediska plánování může být projekt customizace rozdělen do tří fází:

11.4.1.1 První fáze: Plánování a příprava

V této fázi je třeba zohlednit kromě jiného následující aspekty:

385 **Definice cílů projektu a kritérií validace**

stanovit kvalitativní požadavky na výstup strojového překladu

definovat chybové kategorie a jejich závažnost při evaluaci

Výběr tréninkového materiálu

výběr vhodného tréninkového materiálu určují obor činnosti, styl, idiomatika a terminologie

390 Sběr a ověření materiálu: Jaké množství dat je k dispozici? Tip: Množství dat se může při pozdějším zpracování dat v systémech strojového překladu před tréninkem ještě snížit. V některých případech to vede k tomu, že již není dosaženo minimálního objemu dat.

Kvalita dat: Odpovídají data požadavkům na kvalitu ve vztahu k jazykové variantě, druhu textu, obsahu, ortografii, konzistenci, formátu atd.?

395 **Evaluace poskytovatele strojového překladu**

ověřit dostupné jazykové kombinace

otestovat kvalitu jazyka

porovnat přímé a nepřímé jazykové dvojice

porovnat ceny

400 ověřit kritéria zabezpečení dat

případně vzít v úvahu další projektová kritéria či požadavky klienta

11.4.1.2 Druhá fáze: Zpracování a generování dat

Zpracování dat

405 Při zpracování dat jsou připravena tréninková data pro použití v konkrétním systému strojového překladu. Vedle konverze dat do formátu, který podporují poskytovatelé, se na datech často testují různá kvalitativní kritéria, případně jsou data pročištěna. Zohledňují se přitom následující aspekty:

Sémantická správnost: Aby se zabránilo významovým posunům ve výstupu strojového překladu, je třeba, aby si segmenty ve výchozím a cílovém textu bezvýhradně odpovídaly. Častým zdrojem chyb jsou segmenty, kde byla výchozí věta při překladu rozložena do několika vět, nebo prázdné segmenty.

Konzistence: Přinejmenším v segmentech v cílovém jazyce by mělo být prověřeno konzistentní používání terminologie. Inkonzistence ve výchozím textu jsou ale pro překladač spíš přínosné, neboť ten tímto způsobem může rozpoznávat synonyma, která jsou v cílovém jazyce konzistentně reprezentována jedním pojmenováním.

Upozornění: Mají-li být inkonzistence ve výchozím textu ponechány, nelze tréninkový materiál použít pro překlad v obou směrech! Aby byla tréninková korpora použitelná v obou

jazykových směrech, musí být buď konzistentní, nebo musí existovat pro každý jazykový směr samostatný korpus.

420 **Segmentace:** Neúplné gramatické jednotky (například výčty) negativně ovlivňují strojové zpracování správně zformulovaných vět. Pokud to není v rozporu s požadavky projektu, měla by korpora obsahovat výhradně celé věty.

Formátování: Vymazání formátovacího tagu, například prostřednictvím skriptů, většina poskytovatelů nepožaduje; v segmentech bez tagů je však mnohem snadnější zkoumat výše uvedené
425 aspekty.

Anonymizace a pseudonymizace osobních dat: Údaje objevující se v textech, jako jsou vlastní jména, uživatelská jména, data narození, adresy, telefonní čísla, emailové adresy atd., by měly být anonymizovány nebo pseudonymizovány, aby se nestaly součástí překladového modelu.

Generování dat

430 Pokud není k dispozici dostatek vhodných tréninkových dat, nabízí se možnost taková data vygenerovat. Pomocí zástupných znaků lze například různými způsoby reprodukovat stávající dvojjazyčné texty nebo texty vytvořené v malém počtu. K tomu je nutný následující postup:

Rozpoznání určitých, stále se opakujících elementů v textu, které nemají centrální význam: například příjmení, názvy měst atd.

435 **Nahrazení** těchto elementů pro ně definovanými entitami: například {surname} pro všechna příjmení a {city} pro všechny názvy měst.

Definování hodnoty, kterou budou zástupné znaky (automaticky) nahrazeny: Pro entitu {city} bude například uložen seznam hodnot čítající 50 různých názvů měst.

Generování textů, kde budou entity obsazeny různými hodnotami (čím víc entit je v jedné větě
440 rozpoznáno, tím častěji se dá tato věta různými způsoby reprodukovat).

Původní věta:

německy: Übermittlungscod 16318, Dein Name ist Maier und ich heisse Schmitz.

francouzsky: Code d'autorisation 16318, tu t'appelles Maier et moi, je m'appelle Schmitz.

česky: Přenosový kód 16318, Tvé jméno je Maier a já se jmenuji Schmitz.

Věta s entitami:

německy: Übermittlungscod {code}, Dein Name ist {surname} und ich heisse {surname}.

francouzsky: Code d'autorisation {code}, tu t'appelles {surname} et moi, je m'appelle {surname}.

česky: Přenosový kód {code}, Tvé jméno je {surname} a já se jmenuji {surname}.

Automaticky vygenerované věty:

německy Übermittlungscod 16318, Dein Name ist Walter und ich heisse Lehmann.

francouzsky: Code d'autorisation 16318, tu t'appelles Walter et moi, je m'appelle Lehmann.

česky: Přenosový kód 16318, Tvé jméno je Walter a já se jmenuji Lehmann.

německy Übermittlungscod BHG-12344, Dein Name ist Maier und ich heisse Bergmann.

francouzsky: Code d'autorisation BHG-12344, tu t'appelles Maier et moi, je m'appelle Bergmann.

česky: Přenosový kód BHG-12344, Tvé jméno je Maier a já se jmenuji Bergmann.

německy Übermittlungscod F-633, Dein Name ist Schmitze und ich heisse Katz.

francouzsky: Code d'autorisation F-633, tu t'appelles Schmitze et moi, je m'appelle Katz.

česky: Přenosový kód F-633, Tvé jméno je Schmitze a já se jmenuji Katz.

(Příklad customizace v projektu společnosti *Deutsche Bahn AG*)

Před generováním textu se vždy doporučuje zkontrolovat, zda jsou všechny zástupné znaky entit, jakožto i entity samotné, správně umístěny a správně rozloženy ve výchozím a cílovém textu. V opačném případě hrozí, že se vygenerují nesprávná tréninková data.

- 445 Používání entit a generování textu lze automatizovat pomocí skriptů. Vytvoření a přizpůsobení skriptu ovšem vždy znamená, že se zvýší počáteční náklady a náklady na údržbu, neboť i skripty, které jsou již k dispozici, musí být přizpůsobeny textům konkrétního projektu.

11.4.1.3 Třetí fáze: Trénink, evaluace a retraining

- 450 Když jsou tréninková data, testovací data a data pro ladění připravena, jsou nahrána přes webovou stránku nebo přes API do systému strojového překladu. U většiny poskytovatelů jsou tréninková data následně shromážděna do korpusu. Pro trénink či retraining překladového modelu se vybere

455 korpus a začne samotný trénink. Jakmile je proces u konce, uživatel dostane oznámení. Oznámení obsahuje většinou dodatečné informace k průběhu tréninku a k výsledku automatické evaluace na základě testovacího datasetu.

460 Při automatické evaluaci přizpůsobených překladových modelů se v současnosti u všech poskytovatelů uplatňuje BLEU skóre. Toto měření je založeno na předpokladu, že strojový překlad je tím lepší, čím víc se podobá referenčnímu překladu, který vyhotovil člověk. Podobnost se počítá na základě shody n-gramů. Hodnota leží mezi 0 a 1 a často se udává v procentech. BLEU skóre slouží jako první indikátor zdařilého tréninku. Jeho výpovědní hodnota je však sporná, protože systémy neuronového strojového překladu mohou generovat alternativní překlady, které sice nejsou podobné referenčním překladům, ale mohou být přesto správné. Alternativně mohou být provedeny i další automatické evaluace jinými způsoby měření (například TER, WER, HTER, hLEPOR atd.). K ověření specifických kvalitativních kritérií, jako je dodržení terminologie, se v praxi přidávají i další automatické procesy, jako je kontrola kvality, terminologie, nebo speciální skripty.

470 Náklady za následnou manuální kategorizaci a zvážení chyb mohou být automatizováním evaluace podstatně sníženy. Manuální analýza je samozřejmě nutná, aby se konkretizovaly slabiny překladače a aby se pro ně v příštím retrainingu generovala cílená data. Na tomto principu se dají kontrolovat také identifikační čísla a minimální požadavky, které zaručují dlouhodobou porovnatelnost vývoje překladače a které rozhodují o počtu dalších retrainingů.

Na principu výsledků evaluace, tj. na počtu chyb v jednotlivých kategoriích a jejich závažnosti, může pak být cíleně sesbírán či vygenerován další tréninkový materiál k požadované optimalizaci a korektuře určitých chyb. Proces tréninku a následné evaluace pak začíná znovu.

