



Univerzita Karlova

Filozofická fakulta

Ústav translatologie

Diplomová práce

Kategorizace úprav strojového překladu při posteditaci:

jazyková kombinace angličtina – čeština

Classification of Edit Categories in Machine Translation

Post-Editing: English-Czech Language Combination

Bc. Klára Kopecká

Překladařství a tlumočnictví: čeština – angličtina

Prohlášení:

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

V Praze dne 2. srpna 2021

Bc. Klára Kopecká

Poděkování:

Děkuji společnosti Memsource, zejména Mgr. Aleši Tamchynovi, Ph. D., za poskytnutí dat pro výzkum v empirické části a za odborné konzultace v oblasti strojového překladu. Vedoucímu práce PhDr. Bc. Tomáši Svobodovi, Ph.D. děkuji za cenné rady a připomínky v průběhu psaní této práce, ale také za jeho činnost v oblasti překladových technologií na ÚTRL FFUK. V neposlední řadě děkuji své rodině a přátelům za asistenci s technickou úpravou datasetu, technickou a jazykovou úpravou této práce a za celkovou péči a podporu, kterou mě v průběhu psaní této práce obklopovali.

ABSTRAKT

Práce překladatele dnes neznamena jen převádění obsahu z jednoho jazyka do druhého za pomoci vlastních znalostí dvou jazyků a obsahu. Dnes překládání často znamená práci s nápovědami a návrhy řešení z různých zdrojů, včetně strojového překladu. Posteditace strojového překladu (MTPE) je stále žádanější forma překladu, proto by si studenti na translatologických oborech měli tuto dovednost osvojit, než vstoupí na trh. Pro práci se strojovým překladem je potřeba nejen znát základní principy fungování překladačů a překladových technologií, ale také dokázat vyhodnotit relevanci jednotlivých lingvistických úprav vzhledem k zadání a účelu překladu. Cílem této práce je analyzovat lingvistické úpravy provedené při posteditaci strojového překladu ve směru z angličtiny do češtiny. Identifikované úpravy jsou kategorizovány a výsledkem je výčet lingvistických kategorií úprav při posteditaci strojového překladu ve směru z anglického do českého jazyka.

KLÍČOVÁ SLOVA

MTPE, PEMT, posteditace, strojový překlad, kategorizace, úpravy

ABSTRACT

Today, a translation job does not only mean transferring content from one language to another by utilizing one's own knowledge of two languages and subject matter expertise. Today, translation often means working with suggestions from various resources, including machine translation. The popularity of machine translation post-editing (MTPE) as a form of translation is growing. That is why this skill should be acquired by translation students prior to entering the market. In order to work with machine translation efficiently, not only knowledge of the basic principles of how machine translation engines and translation technology work is needed, but also being able to assess the relevance of each linguistic edit with regards to the assigned instructions and purpose of the translated text. The aim of this master's thesis is to analyze linguistic edits carried out during an MTPE job from English to Czech. Identified edits are then classified, resulting in a list of linguistic edit categories in English-to-Czech MTPE.

KEY WORDS

MTPE, PEMT, post-editing, machine translation, classification, edits

OBSAH

SEZNAM ZKRATEK.....	6
ÚVOD.....	7
1 TEORETICKÁ ČÁST.....	10
1.1 Překladové technologie.....	10
1.1.1. Strojový překlad.....	11
1.1.2. Nástroje počítačem podporovaného překladu.....	22
1.2 Posteditace.....	24
1.2.1 Definice a základní pojmy.....	24
1.2.2 Kdo jsou profesionální posteditoři.....	26
1.3 Výuka překladových technologií na univerzitách a vysokých školách.....	28
1.3.1. Výuka nástrojů počítačem podporovaného překladu.....	29
1.3.2 Zkušenosti s kurzy zaměřenými na posteditaci.....	39
1.4 Shrnutí teoretické části.....	42
2 PRAKTICKÁ ČÁST.....	44
2.1 Analyzovaný dataset.....	44
2.1.1 Vznik datasetu.....	44
2.1.3 Zdrojové texty.....	44
2.1.3 Překladače.....	46
2.1.4 Posteditoři a revizoři.....	47
2.1.5 Anotovaný výsek dat.....	49
2.2 Kategorizace úprav.....	51
2.2.1 Postup.....	51
2.2.2 Kategorie úprav ve fázi posteditace.....	52
2.2.3 Kategorie úprav ve fázi revize.....	75
2.3 Diskuze.....	76
ZÁVĚR.....	78
BIBLIOGRAFIE.....	80
PŘÍLOHA.....	84

SEZNAM ZKRATEK

AMTA – Association for Machine Translation in the Americas
CAT – počítačem podporovaný překlad (computer-assisted / computer-aided translation)
CNL – řízený přirozený jazyk (controlled natural language)
DGT – Generální ředitelství pro překlady (Directorate-General for Translation)
EAMT – European Association for Machine Translation
EBMT – příkladový strojový překlad (example-based machine translation)
EMT – European Master's in Translation
LSP – poskytovatel jazykových služeb (language service provider)
MT – strojový překlad (machine translation)
MTPE – posteditace strojového překladu (machine translation post-editing)
NLP – zpracování přirozeného jazyka (natural language processing)
NMT – neurální strojový překlad (neural machine translation)
PEMT – posteditace strojového překladu (post-editing machine translation)
QA – kontrola kvality (quality assurance)
RBMT – pravidlový strojový překlad (rule-based machine translation)
SMT – statistický strojový překlad (statistical machine translation)
TM – překladová paměť (translation memory)
TMS – systém pro správu překladů (translation management system)
ÚFAL – Ústav formální a aplikované lingvistiky

ÚVOD

Nacházíme se v době, kdy dennodenně vzniká více psaného obsahu než kdy předtím. S postupující globalizací je také stále větší množství obsahu zadáváno k překladu a poskytovatelé jazykových služeb (LSP) jsou nuceni nacházet způsoby, jak vyhotovovat překlady rychleji a efektivněji. Jistě není překvapením, že podobně jako v jiných oblastech i s překlady začaly lidem pomáhat informační technologie. Vedle obecných technologických pomocníků uplatnitelných i mimo překladatelství, jakými jsou například internet, textové editory s funkcí automatické kontroly pravopisu nebo elektronické slovníky, se pro překladatele staly nepostradatelnou součástí práce i nástroje počítačem podporovaného překladu (nástroje CAT).

Nástroje CAT jsou systémy vyvinuté specificky za účelem zvýšení produktivity překladatele, redukci repetitivních úkonů a ke konzistenci překladu napříč texty. Na tyto podpůrné nástroje, které napomáhají překladatelům v tom, aby rychleji a efektivněji vytvořili překlad, je dnes napojován i strojový překlad (MT), jenž vyhotovuje překlad z jednoho jazyka do druhého automaticky pomocí počítače.