475 **11.4.1.4 Ukončení projektu**

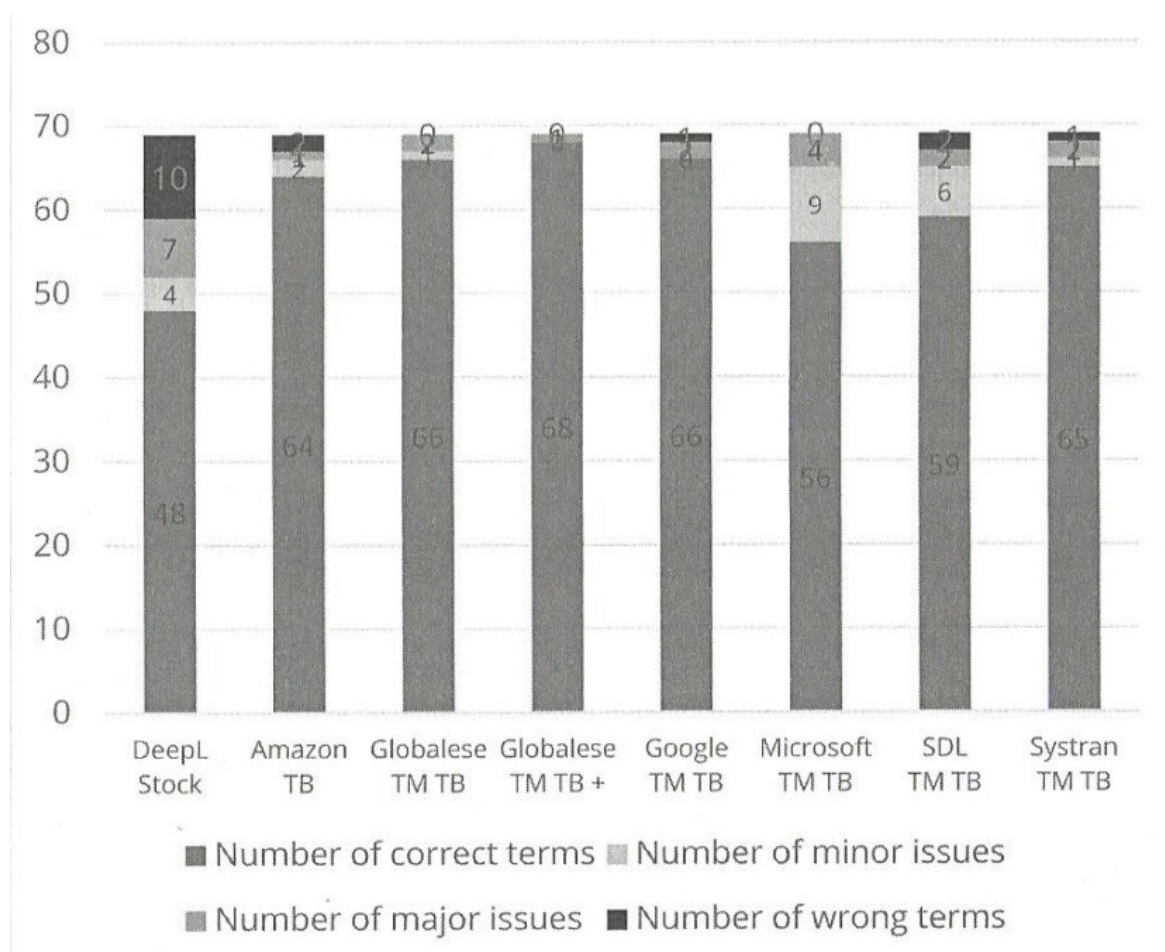
Jakmile je dosaženo požadované kvality překladu, může být nasazen přizpůsobený překladač. Ten musí být u jednotlivých poskytovatelů nejdříve aktivován, respektive zprovozněn.

480 Než jsou tréninkový a customizační projekt ukončeny, doporučuje se zdokumentovat získané zkušenosti z hlediska jazykového páru, množství dat, výsledku evaluace, času, nákladů atd., aby nabyté poznatky mohly sloužit ke konstantnímu zlepšování budoucích projektů.

11.4.2 Výsledky a možnosti optimalizace

Že se trénink nebo retraining překladových modelů vyplatí, je mezi odborníky na strojový překlad všeobecně známo. Míra optimalizačního potenciálu závisí na různorodých faktorech. Patří k nim mimo jiné například jazykový pár (příbuzné a vzdálené jazyky), množství dat a také jejich kvalita. V porovnání s generickými překladovými modely, které v závislosti na oboru správně přeloží jen mezi 5 % a 60 % pojmů, se dá dodržování terminologie pomocí customizace optimalizovat na víc než 90 % (viz Zielinski / Vardaro 2019).

490 Terminologické chyby v generickém Překladači DeepL a v přizpůsobených modelech jiných poskytovatelů



Porovnání dodržování terminologie v generickém překladovém modelu DeepL vs. v přizpůsobených překladových modelech jiných poskytovatelů v jazykové kombinaci angličtina – němčina. (viz Zielinski / Vardaro 2019)

495 V jednom projektu společnosti *Deutsche Bahn AG* mohlo být například díky retrainingu s pouhými 16 000 segmenty (z toho 453 segmentů s odpovídajícím terminologickým párem) zvýšeno správné používání německého *Gegengleis* (traťová kolej pro opačný směr jízdy) a francouzského *voie*

de sens contraire z 8 % (4 z 50) na 86 % (43 z 50). Další trénink s 71 000 segmenty (z toho 2012 segmentů s odpovídajícím terminologickým párem) zvýšil správné používání na 90 % (45 z 50).

500 Tento příklad názorně ukazuje, že optimalizace není proporcionální vůči množství nasazených tréninkových dat. Z důvodu komplexity jazyka a z toho vyplývajícího neomezeného množství možných příčin dalších chybných výsledků (10 %) se náklady na další optimalizaci naceňují vysoko, takže možnou dosažitelnou dokonalost neberou v úvahu.

11.5 Shrnutí a vyhlídky

505 Retraining modelů strojového překladu se ukázal být efektivním prostředkem k přizpůsobování výkonu generických překladových modelů požadavkům odborné řeči. Roste přitom také používání tréninku a customizace, zpravidla v závislosti na stupni odbornosti textů určených k překladu.

Navzdory retrainingu nemůže být stoprocentní správnost používání terminologie u strojových překladů garantována. I při použití větších, kvalitativně lepších tréninkových korpusů dochází 510 při překladech stále ještě k různým chybám v oblasti významu, slovních druhů, flexe, syntaxe, pravopisu atd. To je způsobeno tím, že jazyk je flexibilní systém a ne všechny kombinace, respektive účely použití termínů, jsou zahrnuty v tréninkovém materiálu.

Předpokladem pro přizpůsobení překladačů je dostupnost dostatečného množství tréninkových dat. K tomu jsou vhodné v první řadě paralelní texty a překladové paměti. Méně vhodná jsou 515 terminologická data ve formě glosářů a terminologických databank. Je to jednak proto, že terminologická data jsou sice vícejazyčná, ale většinou bez kontextu, jednak z toho důvodu, že v mnoha systémech je terminologie vložena do překladu prostřednictvím hledání a nahrazení a bez jazykového a kontextuálního přizpůsobení (viz kapitola 11.4). Terminologická data jsou ale dál nezbytná pro automatizované ověření terminologie v překladech.

520 Přizpůsobení systémů strojového překladu se pojí s vysokými náklady na čas a finance, které by se neměly podceňovat. Platí to především pro analýzu, sběr, generování a zpracování dat, jakožto i pro následnou evaluaci strojových překladů, které byly vyhotoveny pomocí přizpůsobeného překladového modelu. V porovnání s tím je samotný trénink modelů strojového překladu mnohem méně náročný, neboť obsluha komerčních systémů strojového překladu se stává uživatelsky 525 přívětivější.

Ze zmíněných důvodů se doporučuje zkontrolovat před zahájením přizpůsobování systému strojového překladu projektový záměr (*business case*). Pokud je k dispozici, mělo by se k přizpůsobení modelu strojového překladu přistupovat jako k projektu, čemuž by mělo odpovídat

530 i plánování a rozpočet. Zkušenosti z jednotlivých projektů by měly být přehledně dokumentovány, aby bylo možné učit se z pokroku modelů strojového překladu a etablovat osvědčené postupy.

A konečně je třeba upozornit na to, že poskytovatelé systémů strojového překladu čas od času aktualizují své generické překladové modely. V těchto případech se doporučuje provést retraining, aby se mohla uplatnit zlepšení nových generických modelů. U některých poskytovatelů jsou odpovídající upozornění zobrazena v rozhraní správy účtu. Trénink a customizace modelů
535 strojového překladu se tím v oblasti překladatelství stávají novou, nepřetržitou výzvou.

11.6 Bibliografie

- Arntz, R. / Picht, H. / Schmitz, K.-D. (2014): Einführung in die Terminologearbeit. Hildesheim: Georg Olms Verlag.
- Arntz, R. (1999): Terminologie in der Terminologie. In: Snell-Hornby et. al (Hrsg.): Handbuch
540 Translation, S. 81. Tübingen: Stauffenburg-Verlag.
- Chen, B. / Kuhn, R. / Foster, G. / Cherry, C. / Huang, F. (2016): Bilingual methods for adaptive training data selection for machine translation. In: The Twelfth Conference of The Association for Machine Translation in the Americas, S. 93–106.
- Chenhui Chu / Rui Wang (2018): A Survey of Domain Adaptation for Neural Machine Translation.
545 In: Proceedings of the 27th International Conference on Computational Linguistics, S. 1304–1319. <https://www.aclweb.org/anthology/C18-1111.pdf>, naposledy načteno: 12. 1. 2020.
- Deutscher Terminologie Tag e.V. (Hrsg.) (2014): Terminologearbeit – Best practices 2.0
- Dougal, D. K. (2018): Improving the Quality of Neural Machine Translation Using Terminology Injection, S. 21–27. <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=8025&context=etd>,
550 naposledy načteno: 14. 1. 2020.
- Durrani, N. / Sajjad, H. / Joty, S. / Abdelali, A. / Vogel, S. (2015): Using joint models for domain adaptation in statistical machine translation. In Proceedings of MT Summit XV, S. 117–130.
- Etchegoyhen, T. / Fernández Torné, A. / Azpeitia, A. / Martínez Garcia, E. / Matamala, A. (2018): Evaluating Domain Adaptation for Machine Translation Across Scenarios. LREC.
555 <https://www.aclweb.org/anthology/L18-1002.pdf>, naposledy načteno: 9. 1. 2020.
- Fei-Fei Li (2018): Empowering businesses and developers to do more with AI. Google Blog. <https://blog.google/products/google-cloud/empowering-businesses-and-developers-to-do-more-with-ai/>, naposledy načteno: 13. 1. 2020.

- Freitag, M. Al-Onaizan, Y. (2016): Fast Domain Adaptation for Neural Machine Translation.
560 <https://arxiv.org/abs/1612.06897>, naposledy načteno: 9. 1. 2020.
- ISO 18587 (2017): Translation services – Post-editing of machine translation output – Requirements. <https://iso.org/standard/62970.html>, naposledy načteno: 12. 1. 2020.
- Koehn, P. (2010): Statistical Machine Translation. Cambridge, Cambridge University Press.
- Microsoft (2018): Customize Microsoft Translator’s Neural Machine Translation to translate just
565 the way you want: Custom Translator now in General Availability. Microsoft Translator Blog.
<https://www.microsoft.com/en-us/translator/blog/2018/12/04/customize-microsoft-translators-neural-machine-translation-to-translate-just-the-way-you-want-custom-translator-now-in-general-availability/>, naposledy načteno: 13. 1. 2020.
- Moore, Robert C. / Lewis, W. (2010): Intelligent selection of language model training data. In:
570 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Volume
2: Short Papers, S. 220–224.
- Nitze, J. / Hansen-Schirra, S. / Canfora, C. (2019): Risk management and post-editing competence.
In: The Journal of Specialised Translation, S. 239–259.
- Risku, H. (2004): Translationsmanagement. Interkulturelle Fachkommunikation im
575 Informationszeitalter. Tübingen, Gunter Narr Verlag.
- Savenkov, K. (2018): State of the domain-adaptive machine translation. Intento.
<https://www.slideshare.net/KonstantinSavenkov/state-of-the-domain-adaptive-machine-translation-by-intento-july-2018>, naposledy načteno: 13. 1. 2020.
- Sennrich, R. / Schwenk, H. / Aransa, W. (2013): A multi-domain translation model framework for
580 statistical machine translation. In: Proceedings of the 51st Annual Meeting of the Association for
Computational Linguistics, Volume 1: Long Papers, S. 832–840.
- Wang, R. / Finch, A. / Utiyama, M. / Sumita, E. (2014): Neural network based bilingual language
model growing for statistical machine translation. In: Proceedings of the 2014 Conference on
Empirical Methods in Natural Language Processing (EMNLP), S. 189–195.
- 585 Zhou, X. / Cao, H. / Zhao, T. (2015): Domain adaptation for SMT using sentence weight. In:
Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated
Big Data, S. 153–163.