Mezijazykový strojový překlad se stal významnou průkopnickou technologií v oboru překladatelství již v devadesátých letech dvacátého století. Dodnes nebyl význam strojového překladu coby inovátora překladatelské práce překonán, a to i díky tomu, že od zmíněných devadesátých let dochází neustále k vývoji strojového překladu, např. k posunu od využívání statistických modelů (tedy od statistického strojového překladu (SMT)) k využívání neuronové sítě (tedy k neuronovému strojovému překladu (NMT)). Příchod NMT znamenal významné zvýšení kvality strojového překladu, a tím i četnější využití v komerční překladatelské praxi.

Technologický pokrok změnil způsob, jakým překladatelé vykonávají svou profesi. Práce překladatele (zejména neliterárních textů) se dnes již nesestává čistě z porozumění zdrojového textu a jeho převedení do cílového jazyka na základě vlastních znalostí zdrojového a cílového jazyka a znalosti tématu. Dnes se překladatel při překládání navíc soustředí ještě na práci se zdroji, které mu nabízí právě zmíněné podpůrné nástroje, ať jsou to překladové paměti, glosáře nebo právě řešení navržené strojovým překladačem. Výstup strojového překladače, který dále není upraven člověkem, má dnes v určitých kontextech své využití, ale ve většině případů musí být pro oficiální publikování textu zrevidován a upraven člověkem. Pokud překladatel vychází z překladu navrženého strojem a upravuje ho, hovoříme o posteditaci strojového překladu (MTPE či PEMT).

V posledním roce se o poskytování MTPE začal mezi poskytovateli překladových služeb projevovat velký zájem. Z průzkumu EUATC z roku 2020 vyplynulo, že 78 % dotazovaných poskytovatelů překladových služeb chce začít tuto službu poskytovat nebo ji začít poskytovat častěji než dosud.¹ Co to znamená pro překladatele? Potřebují se za účelem kvalitního poskytování MTPE učit něco jiného nebo něco navíc oproti studiu překladatelství? V současné době jsou zmapované obecné rozdíly mezi kompetencemi, které ke své profesi potřebuje překladatel a které posteditor, jako například v článku Vivety Geneové². Ze zdrojů, jejichž shrnutí bude rozpracováno v teoretické části, vyplývá, že pro připravenost překladatelů na vykonávání práce posteditora je klíčová praktická zkušenost s touto činností, tudíž věříme, že by se s posteditací měli aspirující profesionálové setkat již při studiu překladatelského oboru. Vedle obecných žádoucích kompetencí je ale potřeba i konkrétní představa o lingvistických úkonech, které při takové práci posteditor provádí. Taková studie a rozbor lingvistických kategorií, kterými se posteditoři zabývají při práci s konkrétní jazykovou kombinací angličtina – čeština, v současné době chybí.

Proto je cílem této diplomové práce zaměřit se na oblast posteditace strojového překladu ve směru z angličtiny do češtiny. Práce má dvě hlavní části – teoretickou a empirickou. Úkolem teoretické části bude zmapovat současně dostupné relevantní zdroje jak z tuzemských, tak i zahraničních akademických pracovišť, a na základě poznatků charakterizovat současný stav této problematiky.

Ve druhé, empirické části budeme analyzovat dataset složený ze souboru textů v originále (v anglickém jazyce), jejich strojových překladů do češtiny v několika verzích z několika překladačů, dále jejich posteditací od skupiny profesionálních překladatelů a následných revizí další skupinou profesionálních překladatelů. Tento dataset vznikl primárně na zadání společnosti Memsource s cílem zkoumat souvislost mezi kvalitou strojového překladu a rychlostí posteditace. Řešitelka této práce je zaměstnaná ve společnosti Memsource a spolu se zástupcem společnosti se domluvili, že Memsource poskytne tento dataset řešitelce i pro další zkoumání. Posteditory a revizory zajistila společnost České překlady s.r.o. S kontaktní osobou na této straně je řešitelka taktéž ve spojení.

¹ MARKING, Marion. Survey Reveals Impact of Covid-19 on Europe's Language Industry. *Slator*. [online] [cit. 11.05.2021] Dostupné z: <https://slator.com/industry-news/survey-reveals-impact-of-covid-19-on-europes-language-industry/>

² GENE, Viveta. Changing Roles: From Translation to Post-Editing. *Intertranslations.com* [online] [cit. 02.04.2021]. Dostupné z: <https://www.intertranslations.com/downloads/post-editing-whitepaper/>

Předmětem zkoumání této práce budou úpravy provedené jak posteditory, tak i změny provedené v následné fázi revize. Budeme se snažit tyto úpravy klasifikovat a kategorizovat, a dále pozorovat, zda případné kategorie úprav budou mít odlišné zastoupení ve fázi posteditace a revize. Tímto budeme chtít zjistit, zda je možné definovat kategorie chyb, které posteditoři nejvíce opravují a které naopak mají tendenci přehlížet. Důležité pro tento výzkum bude i zadání, které posteditoři dostali. Téma úrovně MTPE bude v této práci okomentováno. Rovněž se práce bude věnovat i problematice zkušenosti a odbornosti posteditorů. Každý z nich totiž v rámci prvotního výzkumu také vyplnil krátký dotazník o svých předchozích profesních zkušenostech. Odpovědi z těchto dotazníků bude mít řešitelka také k dispozici a v práci je okomentuje.

Jelikož se patrně momentálně na žádné české vysoké škole nebo univerzitě v rámci programů zaměřených na překladatelství angličtiny nevyučuje samostatný předmět věnovaný posteditaci strojového překladu (toto tvrzení bude v práci dále podloženo v kapitole 1.3.1.3), mohla by zjištění vyplývající z této práce pomoci při zavádění takového kurzu nebo zařazení systematické přípravy na práci posteditora jako součást stávající výuky.

1 TEORETICKÁ ČÁST

V teoretické části bude úkolem podat přehled současného stavu v oblasti překladových technologií, dále zmapovat oblast posteditace, včetně role profesionálů, kteří tuto činnost vykonávají. Dále poskytneme výťah ze zkušeností některých vysokých škol a univerzit, kde již proběhly kurzy zaměřené na výuku posteditace strojového překladu. Na základě těchto poznatků zhodnotíme důležitost praktické zkušenosti studentů s posteditací. Dále navrheme způsob zkoumání dostupného datasetu textů pro účely zmapování reálné práce profesionálních překladatelů z pohledu lingvistických úkonů. Samotné zkoumání bude předmětem empirické části.

1.1 Překladové technologie

V této podkapitole budou shrnuty základní koncepty a pojmy z oblasti překladových technologií. Pro tuto kapitolu bude zapotřebí úžeji vymezit, jak přesně pojem překladové technologie chápeme, protože svým způsobem by se mezi překladové technologie daly zařadit i obecné nástroje, které používáme téměř každý den, např. online slovníky nebo nástroje automatické kontroly pravopisu či gramatiky. Zároveň pokud se chceme zaměřit na nástroje, které napomáhají při vzniku mezijazykového překladu profesionálním překladatelům, hovořili bychom např. o systémech pro správu překladů (TMS), překladové paměti nebo terminologické databázi (Qian, 2013). Podobně situaci vnímá i Svoboda (2012, s. 71):

„V širším slova smyslu lze za nástroje počítačem podporovaného překladu považovat všechny počítačové programy a funkce využívané pro práci na překladu. [...] Nástroje počítačem podporovaného překladu v užším/vlastním slova smyslu, tedy takzvané nástroje CAT (Computer Assisted Translation či Computer Aided Translation), se vyznačují především využíváním technologie překladové paměti, tedy databázového systému, kam se ukládají vždy dvojice segmentů textu – výchozího a jemu odpovídajícího překladového segmentu.“

Profesor Chan v publikovaném rozhovoru (Qian, 2013) nabízí ještě jeden způsob, jak chápat překladové technologie, a to jako počítačové technologie pro překlad, mezi které spadá počítačový (tedy strojový) překlad a počítačem podporovaný překlad. V této kapitole se budeme zabývat právě těmito dvěma hlavními kategoriemi překladových technologií, tedy: a) strojovým překladem, b) nástroji počítačem podporovaného překladu (nástroje CAT) ve smyslu počítačových programů vyvinutých specificky za účelem zefektivnění práce překladatelů pomocí překladových pamětí a terminologických databází.

1.1.1. Strojový překlad

1.1.1.1 Historický vývoj strojového překladu na pozadí vývoje translatologie

Vývoj translatologie jako akademické disciplíny bývá popisován pomocí hlavních milníků, které zásadně změnily vnímání oboru. Tyto milníky se v translatologii nazývají obraty a jsou zevrubně popsány v publikaci *The Turns of Translation Studies: New paradigms or shifting viewpoints?* (Snell-Hornby, 2006). Prvním zásadním obratem byl v období po druhé světové válce obrat pragmatický, během kterého se disciplína translatologie vymanila z obecného vnímání coby součást lingvistiky nebo součást srovnávací literatury. Podle Snell-Hornbyové tomuto obratu velkým dílem přispěl velký zájem vědců právě o strojový překlad, který byl ale zpočátku vnímán velmi naivně a nerealisticky:

„Vědci v té době ve vši vážnosti věřili, že podrobná syntaktická pravidla, rozsáhlá slovní zásoba a vysokorychlostní zpracovávání nejednoznačností budou tvořit dostatečný základ pro dokonalý strojový překlad, známý jako FAHQT (fully automatic high quality translation)“.³

Zcela pochopitelně se z řad lingvistů zvedla proti takovým předčasným závěrům vlna odporu, ať už se jednalo o lingvistu Bar-Hillela, který již v roce 1960 oponoval tím, že proces překladu nelze zúžit pouze na mechanickou aktivitu, ale že je zapotřebí i lidských znalostí (Bar-Hillel, 1960), a nebo o rezolutní zprávu z roku 1966 vydanou poradním výborem ALPAC (Automatic Language Processing Advisory Committee), podle kterého by se v zájmu vědy v rozvoji těchto snah sice mělo pokračovat, ale že v dohledné době nelze spoléhat na to, že se povedlo zásadním způsobem usnadnit praktickou překladatelskou činnost (Snell-Hornby, 2006, s. 35). Tento prvotní neúspěch strojového překladu je mimo jiné vnímán jako příčina následného vzniku nástrojů počítačem podporovaného překladu (Chan, 2015).

V dalších letech jako by se translatologie vyvíjela paralelně na poli lingvistů a na poli matematiků a počítačových lingvistů, kteří převzali snahy o vývoj překladových technologií. V translatologii jako takové hovoříme o dalším obratu v souvislosti s vývojem v 80. letech 20. století, a to o obratu kulturním. Ten je chápán jako obrat, během kterého se disciplína, v předchozím obratu vyčleněná z obecného pojmu lingvistika, pevně zapisuje jako samostatná vědní disciplína, a to hlavně díky publikační činnosti teoretiků překladu, kteří svými vědeckými publikacemi započali vznik nejrůznějších překladatelských konceptů a pojmů, jako jsou např. polysystém, translát nebo norma, a překladatelských škol a směrů, např. manipulační škola nebo

³ Snell-Hornby, 2006, s. 35, překlad KK.

teorie skoposu. Slovy Snell-Hornbyové (Snell-Hornby, 2006, s. 47) se v tomto období translologie z rukou „průkopníků“ dostala do rukou „mistrů“.

Nadcházejícími obraty podle dělení Snell-Hornbyové byly poté obrat empirický, zaměřený zejména na oblast tlumočení, a dále obrat globalizační, ve kterém je reflektováno vnímání nástupu technologií nejen jako obecného prostředku k procesu globalizace, ale také už velmi konkrétně jako pomocníka při překladatelské činnosti. Na tomto místě je vhodné nastínit, že dělení vývoje translologie podle milníků navržených Snell-Hornbyovou není jediné existující dělení. Jiný přístup má k dějinám translologie např. Peter Newmark, který hovoří o čtyřech etapách vývoje, a v souvislosti se začátkem nového tisíciletí hovoří o počátku obratu etického.⁴ Naproti tomu čínský translolog Chan v roce 2002 označil dobu rychlého technologického vývoje a globalizace jako obrat technologický (Qian, 2013, s. 2). Dnes, v roce 2021, můžeme s jistotou konstatovat, že od 90. let 20. století do současnosti je největším hybatelem neliterární překladatelské praxe stále překladová technologie, ať už ji definujeme jakkoliv (viz úvod kapitoly 1.1).

1.1.1.2 Vývoj technologických přístupů ke strojovému překladu

Abychom mohli shrnout vývoj technologických přístupů ke strojovému překladu alespoň pomocí základních konceptů, bude zapotřebí nejprve uvést pojmy, které strojovému překladu předcházejí. Ještě než se v oblasti počítačové lingvistiky začalo hovořit o počítačovém překladu mezi jazyky, bylo vůbec otázkou, jak naučit počítač „rozumět“ přirozenému jazyku do té míry, aby ho uměl zpracovat a následně například odhalit gramatické chyby nebo potažmo i převést texty do jiného jazyka. Právě počítačové zpracovávání přirozeného jazyka (*natural language processing*, NLP) je systém, kterým se počítačová lingvistika zabývá.