Zielinski, D. / Vardaro, J. (2019): Das Terminologieproblem in der maschinellen Übersetzung.
Übersetzen und Dolmetschen 4.0 – Neue Wege im digitalen Zeitalter, Wold Conference Center
590 Bonn, 22.–24. 11. 2019.

3. Komentář

Teoretickou částí této bakalářské práce je komentář k překladu článku *Terminologie in der neuronalen maschinellen Übersetzung*. Zde provedu překladatelskou analýzu výchozího textu podle modelu Christiane Nordové (2009, s. 39–157), z hlediska české funkční stylistiky (Čechová, 2008, s. 192nn.) a zmíním i dělení funkcí textu podle Romana Jakobsona (1995, s. 78–82). V další podkapitole budu popisovat metodu překladu, typologii překladatelských problémů, zvolenou strategii při řešení problémů, na které jsem narazila, a také okomentuji překladatelské posuny, ke kterým při překladu došlo.

3.1 Překladatelská analýza originálu

Článek *Terminologie in der neuronalen maschinellen Übersetzung*, který napsali Tom Winter a Daniel Zielinski, je odborným textem s populárně-naučnými prvky. Tento příspěvek, napsaný pro publikaci *Maschinelle Übersetzung für Übersetzungsprofis*, předpokládá základní znalost překladatelského prostředí, zkušenost s prací s nástroji CAT a orientaci v terminologii modelů strojového překladu. Publikace obsahující německy a anglicky psané články byla vydána v roce 2020, informace v ní tedy nejsou příliš zastaralé. Článek doposud nebyl přeložen do českého jazyka.

3.1.1 Vnětextové faktory

3.1.1.1 Vysílatel a autor textu

Zadavatelem textu je editor publikace *Maschinelle Übersetzung für Übersetzungsprofis* Jörg Porsiel. Ten vystudoval angličtinu a francouzštinu v kombinaci s právy v ústavu pro překlad a tlumočení Ruprecht-Karlový univerzity v Heidelbergu. Od ukončení studia v roce 1992 je činný v oblasti cizojazyčné odborné komunikace, přičemž od roku 2005 zodpovídá za strojový překlad u společnosti Volkswagen AG ve Wolfsburgu. Publikace je už druhou sbírkou odborných textů na téma strojového překladu, kterou Porsiel vydal. V předmluvě, kterou sbírku odborných článků otevírá, vyzdvihuje rozmanitost oboru překladatelství, respektive strojového překladu. Publikace dále obsahuje příspěvky mezinárodních expertů na poli výzkumu a vývoje, ale také překladatelů, poradců a profesorů, kteří se zabývají tématy, jako je řízený jazyk, neuronový strojový překlad, preeditace a posteditace, ale také například setkání se strojovým překladem v praxi z pozice překladatele na volné noze nebo z pozice pracovníka překladatelské agentury.

Autoři příspěvku *Terminologie in der neuronalen maschinellen Übersetzung*, Tom Winter a Daniel Zielinski, se zabývají jazyky a strojovým překladem jak na akademické, tak na profesní úrovni.

Tom Winter získal titul Master of Arts z oboru Terminologie a jazykové technologie a pracuje v jazykové správě společnosti Deutsche Bahn AG. Kromě toho se zabývá různými scénáři integrace strojového překladu, jeho hodnocením, optimalizací a rozvojem udržitelných strategií práce s jazykovými daty. Předsedá Německé terminologické asociaci (Deutscher Terminologie-Tag) a je členem v Radě německojazyčné terminologie (Fachkonferenz Sprache und Recht, 2022). Daniel Zielinski působí jako vědecký pracovník na Univerzitě Johannese Gutenberga v Mohuči, kde se soustředí na jazykové a překladové technologie v rámci kurzů obecné a aplikované lingvistiky (Universität Mainz, 2022).

3.1.1.2 Adresát textu

Jak napovídá název publikace, v níž článek vyšel, je i tento text určen primárně profesionálním překladatelům. Relevantní může být nejen pro studenty překladatelství, ať už bakalářských či magisterských programů, ale samozřejmě i pro již činné překladatele. Zkušenost s posteditací strojového překladu či s prací na překladech odborných textů je jistě výhodou, není však nezbytným předpokladem pro porozumění textu. Autoři nicméně předpokládají znalost anglického a částečně také francouzského a italského jazyka, jakožto i obeznámenost s překladači, které neuronový strojový překlad využívají (Google, Microsoft, DeepL).

Příjemce textu nemusí rozumět procesům na pozadí práce se systémy neuronového strojového překladu, měl by však mít zkušenost s nástroji CAT, protože na terminologii a příklady z tohoto prostředí je často odkazováno. Především je ale text určen zájemcům o pochopení praktické stránky problematiky práce s terminologií v systémech neuronového strojového překladu, a díky jeho ucelenosti ho může přijmout i publikum, které se překladatelstvím profesně nezabývá.

3.1.1.3 Intence

Cílem textu je informovat čtenáře o možnostech optimalizace výstupu komerčních systémů strojového překladu ve vztahu k terminologii. Autoři se zaměřují nejprve na význam terminologie v překladu, popisují různé typické chyby v terminologii, které vznikají při strojovém překladu, a jejich závažnost. Hlavní poselství příspěvku však tvoří kapitola 11.4, v níž je pozornost věnována přizpůsobování překladových modelů tak, aby bylo ve strojovém překladu optimalizováno používání korektní terminologie.

3.1.1.4 Médium

Článek *Terminologie in der neuronalen maschinellen Übersetzung* je součástí publikace *Maschinelle Übersetzung für Übersetzungsprofis* vydané v roce 2020. Sborník, v němž jsou otištěny

příspěvky v angličtině nebo v němčině, je k dispozici v tištěné a po udělení souhlasu autorů příspěvků i v elektronické podobě. Publikaci vydalo nakladatelství BDÜ Weiterbildungs- und Fachverlagsgesellschaft mbH se sídlem v Berlíně. Toto nakladatelství, založené v roce 2003, spadá pod Bundesverband Dolmetscher und Übersetzer (Spolková asociace tlumočnicků a překladatelů), a vydávané publikace často pochází z řad členů této asociace, nebo jsou těmto členům určeny. Kromě nakladatelské činnosti se BDÜ Weiterbildungs- und Fachverlagsgesellschaft mbH zabývá pořádáním seminářů a webinářů na téma překladatelství a tlumočnictví, čímž aktivně doplňuje nabídku regionálních překladatelských a tlumočnických spolků.

Vzhledem ke svému specifickému zaměření není kniha běžně dostupná v kamenných knihkupectvích, na internetu je však možné zakoupit ji jak u nakladatele, tak u větších distributorů (Amazon, Google Books). Dále je možné vypůjčit si knihu v univerzitních knihovnách (Heidelberg, Lipsko), neboť jak již bylo uvedeno výše, *Maschinelle Übersetzung für Übersetzungsprofis* je relevantní také pro studenty translatologických oborů.

Překlad článku *Terminologie in der neuronalen maschinellen Übersetzung* by mohl být vydán jak v tištěné, tak v elektronické podobě, a to buď v rámci překladu celé publikace, nebo samostatně. Oslovila jsem *Časopis ToP* Jednoty tlumočnicků a překladatelů a článek vyjde v letním čísle v červnu 2022.

3.1.1.5 Místo

Text originálu vznikl ve Spolkové republice Německo. Autoři textu, působící v Mohuči (Zielinski) a Bonnu (Winter), mohli na práci spolupracovat prostřednictvím on-line komunikace, možná se setkali na univerzitní půdě, ve velkých metropolích, jako je Kolín nad Rýnem nebo Frankfurt nad Mohanem, ale nejspíše šlo o kombinaci těchto možností.

Místem vzniku překladu je Česká republika. Originální text je vázaný na jazykovou terminologii (německý a francouzský jazyk, přičemž volbu tohoto jazykového páru lze připsat hospodářské provázanosti mezi Německem a Francií) i na realie (některé zkratky), a bude proto potřeba text lokalizovat a doplnit o české překlady, případně o jazykové či kulturní ekvivalenty, nebo alespoň vysvětlivky.

3.1.1.6 Čas

Původní text vznikl v druhé polovině roku 2019, dokončen byl na začátku roku 2020. To lze usuzovat na základě údajů o posledním načtení webových verzí překladačů, které převážně pocházejí z ledna 2020. Překlad byl po čtyřměsíční práci dokončen v dubnu 2022. Časový rozestup

mezi originálem a překladem tvoří dva roky. Za tu dobu prošly překladače (Google, DeepL, Microsoft) vývojem a vylepšeními. Vzhledem k demonstrativnímu charakteru chybných výstupů použitých překladačů nebudou tyto údaje aktualizovány.

3.1.1.7 Motiv

Motiv vydání publikace *Maschinelle Übersetzung für Übersetzungsprofis* je zřejmý – editor Jörg Porsiel chtěl tímto sborníkem navázat na úspěch svého prvního sborníku, *Maschinelle Übersetzung: Grundlagen für den professionellen Einsatz*, který byl vydán roku 2017 a dočkal se velice pozitivních recenzí.

Samotný článek *Terminologie in der neuronalen maschinellen Übersetzung* je motivován skutečností, že překladatelé se dnes při výkonu svého povolání setkávají s posteditací odborných textů, pro jejichž překlad byl využit systém strojového překladu. Pro překladatele je tedy relevantní získat přehled o tom, k jakým chybám nejčastěji dochází a jak je možné trénovat systémy tak, aby se překlad zlepšil a práce překladatele se tím usnadnila.

Následující dvě kapitoly nespadají do modelu Christiane Nordové.

3.1.1.8 Funkční styl

Překládaný článek vyšel v publikaci určené pro profesionální překladatele, dal by se tedy kategorizovat jako prakticky odborný text. Podle Čechové spadají takové texty do funkčního stylu odborného, jehož cílem je podat co nejpřesnější, nejjasnější a nejúplnější informace o daném tématu za účelem poučení adresáta textu. Autor textu by měl navíc disponovat odbornými znalostmi. Prakticky odborný text předává poznatky, které se pojí s praktickou činností v daném oboru, na rozdíl od vědeckého textu, který podává výsledky vědeckého výzkumu a je mnohem více teoretický.