K prvním úvahám o automatickém překladu jednoho jazyka do druhého přispěly dvě zásadní události (Hajič, Panevová, Hladká, 2012), a sice úspěch matematika Alana Turinga a jeho dešifrovacího stroje, jenž během druhé světové války pomohl rozklíčovat německé příkazy nejvyššího velení, a také vznik nového vědního oboru v rámci matematiky – teorie informace zabývající se přenosem informace, kterou popsal Američan Claude Shannon (*The Mathematical Theory of Communication*, 1949).

⁴ Teoretické činnosti Petera Newmarka se podrobněji zabývá ve své diplomové práci Mgr. Ondřej Klbal (2014).

„[N]a základě Shannonovy teorie informace lze totiž problém překladu formulovat jako problém šifrování (či spíše dešifrování) – věta v ruštině se považuje za šifrovanou větu v angličtině, a cílem překladu je vlastně tedy dešifrování původní (zde anglické) věty.“ (Hajič, Panevová, Hladká, 2012)

Prvním vědcem, jenž navrhl použití počítačů k mezijazykovému překladu, byl Warren Weaver v roce 1947. Na základě memoranda „Translation“, které vydal o dva roky později, se spustily první praktické pokusy o mezijazykový počítačový překlad (Qun, Xiaojun, 2015, s.105). První výsledek těchto pokusů demonstrovali zaměstnanci společnosti IBM ve spolupráci s vědci z Georgetown University. Jednalo se tehdy o překlad 49 vět z ruštiny do angličtiny za použití systému, který obsahoval 250 slov a 6 gramatických pravidel. Následoval velký zájem komerčních subjektů, díky kterým se od té doby vývoj technologií pro strojový překlad dosud nezastavil.

Technologický přístup ke strojovému překladu prošel vývojem, který by se dal postihnout popsáním čtyř hlavních typů systémů, na kterých strojový překlad fungoval nebo funguje. Jsou to tyto přístupy: pravidlový strojový překlad (*Rule-Based Machine Translation*, RBMT), příkladový strojový překlad (*Example-Based Machine Translation*, EBMT), statistický strojový překlad (*Statistical Machine Translation*, SMT) a neurální strojový překlad (*Neural Machine Translation*, NMT).

Nyní popíšeme jednotlivé systémy. Pravidlový strojový překlad (RBMT) funguje na základě gramatických pravidel a slovníků, které do systému vložili odborníci. Kromě pravidel definovaných vývojáři existují i ta, jež jsou generovaná automaticky. Ta se pak rovněž využívají i v některých systémech EBMT a SMT, nejedná se v takových případech o pravidlový strojový překlad právě proto, že je do systému nenahrávají lidé (Qun, Xiaojun, 2015, s. 111). Vývoj pravidlového strojového překladu je velmi časově i personálně náročný z hlediska nároků na manuální nahrávání a správu dat. Zároveň není v lidských silách pokrýt všechna pravidla včetně všech výjimek tak, aby byly pokryty veškeré lingvistické fenomény a stroj tak byl schopen zpracovat autentické texty z práce. Kvalitativní nedostatky, navíc v kombinaci s kritikou například ze strany odborného výboru ALPAC (viz kapitola 1.1.1.1), vedly k tomu, že se ve své době systémy RBMT samy o sobě pro komerční překladatelskou praxi neuchytily.

Systémy EBMT překládají obsah na základě analogie. To znamená, že existuje paralelní korpus vět ve zdrojovém a cílovém jazyce, a na základě těchto příkladů potom stroj překládá podobné věty tak, jak už je „zná“ z minulosti. Nevýhodou toho systému je, že se dá zlepšovat pouze rozšiřováním paralelního korpusu. I tak ale bude mít kvalitní výsledky pouze pro obsah, který

je do dostatečné míry podobný tomu, co už je obsaženo v korpusu. Pokud chceme tímto systémem přeložit obsah, který dostupným příkladům podobný není, bude kvalita překladu velmi nízká (Qun, Xiaojun, 2015, s. 112). Dnes na principu ukládání již přeloženého obsahu fungují takzvané překladové paměti (TM), které budou blíže popsány v kapitole o nástrojích počítačem podporovaného překladu (1.1.2).

Rozpuk matematické statistiky v oblasti počítačové lingvistiky v 90. letech 20. století dal vzniknout novému technologickému přístupu ke strojovému překladu – statistickému strojovému překladu (SMT). Zde považujeme za zajímavé zmínit úvahu nad osudem spojení statistiky s lingvistikou. Statistika totiž začala do oboru lingvistiky pronikat již na přelomu 19. a 20. století. Otázkami statistických údajů a frekvence jazykových jevů se poté zabývali i lingvisté Pražského lingvistického kroužku – Vilém Mathesius a Bohumil Trnka. V 60. letech se ve svých publikacích zabývá matematickými modely a statistikou i přední český teoretik překladu Jiří Levý, např. v souvislosti s popisem verše a rýmovými možnostmi různých jazyků (Levý, 2012, s. 203-339; Levý, 1971, s. 264-288) nebo za účelem porovnávání poměrné důležitosti motivů uplatňovaných při překladu (Levý, 1971, s. 88). V zahraničí se ale už v 50. letech 20. století pod vlivem uznávaného a významného lingvisty Noama Chomského uchýlila zahraniční lingvistika směrem formálního popisu jazyka podle jeho modelů formálních gramatik na základě pravidel (Hajič, Panevová, Hladká, 2012) a minimálně v oblasti počítačové lingvistiky se statistika přihlásila o slovo až o několik desítek let později.

„Chomsky ostatně velmi striktně odmítl možnost jakéhokoli statistického pohledu na přirozený jazyk. Tento nepřátelský postoj se podařilo prolomit až na počátku devadesátých let.“ (Hajič, Panevová, Hladká, 2012).

Statistický strojový překlad je takový systém, ve kterém je překlad generován na základě statistických modelů. Proces překladu pomocí SMT se sestává z následujících fází: sestavení bilingvního korpusu a zarovnání na úrovni slov či vět, porovnání frekvence výskytů slov ze zdrojového textu oproti korpusu, vygenerování nejpravděpodobnější verze v cílovém jazyce, úprava slovosledu dle pravděpodobnosti podle monolingvního korpusu, a nakonec vygenerování výsledného překladu (Qun, Xiaojun, 2015, s. 128). Stručně shrnuto, tento systém vždy vyprodukuje překlad na základě nejvyšší statistické pravděpodobnosti. Systémy SMT se na relativně dlouhou dobu staly nejen stěžejním tématem vědeckého bádání v oblasti strojového překladu, ale také i dominantními systémy využívanými komerčně. Vědci po určitou dobu věřili, že budoucnost strojového překladu tkví v hybridním využití dvou nebo více systémů (např. SMT + RBMT).