Článek je jednoznačně textem prakticky odborným, protože jeho stěžejní část tvoří seznámení s praktickým procesem tréninku, retrainingu či customizace systémů strojového překladu. Oba autoři se strojovým překladem profesně zabývají, a adresát tak může pravdivosti poskytnutých informací důvěřovat. Ačkoliv je text odborný a nevyhnutelně tedy využívá terminologii a předpokládá jistý stupeň obeznámenosti s tématem, zvolený jazyk výkladu (jakožto i zvolené příklady v ukázkách chybných zpracování terminologie překladačem) je dostatečně čitelný – je tedy vhodný například pro studenty, kteří se s terminologií již alespoň okrajově setkali. Nelze nicméně hovořit o tom, že by text spadal do kategorie popularizačního odborného stylu.

Německý originál se navíc vyznačuje věcností, přesností a nominálním charakterem textu, což jsou typické znaky tzv. *vědeckého jazyka*. Text neobsahuje subjektivní hodnocení, humor či expresivní výrazy (*hovorový jazyk*) a není patrný ani žádný individuální styl autora (*literární jazyk*).

3.1.1.9 Funkce textu podle Jakobsona

Článek má podle Jakobsonova modelu (1995, s. 78–82) funkci referenční, což je patrné z toho, že se nám autoři snaží předat poznatky o problematice trénování systémů neuronového strojového překladu, které získali na základě studia a praxe v oboru. Text má také funkci konativní, neboť jeho důležitou částí je právě kapitola o trénování a customizaci překladových modelů, přičemž podle autorů by bylo vhodné takové postupy používat v praxi pro optimalizaci překladu terminologie strojovými překladači. Dalo by se říct, že autoři vyjadřují svůj vztah k tématu například výběrem ukázek z prostředí, v němž pracují (Deutsche Bahn AG), a tím by se textu dala přisoudit také funkce emotivní.

3.1.2 Vnitrotextové faktory

3.1.2.1 Téma a obsah

Jak vyplývá z názvu článku, jeho tématem je otázka tréninku neuronových systémů strojového překladu tak, aby se zlepšila kvalita výstupů a zpřesnilo správné používání terminologie při strojovém překládání odborných textů.

V úvodu se autoři vyjadřují k stále narůstající důležitosti strojových překladů ve všech oblastech byznysu. Velkou roli zde hraje globalizace a potřeba oslovit zákazníka po celém světě. Dále jsou v úvodu vysvětleny základní pojmy, rozlišeny druhy systémů strojového překladu a nastíněny možnosti, které nabízí využívání neuronového strojového překladu a které u statistických systémů chyběly. Neuronový strojový překlad přinesl sice zvýšení kvality překladů, nadále však zůstává potřeba posteditace. Na poli překladu odborných textů, kterým se článek věnuje, je role posteditora stále klíčová, protože chyby v terminologii jsou, samozřejmě v závislosti na své závažnosti, nepřijatelné. Úvodní slovo je zakončeno výčtem nejdůležitějších systémů strojového překladu, které jsou následně používány pro demonstraci příkladů v článku samotném, a krátkým představením obsahu článku.

První kapitola se věnuje představení problematiky terminologie v překladu. Terminologie hraje v odborných textech klíčovou roli a je nezbytná k porozumění výchozímu textu. Překladatel ale může narazit na nejednoznačnost či nekonzistentní používání terminologie, což porozumění značně ztěžuje, nemluvě o tom, že překladatel sice disponuje určitými odbornými znalostmi, případně má

způsoby, jak se se slovní zásobou z dané odborné oblasti seznámit, ale často je v případě pochybností či nejasností nutné oslovit odborníka. Jakmile překladatel porozumí výchozímu textu, může začít s hledáním a nasazováním ekvivalentů v cílovém jazyce. Ne vždy existuje přesně odpovídající pojmenování, a proto se překladatel často musí uchýlit k jiným metodám, jako je popis jinými slovy, použití přejímky či vytvoření kalku. Závěrem první kapitoly představují autoři generické modely strojového překladu, respektive problematiku překládání terminologie takovými modely. Zdůrazňují, že pro generický model strojového překladu je důležitý bezprostřední kontext slova určeného k překladu, na základě kterého je zvolen co nejvhodnější překlad. Tím však zákonitě dochází k chybám, protože nezřídka je ke správnému překladu termínu potřeba znát širší kontext.

V druhé kapitole autoři plynule navazují na výskyt chyb ve strojových překladech odborných textů. Terminologické chyby mohou být málo závažné – překladač například použije jiný, méně preferovaný pravopis, nezvolí vhodné synonymum nebo přejde z odborné řeči do hovorové. Častým zdrojem chyb je také složité názvosloví, což je aktuální případ právě v němčině. Čím více má kompozitum součástí, tím se umocňuje množství variant, které je systém strojového překladu schopen vygenerovat. Ne všechny jsou samozřejmě správné, což je demonstrováno na experimentu překladu vět obsahujících termín *Gleisfreimeldeanlage*. Kapitola se dále věnuje konkrétním druhům chyb – významovým chybám (ukázáno na příkladech homografů a akronymů), vynechávání částí pojmů, které může vést k sémantickému posunu, chybám ve flexi a inkonzistenci překladu. Zmíněno je také vyhodnocování závažnosti chyb.

Třetí kapitola, věnující se tréninku a customizaci překladačů, představuje nejdůležitější část článku. Po úvodním vysvětlení termínů (trénink, customizace, retraining, testovací dataset) následuje nastínění reálného plánování tréninkového a customizačního projektu: v první fázi je třeba projekt naplánovat a připravit data, tj. vybrat vhodný tréninkový materiál a vhodného poskytovatele strojového překladu, ve druhé fázi jde o zpracování a generování dat. Při zpracování dat se zohledňují aspekty, jako je sémantická správnost, konzistence, segmentace, formátování a anonymizace, případně pseudonymizace osobních dat. Generování dat probíhá na základě rozpoznání elementů v textu, jejich nahrazení definovanými entitami, definování hodnoty zástupného znaku a generování nových textů, kde jsou entity obsazeny různými hodnotami. Třetí fází projektu je samotný trénink, evaluace a případný retraining systému strojového překladu. Jakmile je dosaženo požadované kvality překladu, přizpůsobený překladač může být nasazen a projekt ukončen. Součástí kapitoly je také prezentace výsledků tréninku překladových modelů a možnosti optimalizace.

V závěru článku prezentují autoři jeho shrnutí, které doplňují upozorněními či náměty na další optimalizaci a vylepšení kvality strojového překladu nejen v oblasti terminologie.

3.1.2.2 Presupozice

Tento odborný článek předpokládá od svých adresátů poměrně velké množství znalostí z oblasti jazykovědy – termíny jako *Kompositum*, *Benennungsbestandteil*, *Synonym*, *Semantik*, *Flexion*, *Entlehnung*, ale i *Fachsprache* či *Allgemeinsprache* by měl čtenář již znát, protože v článku jsou používány bez obsáhlejšího vysvětlení. Také s terminologií z oblasti strojového překladu je žádoucí být obeznámen – *maschinelle Übersetzung*, *MÜ-Engine*, *Light* a *Full Post-Editing*, *Übersetzungsmodelle*, *Training*, *Retraining* jsou jen některými z konceptů, které jsou sice v článku doplněny vysvětlivkou či příkladem, ale pokud se s nimi čtenář dříve nesetkal, článek pro něj může být velice abstraktní a předané informace může vnímat neúplně.

Autoři demonstrují chybné strojové překlady na příkladech mezi angličtinou, francouzštinou, italštinou a němčinou. Ve většině případů jsou konkrétní chyby vysvětleny blíže pod uvedeným příkladem, přesto je pro lepší pochopení, jak a proč chyba vznikla v první řadě, vhodná základní znalost anglického, francouzského a/nebo italského jazyka. Adresát překladu článku sem může zařadit také německý jazyk.

Vyloženě na kulturu vázané presupozice se v článku nevyskytují. Bylo by možné zařadit sem zmínění francouzského akronymu *ADC* (a německého ekvivalentu *Tf*) označujícího strojvedoucího, nicméně jde o ukázkou typické chyby překladače při zacházení s akronymy, a na pochopení dané ukázky nemá znalost těchto pojmů žádný vliv.

3.1.2.3 Kompozice a členění textu

Sbírka *Maschinelle Übersetzung für Übersetzungsprofis* má 384 stran a skládá se z 22 článků, které jsou rozčleněny do celků *Gegenwart und Zukunft*, *Recht*, *Technik*, *Qualitäts- und Risikomanagement*, *Terminologiemanagement*, *kontrollierte Sprache*, *Anwendung* a *Post-Editing*. Článek *Terminologie in der neuronalen maschinellen Übersetzung* je jedenáctou kapitolou sbírky, náleží celku *Terminologiemanagement, kontrollierte Sprache, Anwendung* a má 23 stran. Je rozčleněný na úvod (11.1), kapitolu *Terminologie beim Übersetzen* (11.2), *Terminologiefehler in maschinellen Übersetzungen* (11.3), *Training und Customization von MÜ-Engines* (11.4) a nakonec závěrečné shrnutí (11.5) a bibliografické údaje (11.6). Kapitoly 11.2, 11.3 a 11.4 se dále dělí do podkapitol.

3.1.2.4 Nonverbální prvky

Originální text je psán obyčejným bezpatkovým písmem. Struktura je dobře patrná díky několika velikostním (i významovým) úrovním nadpisů a členění do odstavců. Autoři často používají v textu kurzívu k vyznačení cizojazyčných termínů či příkladů.

Důležitou součástí článku jsou tabulky, které usnadňují orientaci v jednotlivých příkladech chybného použití terminologie, snímek obrazovky webové verze překladače DeepL a graf znázorňující poměr terminologických chyb v tomto překladači a v překladačích jiných poskytovatelů.

3.1.2.5 Lexikální rovina

Záměrem autorů je předat znalosti a objasnit důležitost a princip tréninku neuronových systémů strojového překladu, a to prakticky odbornou formou. Základní pilíř textu tedy tvoří termíny z oblasti jazykovědy, translatologie a strojového učení.

Dále se autoři německého originálu uchylují k nominálnímu stylu, který je pro odborné texty v němčině typický. Nápadné je kromě velkého zastoupení substantiv také tvoření kompozit, přičemž se nejedná vždy nutně o terminologii. I to je znakem odborného charakteru textu. Při překladu bude nutné zvolit vhodné překladačské postupy, neboť ne všechna kompozita je možné přeložit jako termín, ale třeba je rozvolnit jiným způsobem.