To se však změnilo kolem roku 2016 s příchodem nového pojetí strojového překladu, a sice NMT, systému založeném na neuronových sítích a strojovém učení (*machine learning*).

„Jedná se stále o pravděpodobnostní statistické modely, avšak díky aplikaci tzv. spojených reprezentací slov pomocí vektorů a tzv. embeddingů dokážou jazykové jednotky reprezentovat včetně jejich syntaktických, sémantických a morfologických vlastností a díky neuronovým sítím a hlubokému učení (*deep learning*) jsou schopny učit se z dat výrazně rychleji a efektivněji než SMT.“ (Hrách, 2020, s. 16)

Příchod NMT znamenal zlepšení kvality strojového překladu do takové míry, že upoutal i pozornost širší veřejnosti, např. prostřednictvím populárně naučných článků jako je tento: „Google skokově vylepšil překlad. Umělá inteligence se zakousla do češtiny“ (Kasík, 2017).

NMT je momentálně nejužívanější technologií pro strojový překlad psaného obsahu (Microsoft, „What is machine translation?“).

1.1.1.3 Strojové překladače

V předchozí kapitole jsme shrnuli základní typy technologií, které se používaly nebo používají k vývoji strojového překladu. V této kapitole se zaměříme na výběr konkrétních poskytovatelů strojového překladu pro komerční využití.

Konkrétní programy, jež zprostředkovávají výstupy MT systémů koncovým uživatelům skrz počítačové rozhraní, se nazývají strojové překladače (nebo jen překladače). Abychom uvedli alespoň několik nejrozšířenějších zástupců překladačů dostupných na současném trhu, budeme vycházet z žebříčku sestaveného nezávislým webovým portálem G2⁵, který shromažďuje recenze a uživatelská hodnocení nejrůznějších softwarů. Tento žebříček je sestavován na základě algoritmu⁶, který vyhodnocuje data o spokojenosti komunity uživatelů G2 na základě jimi napsaných recenzí v kombinaci s daty o tržní pozici daného softwaru (zde rozhodují kritéria jako rychlost růstu firmy, přítomnost na webu apod.). Ze 45 uvedených překladačů jsou v době vzniku tohoto textu na prvních 10 příčkách tyto překladače: Google Translate, Microsoft Translator, Yandex.Translate, IBM Watson Language Translator, Amazon Translate, Bing Translator, Google Cloud Translation API, Babylon, Slate Desktop a Unbabel. Češtinu podporují z těchto uvedených překladačů všechny s výjimkou překladače Unbabel.

⁵ „Best Machine Translation Software“. G2. [online] [cit. 18.6.2021] Dostupné z: <https://www.g2.com/categories/machine-translation>

⁶ „G2 Research Scoring Methodologies“. G2. [online] [cit. 18.6.2021] Dostupné z: <https://research.g2.com/g2-scoring-methodologies>

Specificky pro český trh jsou ale významnější ještě další překladače. Uvedme například překladač LINGEA od stejnojmenné české firmy. Z informací dostupných na jejich webové stránce⁷ se dozvíme, že jejich překladač je zejména kvalitní v jazykové kombinaci angličtina – čeština, a co se týče obsahu, je jejich překladač připravený konkrétně k překladu průvodců, a tak bude pro tento typ textů v této jazykové kombinaci nabízet nejkvalitnější výstup.

Dalším důležitým subjektem na českém trhu je od letošního roku i překladač DeepL německé firmy LINGUEE, který od března 2021 podporuje i češtinu, a to údajně ve vysoké kvalitě:

„Na překvapivě dobré překlady služby DeepL upozornil i Viktor Janiš, český překladatel převážně anglické beletrie: „Porovnával jsem Google Translate a DeepL. Nebe a dudy. DeepL překládá překvapivě dobře. Je to skvělý nástroj pro překlad textů administrativního charakteru, akademických studií, jednodušších novinových článků a podobně.““ (Kasík, 2021)

Hlavním výzkumným pracovištěm zabývajícím se strojovým překladem je v České republice Ústav formální a aplikované lingvistiky (ÚFAL) Matematicko-fyzikální fakulty Univerzity Karlovy. Mezi největší úspěchy tamních vědců (ve spolupráci s vědci z Oxfordské univerzity a týmu Google Brain) patří nepochybně **překladač CUBBITT** (Charles University Block-Backtranslation-Improved Transformer Translation), pro osobní nekomerční užití dostupný na <https://lindat.cz/cubbitt> pro jazyky čeština, angličtina, francouzština, němčina, polština, ruština a jako cílový jazyk navíc ještě hindština. Jedná se o překladač, který funguje na principu neuronových sítí, které byly natrénované pomocí hlubokého učení. K natrénování neuronové sítě pro pár angličtina – čeština byl použit česko-anglický paralelní korpus o 58 milionech párů vět. Kromě lidmi přeložených anglicko-českých textů byla použita i tzv. synteticky paralelní data. Syntetická paralelní data vznikla pomocí tzv. zpětného překladu (*backtranslation*), tzn. k textům dostupným pouze v cílovém jazyce byl vytvořen „zdrojový“ text strojovým překladačem. Díky tomuto postupu vzniklo mnohem větší množství trénovacích dat, než kdyby byly použity pouze autentické paralelní korpusy, kterých je pro vývoj opravdu kvalitního překladače stále nedostatečné množství (Popel *et al.*, 2020, s. 2). Inovace při trénování tohoto modelu spočívala v tom, jak byly tyto dva typy bloků prokládány. Autentické a syntetické bloky dat byly neuronové síti předkládány ve specificky vyváženém rytmu, což se ukázalo jako

⁷ „Nový překladač LINGEA“. *LINGEA*. [online] [cit.18.6.2021] Dostupné z: <https://www.lingea.cz/aktuality/novy-prekladac-lingea>

výhodné a vedlo k výraznému kvalitativnímu zlepšení oproti předchozím verzím („Překladač z Matfyzu dohání v kvalitě běžné překladače“, 2020).

Autoři s překladačem CUBBITT realizovali na Matematicko-fyzikální fakultě Univerzity Karlovy experiment, jehož výsledky publikoval i přední vědecký časopis Nature Communications (POPEL *et al.*, 2020). V experimentu hodnotitelé v zaslepeném testu hodnotili kvalitu výstupů překladače CUBBITT a profesionálních překladačů při překladu novinových článků. Celkově byly hodnoceny 53 novinové články, které byly tematicky rozřazeny do 7 kategorií: business (včetně ekonomie), kriminalita, zábava (umění, film a jeden článek o architektuře), politika, věda a technologie, sport a svět. Při evaluaci se hodnotitelé zaměřovali na dva aspekty, a sice významovou přesnost (*accuracy*) překladu a plynulost (*fluency*), tedy jak dobře působí přeložený text bez ohledu na zdroj.