Anglicismy jsou nedílnou součástí nejen tohoto textu, ale nezanedbatelného procenta odborných textů obecně, neboť anglické termíny jsou uzuální a často nemají v německém jazyce ekvivalent. Zatímco v němčině ekvivalent často neexistuje z důvodu, že není potřebný (anglický pojem zůstává i po přejímce srozumitelný a zapadá do textu, např. není třeba řešit flexi, se kterou se v češtině pojí řada problémů, jako je určení rodu, přiřazení správné koncovky apod.), v českém jazyce bude třeba nalézt vhodné ekvivalenty nebo termín opsat tak, aby text působil přirozeně.

V textu najdeme také řadu verbonominálních vazeb (*Funktionsverbgefüge*), pro které v češtině ne vždy nalezneme odpovídající vazbu, a proto bude nutné přeložit je jiným způsobem.

3.1.2.6 Syntaktická rovina

Z hlediska syntaktického je text rozmanitý. Nepřekvapí nás hojně používání souvětí, kdy kromě souvětí souřadných narážíme samozřejmě i na ta podřadná. Častým druhem vedlejších vět jsou věty přívlastkové a předmětné. V textu se objevují také různé infinitivní konstrukce.

Neobvyklý není ani výskyt několikanásobných větných členů.

V kapitole 11.4.1 (*Trénink a customizace jako projekt*) se navíc setkáváme se seznamem činností imperativního charakteru, kdy jsou slovesa ve větách ponechána v infinitivu. Také zde bude při překladu nutné zvolit vhodný překladatelský postup.

Médium ani intence výchozího a cílového textu se nemění. To znamená, že se při překladu budu snažit zachovávat členění vět, pokud to nebude bránit srozumitelnosti přeloženého textu.

3.1.2.7 Suprasegmentální prvky

Kromě složitějších souvětí, která jsou obohacena středníky, dvojtečkami, závorkami, pomlčkami či prostými čárkami, patří mezi další hojně používané suprasegmentální prvky v textu kurzíva a uvozovky. Do uvozovek autoři dávají cizojazyčné příklady, kurzíva je zvolena v případě anglických termínů nebo cizích slov.

Do kategorie suprasegmentálních prvků patří také odrážkový seznam, kterých najdeme v kapitole 11.4.1 hned několik.

3.2 Překladatelská metoda a postupy

Překladatelskou koncepci jsem zvolila na základě díla Jiřího Levého (2012, s. 82) a vybraných kapitol z publikace *Překládání a čeština* (Skoumalová, Povejšil, 1994). Levý rozlišuje v překladu dvě normy, a sice normu překladatelské věrnosti a volnosti. Překladatelskou metodu věrnou (doslovnou) volí při pracovním postupu překladatelé, jejichž hlavním cílem je přesně reprodukovat výchozí text. Naproti tomu se pomocí překladatelské metody volné (adaptační) snaží překladatel zachovat krásu a estetičnost díla a navodit pocit blízkosti čtenáři, aby ten měl pocit, že čte originál a že k němu promlouvá autor původního textu.

Mým cílem při překladu je zachovat věcný obsah původního textu, avšak použít přitom takové prostředky, které odpovídají české stylistické normě. Ráda bych se pohybovala mezi metodou věrnou a volnou tak, abych zachovala všechny sdělené informace, ale zároveň aby text působil i na českého čtenáře věrohodně a dobře se četl, tj. aby informace byly podány srozumitelně.

V knize *Translating as a Purposeful Activity: Functionalist Approaches Explained* rozlišuje Nordová mezi překladem dokumentárním a instrumentálním. Cílem dokumentárního překladu je doslova dokumentovat komunikaci mezi vysílatelem a příjemcem výchozího textu, instrumentální překlad se už soustředí na převod sdělení vysílatele výchozího textu k příjemci cílového textu.

V případě článku *Terminologie in der neuronalen maschinellen Übersetzung* by český čtenář neměl vnímat, že čte původně cizojazyčný text, jedná se tedy o překlad instrumentální. Z povahy textu (jazykové příklady, jména autorů, reference na společnost Deutsche Bahn AG, ...) lze sice usuzovat, že se jedná o překlad, avšak jde spíše o detaily, nikoli o dominantní prvky textu, které by věrohodný překlad znemožňovaly.

Podle Levého se nabízejí tři překladatelské postupy: překlad pojmový, substituce (náhrada za domácí analog) a transkripce (přepis), přičemž tyto postupy se volí v závislosti na poměru jedinečnosti a obecnosti ve výchozím textu. Jiní teoretici překladu, např. Knittlová (1995, s. 9) hovoří dále o transkripci (přepisu), která je adaptovaná na úzus cílového jazyka, transpozici (nutných gramatických změnách, které způsobuje odlišný jazykový systém), modulaci (změně hlediska), kalku (doslovném překladu), adaptaci (situace popsaná v originále je v překladu nahrazená jinou, adekvátní situací), amplifikaci (rozšíření textu), explicitaci (přidání vysvětlivky či doplňující informace), vynechání a kompenzaci.

Levý dále popisuje proces překladu takto: Nejprve je důležité pochopit předlohu, pak přichází její interpretace a přestylování. Na obtíže může překladatel narazit v každé fázi. Vybrané konkrétní problémy, které jsem řešila při překladu článku, budou předmětem následující kapitoly. K porozumění originálnímu textu mi dopomohly mimo jiné publikace *Čeština a strojový překlad* (Bojar, 2012) a *Technologie ve službách překladatele* (Pošta, 2017).

Asi nejdůležitějším kritériem pro funkčnost překladu je stanovení, za jakým účelem vzniká. Pokud máme přesně stanoveny, kdo bude adresátem nebo v jakém médiu se překlad objeví, můžeme zvolit vhodnou strategii při řešení překladatelských problémů. Mým cílem bylo přeložit text tohoto odborného článku tak, aby mohl vyjít v podobně odborně zaměřeném médiu. Adresát textu zůstává nezměněn, mění se pouze předpoklad, jaký jazyk je jeho mateřským/pracovním.

3.3 Typologie a řešení překladatelských problémů

3.3.1 Lexikální rovina

3.3.1.1 Odborná terminologie

Odborný překlad se dá považovat za zvláštní druh zpracování informace a překládání odborných textů má výrazný analyticko-syntetický charakter. Informace by se měly předávat dostatečně explicitně. Termíny samotné, které hrají v odborných textech centrální roli, jsou vysoce autonomními jednotkami, jsou přesně vymezené a nezávislé na svém kontextu i etymologickém významu (Straková 1994, s. 90)

Hned u názvu článku jsem narazila na problém při překladu adjektiva *neuronal*. Intuitivním ekvivalentem by sice bylo slovo *neuronální* (případně *neurální*, utvořené od anglického *neural*), oficiálním termínem používaným v českých odborných textech je však *neuronový*. Vycházela jsem například z terminologie, kterou používá Bojar (2012).

Training, Retraining, Customisation

Autoři používají tyto tři termíny převzaté z angličtiny. Mou původní snahou bylo nalézt vhodné české ekvivalenty (např. *výcvik*, *opětovný výcvik*, *přizpůsobení*), ty se však v českém kontextu nepoužívají. Zatímco *training* disponuje počestělým ekvivalentem *trénink*, pojem *retraining* se v češtině používá nezměněný. *Customisation* se uzuálně překládá jako *customizace*. Od těchto termínů se odvozují také různá přídavná jména, která jsem se snažila konzistentně používat – *tréninková data*, *customizační projekt* atd.

Übersetzungsmodelle, Übersetzungstools

Na lehký problém jsem narazila při překladu kompozit obsahujících část *Übersetzung*-. Rozhodovala jsem se mezi slovem *překladový* a *překladatelský*. Slovtvorný rozdíl mezi těmito variantami spočívá v tom, od čeho je přídavné jméno odvozeno (Šlosar, 1995, s. 165nn). Je to osoba překladatele, nebo překlad/překládání samotné? Zatímco v některých případech jsem si byla jistá (*Übersetzungsdienste* – *překladatelské služby*), jinde jsem musela konzultovat problém s odborníky (*Übersetzungsmodelle* – *překladové modely*, *Übersetzungstools* – *překladové nástroje*). Modely ani nástroje nejsou odvozeny od osoby překladatele, nýbrž se vztahují k systému strojového překladu.

Light / Full Post-Editing

Zde bylo třeba seznámit se s existující terminologií na poli posteditace (*základní / úplná*) (ISO 18587). Další rozhodnutí jsem musela učinit v případě pravopisu slova *posteditace*, které se v literatuře objevuje jak bez spojovníku, tak s ním (*post-editace*). Korpus SyD mi v tomto ohledu nepomohl, protože neobsahuje dostatečné množství dat pro porovnání. Vyhledávání přes vyhledávač Google sice naznačilo, že „posteditace“ není tak používaná jako „post-editace“ (288 nálezů pro „posteditace“ a 1170 pro „post-editace“, vyhledáváno 22. 4. 2022), ale nakonec jsem se rozhodla řídit územ v literatuře, navíc jsem přihlédla k zápisu ve výše zmíněné normě ISO 18587.

Methode der Merkmalanalyse

V originále se v kapitole 11.2.1 vyskytuje ve větě „Bei der Klärung der Bedeutung von Begriffen, die im Ausgangstext referenziert sind, bedient sich der Übersetzer der Methode der

Merkmalanalyse“ – O 213 termín *Methode der Merkmalanalyse*, který bylo obtížné přeložit. V češtině se jedná o tzv. sémantickou analýzu (Turčová, 2001), což bych bez rešerše nedokázala určit. Dále jsem musela konzultovat pojem *diskriminierende Merkmale*, který se objevuje v navazující větě. Rozhodla jsem se přeložit jej jako *diskriminační charakteristiky*, protože původní nápad (*diskriminační vlastnosti*) by nerozlišoval mezi německými slovy *Merkmal* (charakteristika, specifikum) a *Eigenschaft* (vlastnost).

Engine

Se strojovým překladem se pojí termín *engine*, doslova stroj. Takový překlad by však nebyl úplný, neboť v angličtině (potažmo v němčině) je třeba rozlišovat mezi pojmy *machine* (*Maschine*) a *engine* (*Engine*). Některé české zdroje přejímají, podobně jako němčina, termín (ten) *engine*, v případě tohoto textu, kdy termín téměř výhradně tvoří součást kompozita *MÜ-Engine*, je však lepší zvolit termín **překladač**, protože neživotná koncovka v češtině funguje dobře jako indikátor, že se jedná o stroj, který je za překlad odpovědný. (To by ani v němčině, ani v angličtině nefungovalo, neboť tam by nebylo možné rozlišit, zda *Translator / Übersetzer* je osoba či stroj.) (Šlosar, 1995, s. 113nn.)

Platzhalter

Nejistotu jsem pocítila také při překladu termínu *Platzhalter*. V němčině má slovo několik rozdílných významů, kdy po vyloučení toho obecného (*Person, die für eine andere einen Platz freihält*) bylo třeba rozhodnout, zda je termín myšlen lingvisticky (*grammatische Einheit in einem übergeordneten Teilsatz, die auf etwas im untergeordneten Teilsatz hinweist, selbst keine eigene Bedeutung hat und in vielen Fällen auch einfach weggelassen werden kann*), informaticky (*ein Zeichen oder eine Zeichenkette, die stellvertretend für etwas steht*) či matematicky (*Name für eine Leerstelle in einem logischen oder mathematischen Ausdruck*) (DWDS, 2022). Z kontextu je zřejmé, že se jedná o termín informatický, vyloučila jsem tedy české varianty *náhradní znak*, *kolonka* či *náhradní/zástupný symbol*, neboť ty nespádají do oblasti informatiky. Tam se v češtině používá pojem *žolikový znak* nebo *zástupný znak*. Protože jsem toho názoru, že by *žolikový znak* vyžadoval další vysvětlení (článek přeci jen nepředpokládá znalosti z oboru informatiky), zvolila jsem variantu *zástupný znak*, která je pro adresáta pochopitelná i bez vysvětlivky.

Sprachpaar

Ačkoliv se zvolený překlad *jazykový pár* jeví jako samozřejmá volba, předcházelo mu rozhodování mezi párem a dvojicí. Korpusový průzkum variant SyD ukazuje, že lemma „pár“ se v psaném

jazyce používá výrazně častěji (85,9 %) než lemma „dvojice“ (zadáno 22. 4. 2022). Rozhodla jsem se proto pro *jazykový pár*.

3.3.1.2 Zkratky

MÜ

Tato zkratka termínu *maschinelle Übersetzung* se ve výchozím textu vyskytuje více než třicetkrát. V němčině je velice časté zkracovat opakovaně používaná kompozita či víceslovné názvy na jejich počáteční písmena nejen v případě hovorového jazyka (*PKW, IHK*), ale samozřejmě i v odborné sféře (např. v oblasti právní vědy – *BGB, GG, AVO*). Překlad zkratky samotné problematický není, nicméně kreativní jsem musela být při překladu termínů jako *MÜ-Engine, MÜ-System* či *MÜ-Modell*. Atribut *MÜ* se v těchto kompozitech dostává v češtině do postpozice (*system strojového překladu*), v některých případech bylo možné jej vynechat (*MÜ-Engine* přeložený jako *překladač*).

3.3.1.3 Kompozita

Jak jsem již zmínila výše, kompozita představovala častý zdroj překladatelských problémů. Kompozitum je slovo, které je utvořené ze dvou nebo více slovtvorných základů způsobem zvaným kompozice (skládání) (CzechEncy, 2022). Významově je kompozitum mnohem užší než význam jednotlivých motivujících slov. Zatímco v češtině je kompozit mnohem méně než například slov odvozených, neboť čeština disponuje širokou škálou slovtvorných přípon, v němčině hraje kompozice ve slovtvorbě podstatnou roli (Grammis, 2022). Velice často narazíme v němčině na kompozitum složené ze dvou substantiv, které by v češtině muselo být vyjádřeno pomocí adjektivního nebo substantivního přívlastku.

Německá kompozita jsem do češtiny překládala převážně těmito způsoby:

a) jako substantivum a přívlastek shodný

Sprachpaar – *jazykový pár*

Ausgangstext – *výchozí text*

Zielsprache – *cílový jazyk*

Übersetzungsmodell – *překladový model*

Mehrwortbenennung – víceslovné pojmenování

b) jako substantivum a přívlastek neshodný

MÜ-System – systém strojového překladu

Qualitätsanforderungen – požadavky na kvalitu

Informationsquelle – zdroj informací

c) pomocí vedlejší věty

Složené adjektivum *domänenspezifisch* a *kundenspezifisch*, objevující se v textu jako přívlastek (*domänen- und kundenspezifische Übersetzungsmodelle*), bylo z hlediska překladu problematické. Bylo třeba význam kompozit přenést tak, aby věta nezněla strojeně, což vyžadovalo většinou opis vedlejší větou (např. *překladové modely, které jsou přizpůsobené potřebám klienta či domény*). Při volbě překladu jsem také vycházela z anglického termínu *custom models*.

Dalším příkladem, kdy jsem zvolila vedlejší větu namísto kompozita, je tato:

Die aufgeführten Beispiele zeigen die hohe Fehleranfälligkeit der generischer maschinellen Übersetzung im Bereich der Terminologie. – O 221

Uvedené příklady ukazují, jak často generický strojový překlad na poli terminologie chybí. – P 306

d) jednoslovně

allgemeinsprachlich – hovorový

fachsprachliche Benennung – odborné pojmenování

Gleisfreimeldeanlage – výhybka

3.3.1.4 Anglicismy

Při překladu anglicismů, které v německé terminologii na poli strojového překladu objevují velice často, jsem vycházela z odborné literatury na toto téma, která byla publikována v češtině (Bojar 2012). Přestože jsou v českém jazyce také tendence zůstat u anglicismů (i proto, že většina odborných textů o strojovém překladu vzniká v angličtině, a jazyk se rychlému vývoji tohoto oboru nestihá přizpůsobovat), v některých případech bylo možné nalézt vhodný ekvivalent (například

Translation Memorys – překladové paměti, Custom Translator – vlastní překladač, Training – trénink, Customisation – customizace).

Zajímavá je otázka tzv. hybridních kompozit (CzechEncy, 2022) složených z anglicismu a německého slova, která se v textu objevují poměrně často (*MÜ-Engine, MÜ-Output, Übersetzungstools, Tuning-Daten*). Při jejich překladu jsem musela brát v úvahu existující českou terminologii, která taková kompozita nevytváří. Anglicismy tedy při překladu těchto slov vymizely.

*Bei der Übersetzung von drei Sätzen mit der fachspezifischen Benennung „Gleisfreimeldeanlage“ ins Französische durch jeweils vier **MÜ-Engines** werden insgesamt zehn französische Benennungsvarianten ausgegeben... – O 217*

*Při překladu tří vět, které obsahovaly specifický termín Gleisfreimeldeanlage, do českého jazyka pomocí vždy čtyř **překladačů** se vytvořily čtyři varianty. – P 190*

*Gerade bei der professionellen Übersetzung mit **Übersetzungstools** werden meist nur diejenigen Segmente an das MÜ-System geschickt, für die kein ausreichend hoher Treffer im **Translation Memory** gefunden wurde. – O 215*

*U profesionálního překladu, při němž se využívají **překladové nástroje**, jsou do systému strojového překladu většinou zaslány pouze ty segmenty, pro které nebyl nalezen ekvivalent s dostatečně vysokým počtem výskytů v **překladové paměti**. – P 139*

Pomocí přejímky jsem mohla přeložit spojení *Domain Adaptation* (doménová adaptace).

3.3.1.5 Verbonominální vazby

V originále se vyskytuje množství verbonominálních vazeb neboli Funktionsverbgefüge (FVG). Jedná se o lexikální jednotky, které se skládají z funkčního slovesa, které samo o sobě ale nevyjadřuje přísudek věty, a ze jmenné části, zpravidla jde o substantivum v akuzativu nebo o předložkovou vazbu. Dohromady tvoří tato vazba sémantickou jednotku a funguje ve větě jako přísudek. Zatímco funkční sloveso je nositelem funkce gramatické a sémanticky je částečně vyprázdněné (může totiž stále vyjadřovat stav, změnu stavu, konání změny), hlavním nositelem významu přísudku je jmenná část (Helbig a Buscha, 2017, s. 68–70). FVG jsou často lexikalizované, ale na rozdíl od frazeologismů můžeme verbonominální vazbu rozdělit na jednotlivé součásti (Helbig a Buscha, 2017, s. 69).

V některých případech bylo vhodné přeložit FVG pomocí podobné verbonominální vazby:

Die maschinelle Übersetzung ist nicht zuletzt auch durch die mediale Berichterstattung bei vielen Unternehmen und Organisationen in den Fokus des Interesses gerückt. – O 211

Strojový překlad se stal středem zájmu mnoha společností a organizací, a to nejen kvůli mediálnímu pokrytí. – P 29

Jindy bylo nutné nalézt vhodné sloveso, kterým bylo možné FVG nahradit:

Der Aufwand für das Sammeln und Aufbereiten der Trainingsdaten sowie die Evaluierung des MÜ- Outputs steht in Abhängigkeit der Projektziele. – O 224

Náklady na sběr a zpracování tréninkových dat a na evaluaci výstupu strojového překladu se odvíjejí od cílů projektu. – P 379

Neben der Unterstützung der professionellen Übersetzung kommt sie zur Informationsübersetzung (Gisting) und -gewinnung oder zur Echtzeitübersetzung in Dialogsystemen (Chat, Chatbots, Unified Messaging) zum Einsatz. – O 211

Kromě toho, že podporuje práci profesionálních překladatelů, se používá k překládání a získávání informací (gisting) nebo k překládání v reálném čase v dialogových systémech (chaty, chatboti, jednotné zasílání zpráv). – P 33

3.3.2 Morfologická a syntaktická rovina

3.3.2.1 Pasivum

Při zohlednění žánrově-stylistického zařazení originálu (odborný text) není překvapivé, že se ve větách setkáváme s převažujícími pasivními konstrukcemi. I v česky psaných odborných textech je však relevantní pasivum použít, například v případech, „kdy skutečného původce činnosti není nutno konkretizovat, popř. není znám, není uveden“ (Minářová 2009, s. 51). V mnoha případech jsem tedy mohla pasivní konstrukci ponechat. Do češtiny jsem takové konstrukce převáděla častěji pomocí zvratného (reflexivního) pasiva, které Čechová (2008, s. 152) označuje za stylisticky neutrální.

Dabei wird zwischen Light Post-Editing und Full Post-Editing unterschieden. – O 211

Rozlišuje se základní posteditace (light post-editing) a úplná posteditace (full post-editing). – P 43

Mögliche Übersetzungen für ausgangssprachliche Benennungen werden identifiziert und dann miteinander verglichen, bis eine geeignete Übersetzung gefunden wurde. – O 214

Možné varianty překladu pro pojmenování ve výchozím jazyce se nejprve identifikují, poté se mezi sebou navzájem porovnávají, dokud není nalezeno vhodné překladové řešení. – P 117

V některých případech jsem se rozhodla použít také opisné pasivum, i když to Čechová považuje za lehce knižní (Čechová 2008, s. 152).