„Výsledek srovnání můžeme považovat za průlomový. Automatický překladač sice nepatrně pokulhával za lidskými překladači v hodnocení plynulosti, byl ale v průměru o něco přesnější, pokud jde o obsahovou správnost překladu. Naměřený výsledek byl statisticky signifikantní,“ uvedl hlavní autor studie Mgr. Martin Popel, Ph.D. z MFF UK. („Překladač z Matfyzu dohání v kvalitě běžné překladače“, 2020)

Zde považujeme za důležité zdůraznit, že pozoruhodné výsledky překladače CUBBITT jsou podmíněné typem textů, které jsou jím překládány. Určitě není vhodné zobecněné tvrzení, že CUBBITT překládá lépe než profesionální překladačové. I v samotném článku pro Nature Communications (POPEL *et al.*, 2020, s. 2) autoři uvádějí, že ony průlomové výsledky se týkají především novinových článků, a to zejména v kategorii business nebo politika. Rozhodujícím kritériem pro toto zhodnocení byla přesnost, kterou u výsledků z překladače CUBBITT hodnotili hodnotitelé jako vyšší než u výsledků překladačů. Sám Popel ale uvádí (Houser, 2020), že u překladu beletrie by byly výstupy překladače výrazně horší (vždyť i pro lidské překladače je to náročnější formát) a u populárně naučných knih by byl podle Popela strojový překlad sice použitelný za účelem urychlení překladačského procesu, ale vyžadoval by bezpochyby více než běžnou redakci. A to právě z toho důvodu, že u takových stylistických forem nelze nahlížet pouze na kritérium přesnosti, ale i na kritérium plynulosti, ve kterém si překladač CUBBITT vedl hůře než překladačové. Nicméně stále se zlepšující kvalita překladačů je pro veřejnost zajímavá a láká k testování překladačů i na beletristickém žánru, jako například v článku Ondřeje Nováka (2021), ve kterém je porovnán původní lidský překlad úryvku z knihy *Konec civilizace* s překlady z překladačů Google Translate, DeepL a CUBBITT. Byť z tohoto porovnání nevyšel jeden jasný vítěz, je nepochybně důležité, že se podobné články tématu

strojového překladu věnují i z hlediska technologií, které za překladači stojí, a přibližují veřejnosti principy jejich fungování.

1.1.1.4 Současné trendy v oblasti strojového překladu

V předchozích kapitolách jsme popsali vývoj a současný stav technologických přístupů i současné nejvýznamnější poskytovatele strojového překladu na trhu. V této kapitole nastíníme několik konceptů, které se strojovým překladem souvisí a které se týkají dnešních problémů a výzev, o kterých se mluví v souvislosti s implementací a užíváním strojového překladu v překladatelské praxi.

Jedním z aktuálních témat je problematika **hodnocení kvality strojového překladu** (*MT quality assessment* nebo *MT quality evaluation*). Měření a vyhodnocování kvality strojového překladu jsou důležité nejen pro vědce a autory těchto systémů, aby je mohli dále zlepšovat, ale slouží také koncovým uživatelům k tomu, aby si byli schopni ze široké nabídky různých poskytovatelů kvalifikovaně vybrat, který systém zakoupit. V současné době jsou rozšířené dva hlavní typy hodnocení kvality MT – ruční/lidské a automatické (např. TER (Translation Error Rate) nebo BLEU (Bilingual Evaluation Understudy)). Ovšem každá z metod je vhodná pro hodnocení různých aspektů, a tedy měří výkon překladačů z různých pohledů. V důsledku této rozmanitosti neexistuje konsenzuální standardizovaný postup pro hodnocení kvality strojového (ale koneckonců ani lidského) překladu. „Navíc každá metoda má svá specifická úskalí a nepřesnosti, které mohou věrohodnost měření podstatně snížit“ (Bojar, 2012, s. 105).

Výhody a nevýhody obou metod shrnula ve své studii Irene Rivera-Trigueros (2021, s. 5). Podle ní mezi výhody lidského hodnocení patří fakt, že je považováno za spolehlivější (pokud hodnotitelé splňují určitá kritéria), na druhou stranu mezi nevýhody této metody patří náročnost na čas a lidské zdroje. Bojar (2012, s. 109) mezi nevýhody lidského hodnocení zařazuje subjektivitu (a z ní vyplývající vnitro- i mezihodnotitelské neshody) a principiální nereprodukovatelnost (= nemožnost otestovat zlepšování systému bez změny testovacích vět nebo hodnotitelů). Naproti tomu u automatických metod uvádí Rivera-Trigueros mezi výhody menší lidské úsilí, objektivitu a nižší cenu, zatímco mezi nevýhody řadí fakt, že automatické metody porovnávají podobnost výsledků MT s referenčními překlady navzdory tomu, že jedno jediné správně řešení v naprosté většině případů neexistuje, a dále také zastarávání technologií, na jejichž principu metriky fungují.

Jako příklad automatických metod uveďme již zmíněnou nejpoužívanější metriku, BLEU, která v podstatě určuje škálu shod strojové překladu s referenčním překladem. Nehodnotí tedy, jak je

přenesen význam sdělení, ale pouze porovnává, do jaké míry se strojový překlad shoduje s překladem referenčním (který ale zároveň není jediný správný). Metrika BLEU tedy není schopna vyhodnotit srozumitelnost a gramatickou správnost textů (TAUS eLearning, „Post-Editing/Reviewing“). Jak uvádí Tamchyna (2021), metrika BLEU je sice v současnosti považována za standard mezi výzkumníky i mezi komerčními překladatelskými subjekty, protože je v mnoha případech dostačující, ale je známo několik slabin, kvůli kterým ani tato populární metrika není neochvějná, např. nemá velmi dobré výsledky na úrovni jednotlivých vět nebo má tendenci upřednostňovat plynulost před významovou přesností. Tamchyna (2021) proto doporučuje věnovat pozornost metrice chrF3, která má údajně (minimálně na větné úrovni) významně lepší výsledky než BLEU.

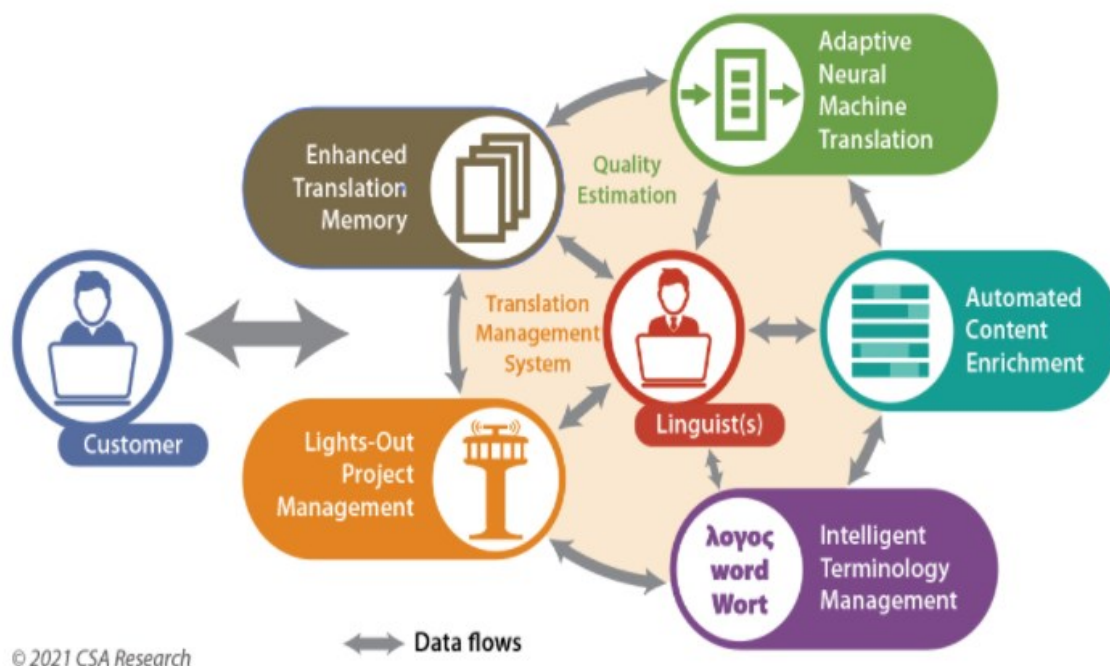
Speciální kategorií pro měření kvality MT jsou metriky, které se zaměřují spíše na zkoumání procesů posteditace (lidmi vykonaných úprav výstupu MT) než na samotné výstupy MT. Při takovém zkoumání je předmětem sledování např. kognitivní zátěž/úsilí (*cognitive load*), časová náročnost apod. Jedna z takových metrik, která vyčísluje úsilí vykonané během posteditace, se nazývá HTER (human-mediated translation error rate).

Problematiku kvality a s ní spojného výběru mezi překladači pojala velmi inovativním přístupem technologická firma Memsorce. Tamější tým zaměřený na umělou inteligenci vyvinul technologii Memsorce Translate fungující na principu algoritmu, který automaticky vybere nejvhodnější překladač (z výběru podporovaných překladačů) vzhledem k překládanému obsahu jak z hlediska jazykové kombinace, tak i z hlediska tematického zaměření (někdy také *domény*) daného textu. Kromě zvolení toho nejlépe vyhodnoceného překladače pro daný text ještě Memsorce Translate obsahuje funkcionalitu automatického hodnocení kvality tohoto překladače, MTQE (machine translation quality estimation). MTQE nevyhodnocuje kvalitu strojového překladu na základě srovnání s referenčním překladem, nýbrž pomocí algoritmu. Automaticky ohodnotí očekávanou kvalitu výstupu překladače na úrovni jednotlivých segmentů (většinou vět). Na základě tohoto ohodnocení se může zpracovatel překladu rozhodnout, zda výstupy překladače využije a provede posteditaci nebo zda raději segment přeloží rovnou sám bez využití strojového překladu. („Introducing Memsorce Translate“, 2020).

Dalším velkým tématem v oblasti strojového překladu je také dotrénování neboli tzv. **customizace překladačů**. Jedná se o proces, při kterém si jednotlivec nebo organizace může generický strojový překladač dotrénovat pomocí vlastního datasetu. Takto natrénovaný překladač potom bude vykazovat lepší výsledky pro obsah podobný datům použitým pro natrénování. Některé firmy začaly customizaci překladačů nabízet jako službu a postavily na tom nový obchodní model. Takový obchodní model může vypadat např. takto: poskytovatelé překladače natrénují customizovaný překladač pro korporátní firmu, spravují ho na svém serveru a poskytují ho firmě k neomezenému překládání na bázi předplatného, které se může pohybovat až ve výši € 200 000 za pětileté předplatné (Larsonneur, 2021, s. 260).

Dalším novodobým přístupem ke strojovému překladu, který považujeme za důležité zde zmínit, jsou tzv. **adaptivní strojové překladače**. To jsou takové překladače, které se od překladačů/posteditorů „učí“ v reálném čase (Mihalache, 2021, s. 26). S adaptivním neurálním strojovým překladem se počítá i v předpovědi toho, jak bude proces překladu vypadat do budoucna. Již v roce 2017 zavedla výzkumná organizace CSA Research termín „rozšířený překlad“ (*augmented translation*), což byla vize, ve kterém je překladač uprostřed překladačského procesu, a kolem sebe má k dispozici nejrozličnější technologii a zdroje, včetně adaptivního neurálního překladače (DePalma, 2017). S konceptem rozšířeného překladu pracuje CSA Research ve své předpovědi i na počátku roku 2021. V nové předpovědi už nemá překladač mít roli příjemce, který čistí výstupy překladače a ostatních zdrojů, ale má být naopak přímo v centru procesu trénování strojového překladače. Adaptivní neurální překladač zde pak zastává roli urychlovače celého procesu, jehož úkolem je produkovat kvalitní výstup a v reálném čase se učit od lidského překladače/posteditora. (DePalma, 2021).

Augmented Translation Ties Humans and Machines Together



Obrázek 1 – Koncept tzv. rozšířeného překladu, kde je překladatel v centru překladatelského procesu a aktivně se podílí na trénování a adaptaci podpůrných zdrojů. (DePalma, 2021)

Překlad – Nadpis: „Rozšířený překlad propojuje lidi a stroje“. Zleva: Zákazník (Customer); Zlepšená překladová paměť (Enhanced Translation Memory); Automatizovaný projektový management (Lights-Out Project Management); Adaptivní neurální strojový překlad (Adaptive Neural Machine Translation); Lingvisté – ve smyslu překladatelů, posteditorů, korektorů, atd. (Linguist(s)); Inteligentní správa terminologie (Intelligent Terminology Management); Automatické obohacování obsahu (Automated Content Enrichment); Datové toky (Data flows).

Jako poslední současný trend v oblasti strojového překladu zde zmíníme **strojový překlad na úrovni dokumentů** (*document-level MT*). Momentálně se v této oblasti dělají velké pokroky, které umožňují překlenout strojové překládání na úrovni jednotlivých vět a překládat dokument jako celek. Tím pádem by do procesu strojového překládání mohl být zahrnut kontext celého dokumentu. Kontext na úrovni dokumentu by u NMT mohl znamenat zlepšení kvality celkového výstupu, protože na úrovni věty „již není mnoho prostoru ke zlepšení, alespoň pro jazykové páry a domény bohaté na trénovací data“ (Jon, 2019, „Abstrakt“).