Welche der Benennungen am Ende für die Übersetzung gewählt wird, hängt vom Kontext ab. – O 215

Jaké pojmenování bude nakonec zvoleno jako vhodný překlad, záleží na kontextu. – P 137

Pokud při překladu vyznívalo zvrtné pasivum neobratně a opisné až příliš toporně, rozhodla jsem se, byl-li znám konatel děje (agens), převést rod z trpného do činného. Jedná se o syntaktickou transpozici.

Fehlerwahrscheinlichkeit und -häufigkeit werden durch die folgenden Faktoren begünstigt. – O 216

Následující faktory ovlivňují pravděpodobnost a četnost výskytu chyb. – P 168

3.3.2.2 Neosobní konstrukce (man, es)

Za neosobní konstrukci považujeme neurčitý personický subjekt, který označuje „nepojmenovanou osobu, skupinu osob nebo člověka obecně, jemuž (jíž) se predikátem věty přisuzuje nějaký děj nebo stav“ (Štícha, 2015, s. 548). Cílem používání takových konstrukcí je zdůraznění děje či činnosti komunikátu na úkor konatele takového děje (konatel je tímto v podstatě anonymizován a odsunut do pozadí). V němčině je neosobní konstrukce vytvořena nejčastěji pomocí neurčitého podmětu man. Do češtiny jsem jej převáděla nejčastěji pomocí první osoby množného čísla:

Verändert man den Kontext im Ausgangssatz und ersetzt green durch blue, wird das Adjektiv jedoch richtig flektiert. – O 220

Změníme-li však kontext ve výchozí větě a místo green použijeme blue, výsledný překlad bude v tomto případě už vyskloňovaný správně. – P 271

Další častou konstrukcí v němčině je zájmeno es, které se může objevovat jako zájmeno nahrazující neutrum, jako korelát, tj. zájmeno odkazující na větu podmětnou či předmětnou či jako formální podmět nebo předmět. Dále jej můžeme postavit na začátek věty před sloveso (verbum finitum),

čímž umožníme podmětu či příslovečnému určení vystupovat ve větě jako réma (Helbig a Buscha, 2017, s. 239–244). Podle toho, jakou funkci *es* v originální větě mělo, jsem se snažila zvolit vhodné překladatelské řešení.

Bevor das Trainings- oder Customization-Projekt abgeschlossen wird, empfiehlt es sich, die gewonnenen Erfahrungen in Bezug auf Sprachpaar, Datenmenge, Evaluierungsergebnis, Zeit, Kosten etc. Zu dokumentieren... – O 228

Než jsou tréninkový a customizační projekt ukončeny, doporučuje se zdokumentovat získané zkušenosti z hlediska jazykového páru, množství dat, výsledku evaluace, času, nákladů atd. – P 483

Während das Retraining mit 20.000 Segmenten bei einigen Anbietern in weniger als einer Stunde abgeschlossen ist, dauert es bei anderen mehrere Stunden. – O 224

Zatímco retraining s 20 000 segmenty může být u některých poskytovatelů hotový za méně než hodinu, u jiných to trvá více hodin. – P 376

Es wurde hier die Benennung lattina verwendet, die eher im Sinne von Dose meist im Kontext von Lebensmitteln verwendet wird. – O 220

Místo toho byl v překladu použit pojem lattina, který se používá spíše ve smyslu plechovka či konzerva v souvislosti s potravinami. – P 263

3.3.2.3 Rozdělování a spojování vět

Pokud to bylo možné (např. pokud bych tím nenarušila srozumitelnost textu), zachovávala jsem při překladu členění vět podle originálního textu. Na některých místech jsem ale cítila potřebu souvětí rozdělit do dvou vět.

Dabei kommen in erster Linie datenbezogene Verfahren zum Einsatz, die es dem Nutzer ermöglichen, domänen- oder kundenspezifische Daten hochzuladen, die dann für die Erstellung oder Anpassung eines MÜ-Modells verwendet werden. – O 212

V první linii jsou využívány procesy vztahující se na data, které umožňují uživateli nahrávat data přizpůsobená doméně či klientovi. Tato data jsou pak použita ke zhotovení nebo přizpůsobení modelu strojového překladu. – P 71

3.3.2.4 Nominálnost textu

Nominální styl je typický především pro odborné texty, protože jeho funkcí je v první řadě ekonomičnost vyjádření (Eroms, 2014, s. 115–131). Na několika místech se objevují například jmenné řetězce sloužící k syntaktické kondenzaci. Zatímco v němčině, která má obecně o něco nominálnější charakter než čeština, nominální vazby nebrání plynulosti textu, v češtině bylo třeba tyto vazby rozvolnit, například vyjádřit pomocí vedlejších vět.

Zusätzlich empfiehlt sich die Überprüfung der Übersetzungskandidaten anhand einer Lister definierter Benennungsbewertungskriterien. – O 214

Při ní se doporučuje podrobit každého kandidáta na překlad kontrole pomocí seznamu definovaných hodnotících kritérií pro jednotlivé pojmy. – P 119

Neben der Unterstützung der professionellen Übersetzung kommt sie zur Informationsübersetzung (Gisting) und -gewinnung oder zur Echtzeitübersetzung in Dialogsystemen (Chat, Chatbots, Unified Messaging) zum Einsatz. – O 211

Kromě toho, že podporuje práci profesionálních překladatelů, se používá k překládání a získávání informací (gisting) nebo k překládání v reálném čase v dialogových systémech (chaty, chatboti, jednotné zasílání zpráv). – P 33

3.3.3 Stylistická rovina

Odborné vyjadřování v německých textech se v porovnání s diskursem v anglických či českých textech liší například v počtu nominalizací, složenin, pasiv, neosobních konstrukcí, dále v textové rovině v mnohosti odboček, které porušují lineárnost textu nebo v četnosti modálních výrazů, které vytvářejí složitou hypotetickou modalitu textu. Německý odborný text tak klade na adresáta poměrně vysoké nároky, protože předpokládá, že čtenář je zodpovědný za dešifraci textu, na rozdíl od angličtiny, kde je za jasnost výkladu zodpovědný autor (Čmejrková, Daneš, Světlá 1999, s. 27)

Při překladu jsem se musela soustředit na to, aby věty, obzvlášť delší souvětí, neztrácely význam a především aby zůstaly srozumitelné. Jak jsem demonstrovala výše, bylo potřeba rozdělit některá souvětí či rozhodnout, co s nominálním charakterem některých odstavců.

V případě následujícího výčtu specifik, která by měla být při kontrole kandidátů na překlad ověřena, jsem se i přes menší toleranci češtiny k nominalitě rozhodla použít substantivní výčet tak, jak je

v originále. Drobnou stylistickou změnou je také přidání reference na objekt v předchozí větě (*kritéria*).

Dazu zählen Eindeutigkeit, Gesetzes- und Normenkonformität, Transparenz, Einheitlichkeit, Gebräuchlichkeit, Zielgruppengerechtigkeit, Kürze, Sprechbarkeit, sprachlich-logische Korrektheit, Konnotationsfreiheit, Ableitbarkeit und Internationalität. – O 214

K těmto kritériím patří jednoznačnost, soulad se zákony/normami, transparence, jednotnost, míra používání, vhodnost v závislosti na cílové skupině, délka, výslovnost, jazykově-logická správnost, konotační svoboda, odvoditelnost a mezinárodnost. – P 120

3.3.4 Pragmatická rovina

Na pragmatické rovině bylo především nutné vyřešit, jak v překladu naložit s vícejazyčnými tabulkami, na nichž jsou demonstrovány chybné výstupy překladačů. Autoři volili příklady v jazykových párech angličtina – němčina, francouzština – němčina a italština – němčina. Pokud bych se ale rozhodla kompletně nahradit němčinu češtinou v jednotlivých příkladech, narazila bych na několik problémů: některé konkrétní chyby v překladu terminologie by nebylo možné demonstrovat, protože se v novém jazykovém páru (např. francouzština – čeština) nevyskytují. Musela bych tedy přeložený text výrazně měnit či opatřovat vysvětlivkami. Některé odstavce by přestaly být relevantní a musela bych je odstranit. Navíc by při vytváření nových jazykových párů mohly vzniknout chyby nové, které by bylo třeba tematizovat. Rozhodla jsem se tedy ponechat jazykové páry tak, jak jsou uvedeny v originále, a vždy jsem pouze doplnila vlastní, doslovný překlad výstupu překladače do češtiny, čímž jsem zachovala názornost daných příkladů a zároveň jsem nebyla nucena měnit text článku jako takový. Přidání sloupce s doslovnými překlady do češtiny jsem doplnila poznámkou překladatele.

V případě výčtu francouzských variant pro slovo *Gleisfreimeldeanlage*, respektive *Weiche*, které byly vygenerovány strojovými překladači, jsem se rozhodla nahradit je variantami českými, neboť cílem daného odstavce bylo pouze demonstrovat důležitost kontextu pro generování překladů a množství variant, které je překladač schopen vygenerovat; o konkrétní jazykový pár nebo gramatický jev, který se s daným jazykovým párem pojí, v tomto případě nešlo. Tuto změnu jsem však opatřila poznámkou překladatele, kde jsem rovněž zmínila zajímavý jev, a sice obrácený poměr množství vygenerovaných variant v porovnání mezi němčinou a francouzštinou a mezi němčinou a češtinou.

3.4 Typologie posunů

Překladačským posunům není možné se při překladu vyhnout. Dochází k nim kvůli jazykovým a kulturním rozdílům mezi jazykem výchozího a cílového textu. Podle Popoviče (1975, s. 122) se objevují jak na makrostylistické úrovni textu (tj. tematická výstavba textu), tak na té mikrostylistické (tj. jazyková výstavba). Dále rozlišuje mezi posuny nevyhnutelnými (konstitučními), které jsou dané objektivní skutečností (např. gramatické rozdíly mezi jazyky) a individuálními, které vychází z překladatelova výrazového idiolektu (Popovič, 1975, s. 132).

3.4.1 Intelektualizace

Hlavní snahou překladatele by mělo podle Levého (2012, s. 132) být, aby domácím čtenáři podal dílo takovou formou, která pro něj bude srozumitelná. Překladatel text tedy nejen překládá, ale i vykládá – zlogičťuje, dokresluje, intelektualizuje. Levý uvádí jako hlavní tři typy intelektualizace formální vyjadřování syntaktických vztahů, vykládání nedořečeného a zlogičťování textu.

V této větě jsem v obou případech slovo *Übersetzung* doplnila podle kontextu, aby předaná informace byla jasnější:

Mögliche Übersetzungen für ausgangssprachliche Benennungen werden identifiziert und dann miteinander verglichen, bis eine geeignete Übersetzung gefunden wurde. – O 214

Možné varianty překladu pro pojmenování ve výchozím jazyce se nejprve identifikují, poté se mezi sebou navzájem porovnávají, dokud není nalezeno vhodné překladové řešení. – P 117

3.4.2 Vynechávky

Stává se, že z jazykových, literárních nebo kulturních důvodů dojde k tomu, že se v překladu něco ztratí nebo v něm naopak něco přibude. Podle Popoviče (1975, s. 118) je ovšem důležité, aby zůstal zachován invariant (jádro) výpovědi.

V níže uvedené větě jsem se rozhodla vypustit anglickou frázi „*good enough*“. Jedním z důvodů bylo, že v češtině by se mi s anglickou frází špatně pracovalo. Podstatnějším důvodem však byla jistá synonymita slov *ausreichend* a „*good enough*“, kde jediný rozdíl spočívá v jisté jízlivosti při použití angličtiny. Uvozovky mohou také naznačovat, že se ze strany autorů jedná o ironii, vypuštěním této informace však text neztrácí na odbornosti a srozumitelnosti.

Während maschinelle Übersetzungen für Informationszwecke oft schon ausreichend oder „good enough“ sind, müssen sie für Veröffentlichungszwecke meist noch post-editiert werden. – O 211

*Přestože už jsou strojové překlady pro účel předání informace **dostatečné**, musejí být před případnou publikací ještě posteditovány. – P 43*

3.4.3 Substituce

V následující větě jsem nahradila německý příklad variability pravopisu (slova se spojovníkem či bez něj), neboť v češtině se s tímto problémem nesetkáváme tak často. Vhodným ekvivalentem je však kolísavé používání písmen s a z ve slovech cizího původu.

*Je mehr natürliche Synonyme ein Begriff in einer Sprache hat oder je mehr Schreibweisen zu einer Benennung existieren (z. **B. Bindestrich-Varianten**), desto unwahrscheinlicher ist es, dass das Datenmaterial, das für das generische Übersetzungsmodell verwendet wurde, konsistent ist. – O 216*

*Čím víc má nějaký pojem v dané řeči přirozených synonym nebo čím víc existuje variant, jak pojem korektně zapsat (**například psaní s či z ve slovech cizího původu – filosofie/filozofie, president/prezident**), tím méně pravděpodobné je, že data, která byla pro generický překladový model použita, budou konzistentní. – P 171*

4. Závěr

Cílem této bakalářské práce bylo vyhotovit funkčně ekvivalentní překlad článku *Terminologie in der neuronalen maschinellen Übersetzung*, provést překladatelskou analýzu výchozího textu, představit překladatelskou metodu, nastínit vybrané překladatelské problémy, které při překladu vyvstaly, a popsat překladatelské posuny, ke kterým došlo.

Jako nejobtížnější se pro překlad ukázala rovina lexikální, neboť zvolený odborný text obsahoval mnoho termínů, pro které nebylo vždy snadné nalézt vhodný český ekvivalent. Tato potíž byla způsobena i tím, že na poli neuronového strojového překladu publikují vědci většinou v angličtině, případně i do českých textů pronikají anglické termíny, které však nejsou oficiální a není stylisticky vhodné je používat ve všech případech. Musela jsem tedy svá řešení konzultovat s odborníky, abych si byla jistá jejich správností.

S odborností výchozího textu se pojily i další problémy – jeho nominální charakter a převaha pasiva se nedaly do češtiny převést mechanicky. Leckdy bylo zapotřebí být kreativní a zajistit, aby se překlad četl plynule a čtenář neměl pocit strojenosti; přitom jsem však musela dbát na zachování odborného tónu.

Zpočátku se jako velká výzva jevíly vícejazyčné tabulky, kde bylo nutné rozhodnout, jestli vypustím německou část a nahradím ji češtinou. Nakonec jsem se rozhodla ponechat příklady z originálu beze změny a pouze jsem je doplnila o sloupec s doslovným českým překladem.

Překlad vybraného textu byl pro mě velkou výzvou, ale jelikož se o téma strojového překladu a strojového učení obecně zajímám a ráda bych se mu věnovala i v budoucnu, považovala jsem za více než přínosné překládat právě tento text. Bavilo mě navíc číst paralelní texty a rozšiřovat své znalosti v tomto oboru. Při psaní komentáře jsem využívala vyučujícími doporučené překladatelské příručky a gramatiky a poznatky, které jsem nabyla za dobu studia na Univerzitě v Lipsku a Univerzitě Karlově v Praze.

Použitá literatura

- BOJAR, Ondřej, 2012. *Čeština a strojový překlad: Strojový překlad našincům, našinci strojovému překladu*. Praha: Ústav formální a aplikované lingvistiky. ISBN 978-80-904571-4-0.
- BOZDĚCHOVÁ, Ivana; WAGNER Roland, 2017. HYBRIDNÍ KOMPOZITUM. *CzechEncy: Nový encyklopedický slovník češtiny* [online]. [cit. 15. 4. 2022]. Dostupné z: <https://www.czechency.org/slovník/HYBRIDNÍ%20KOMPOZITUM>
- BOZDĚCHOVÁ, Ivana; WAGNER Roland, 2017. KOMPOZITUM. *CzechEncy: Nový encyklopedický slovník češtiny* [online]. [cit. 15. 4. 2022]. Dostupné z: <https://www.czechency.org/slovník/KOMPOZITUM>
- ČECHOVÁ, Marie; KRČMOVÁ, Marie; MINÁŘOVÁ, Eva, 2008. *Současná stylistika*. Praha: Nakladatelství Lidové noviny. ISBN 978-80-7106-961-4.
- Český národní korpus – SYN2015. 2015. Ústav Českého národního korpusu FF UK, Praha. [cit. 22. 4. 2022]. Dostupné z: <https://www.korpus.cz/>
- ČMEJRKOVÁ, Světlá; DANĚŠ, František; SVĚTLÁ, Jindra, 1999. *Jak napsat odborný text*. Praha: LEDA. ISBN 80-85927-69-1.
- ČSN ISO 18587, 2021. *Překladačské služby – Posteditace výstupů strojového překladu – Požadavky*. Praha: Český normalizační institut.
- DWDS – Digitales Wörterbuch der deutschen Sprache [online]. *PLATZHALTER*. [cit. 22. 4. 2022]. Berlin-Brandenburgischen Akademie der Wissenschaften. Dostupné z: <https://www.dwds.de/>.
- EROMS, Hans-Werner, 2014. *Stil und Stilistik: Eine Einführung*. 2. přepracované a rozšířené vydání. Berlin: Erich Smidt Verlag.
- Fachkonferenz Sprache und Recht 2022. [online]. *Autor: Tom Winter*. [cit. 30. 5. 2022]. Dostupné z: <https://www.fsr2022.de/>
- Grammatisches Informationssystem grammis. [online]. *KOMPOSITION*. [cit. 14. 4. 2022]. Leibniz-Institut für Deutsche Sprache. Dostupné z: <https://grammis.ids-mannheim.de/systematische-grammatik/585>.
- HELBIG, Gerhard; BUSCHA Joachim, 2017. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. München: Klett-Langenscheidt.
- JAKOBSON, Roman, 1995. *Poetická funkce*. Jinočany: H & H.
- KNITTLOVÁ, Dagmar, 1995. *Teorie překladu*. Olomouc: Univerzita Palackého v Olomouci. ISBN 80-7067-459-8.

- KUFNEROVÁ, Zlata; POLÁČKOVÁ, Milena; POVEJŠIL, Jaromír. SKOUMALOVÁ, Zdena; STRAKOVÁ, Vlasta, 1994. *Překládání a čeština*. Praha: H & H. ISBN 80-85787-14-8.
- LEVÝ, Jiří, 2012. *Umění překladu*. 4., upr. vyd. Praha: Apostrof. ISBN 978-80-87561-15-7.
- NORD, Christiane, 2009. *Textanalyse und Übersetzen: Theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. 4. vydání. Tübingen: Julius Groos Verlag Brigitte Narr. ISBN 978-3-87276-868-1.
- NORD, Christiane, 2018. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. 2. vydání. Milton Park, Abingdon, Oxon; New York: Routledge. ISBN 978-1-138-57336-9.
- POPOVIČ, Anton, 1975. *Teória umeleckého prekladu*. Bratislava: Tatran.
- POŠTA, Miroslav, 2017. *Technologie ve službách překladatele: CAT, strojový překlad, korpusy*. Praha: Nakladatelství Miroslav Pošta – Apostrof. ISBN 978-80-87561-37-9.
- ŠLOSAR, Dušan, 1995. *Slovotvorba*. In: *Příruční mluvnice češtiny*. 2. vydání. Praha: Nakladatelství Lidové noviny. ISBN 9788071061342.
- ŠTÍCHA, František, 2015. *Česko-německá srovnávací gramatika*. Praha: Academia. ISBN 978-80-200-2378-0.
- TURČOVÁ, I., 2001. *Sémantická analýza anglické odborné terminologie v oblasti výchova a aktivity v přírodě*. Diplomová práce. Praha: UK FTSV.
- Universität Mainz. Autor: Daniel Zielinski. [online]. [cit. 30. 5. 2022]. Dostupné z: <https://astt.fb06.uni-mainz.de/author/dzielins/>.
- WINTER, Tom. ZIELINSKI, Daniel, 2020. *Terminologie in der neuronalen maschinellen Übersetzung*. In: *Maschinelle Übersetzung für Übersetzungsprofis*, ed. Jörg Porsiel, BDÜ, s. 210-233. ISBN 9783946702092.

Přílohy

Text originálu

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich zur Anfertigung der vorliegenden Arbeit keine anderen als die angegebenen Quellen und Hilfsmittel und keine nicht genannte fremde Hilfe in Anspruch genommen habe. Alle Zitate aus der Primär- und Sekundärliteratur, einschließlich Zitaten aus Internetquellen und von elektronischen Datenträgern, habe ich an der jeweils betreffenden Textstelle markiert (direkte Zitate in Anführungszeichen) und einschließlich der zugehörigen Seitenangabe belegt. Die verwendete Primär- und Sekundärliteratur habe ich mit vollständigen bibliographischen Daten in der Bibliographie am Ende der Arbeit alphabetisch gelistet. Bei Internetquellen habe ich mein Zugriffsdatum mit angegeben.

Die Arbeit wurde nicht für andere Prüfungszwecke verwendet.

Mir ist bekannt, dass eine unwahrheitsgemäße Erklärung als Täuschung im Sinne von § 13 (3) in Verbindung mit § 21 (1) der Prüfungsordnung für den Bachelorstudiengang Interkulturelle Kommunikation und Translation Tschechisch-Deutsch an der Universität Leipzig vom 8. Oktober 2014 gilt.

Ort: Praha

Datum: 09.05.2022

Unterschrift: Rencova'