1.1.2 Nástroje počítačem podporovaného překladu

V této kapitole pojednáme o nástrojích podporovaného překladu ve smyslu počítačových programů vyvinutých specificky za účelem zefektivnění práce překladatelů pomocí překladových pamětí a terminologických databází. Pro kontext překladatelské profese, ať už z pohledu překladatelů, posteditorů, tak i manažerů překladatelských projektů, jsou v dnešní době takové programy nepostradatelné. Z pohledu právě vyjmenovaných rolí by se překladatelské počítačové programy daly rozdělit do dvou hlavních typů – TMS (*translation management system*) a CAT (*computer-aided* nebo *computer-assisted translation*). Byť tyto dva typy programů od sebe v praxi často nejsou oddělené (naopak nástroje CAT jsou často součástí nástrojů TMS), nabídneme v této kapitole dělení podle toho, jaké profesní role v tom či onom prostředí pracují a k jakým úkonům v nich dochází.

Nástroj TMS je v doslovném překladu systém pro správu překladů. Je to tedy program, v jehož prostředí pracují zejména manažeři překladatelských projektů a v nichž probíhá zadávání zakázek k překladu, správa zdrojů (TM a TB, viz dále), nastavení strojového překladu, zadání kroků v překladatelském projektu (tzv. *workflow steps*) apod.

Naproti tomu nástroje CAT jsou systémy vyvinuté specificky za účelem zefektivnění práce překladatelů. Základní popis nástrojů CAT považujeme pro tuto práci důležitý, jelikož se jedná o prostředí, ve kterém nejčastěji probíhá proces posteditace strojového překladu, a seznámení s tímto konceptem je tedy zásadní pro pochopení kontextu posteditační činnosti. Mezi základní funkce nástrojů CAT se řadí překladová paměť (TM), jež by se dala popsat jako databáze v minulosti přeloženého obsahu, a terminologická databáze (TB, někdy také glosář), což je vlastně slovník, ze kterého může překladatel čerpat terminologii nebo ji do něj přidávat, a rozšiřovat ho s postupně přibývajícimi překlady. Obě tyto základní funkce jsou překladateli k dispozici během překládání v prostředí nástroje CAT a slouží zejména ke zvýšení produktivity překladatele, redukci repetitivních úkonů a ke konzistenci překladu napříč texty. Na tyto podpůrné nástroje, které napomáhají překladateli v tom, aby rychleji a efektivněji vytvořili překlad, je dnes napojován i strojový překlad. Pokud je v projektu využito výstupů z překladače a lidský odborník tyto výstupy upravuje, hovoříme o posteditaci strojového překladu (viz kapitola 1.2).

Mezi nejpoužívanější nástroje CAT (a TMS v jednom) podle průzkumu EUATC (Asociace překladatelských agentur Evropské unie) z roku 2020 patří Memsource, MemoQ, SmartCAT a Studio2019 – Trados. („European Language Industry Survey“, 2020, s. 52).

Součástí všech komerčních nástrojů CAT je dnes už i funkcionalita kontroly kvality (QA). Pomocí této automatické kontroly je možné předejít chybám v překladech, jako je použití terminologie (kontrola využití shod s terminologickou databází), pravopis, gramatika, interpunkce a formátování (např. dvojité mezery) apod. Dále nás QA upozorní, pokud by překladatel omylem vynechal nějaký segment a nepřeložil ho, nebo pokud by se počet a typ tagů (značek, které nesou informaci o formátu textu) v cílovém textu neshodoval se zdrojovým textem (Garcia, 2015).

Kromě automatické kontroly QA, která zajišťuje odhalení překlepů, gramatických a formálních chyb apod. okamžitě v první fázi překladu, se pak správností a kvalitou překladu mohou zabývat ještě další lidé v dalších fázích překladového procesu. Fáze, která následuje po překladu jako takovém, se nazývá revize. Norma ISO (ISO 17100:2015) definuje proces revize jako „dvojazyčné přezkoumání cílového jazykového obsahu z hlediska jeho vhodnosti pro dohodnutý účel v porovnání se zdrojovým jazykovým obsahem“⁸. Pro fázi revize jsou nástroje CAT uzpůsobené například tím, že zaznamenávají provedené změny, takže lze následně snadno porovnávat zdrojový text se všemi verzemi překladu. Další možnost, jak dále využít nástroje CAT nad rámec samotného překládání, je anotace chyb. Anotace chyb se může provádět jak na textu přeloženém člověkem, tak i na strojovém překladu. Cílem takové anotace je poskytnout zpětnou vazbu buď na výkon daného překladatele, nebo na výkon daného strojového překladače. U lidského hodnocení kvality (tak, jak se k němu přistupuje v počítačové lingvistice) je kladen důraz na dvě kategorie, a sice na plynulost a na obsahovou věrnost (v translatoologii je přístup ke kritice a hodnocení překladu mnohem komplexnější, viz např. publikace *Kvalita a hodnocení překladu: Modely a aplikace* (Zehnalová, 2015). Po identifikování chyby je dále podle instrukcí TAUS (TAUS eLearning, „Post-Editing/Reviewing“) vhodné ohodnotit její závažnost na škále od 3 do 5 stupňů (každý stupeň je také vhodné mít dostatečně definovaný).

⁸ČSN ISO 17100.

Z výše vyjmenovaných funkcionalit je zřejmé, že nástroje CAT jsou významné nejen pro překladatele, kterým pomáhají být efektivnější a konzistentní (prostřednictvím TM a TB), ale i pro celkovou kvalitu překladů, protože pomocí funkcí jako je kontrola pravopisu přímo v editoru nebo QA pomáhají automaticky odhalit přirozeně vznikající lidské chyby. Nástroje CAT i TMS se neustále vyvíjejí a uzpůsobují potřebám trhu s překlady, a můžeme předpokládat, že se díky nim bude tvorba i správa překladatelských projektů dále zefektivňovat a zkvalitňovat.

1.2 Posteditace

V této kapitole bude cílem definovat činnost posteditace strojového překladu a popsat kompetence potřebné pro vykovávání této činnosti. Dále poskytneme přehled zkušeností některých vysokých škol a univerzit, kde již proběhly kurzy zaměřené na výuku posteditace. Na základě těchto poznatků zhodnotíme důležitost praktické zkušenosti studentů s posteditací a vytyčíme cíle pro zkoumání dostupného korpusu textů pro účely zmapování reálné práce profesionálních překladatelů z pohledu lingvistických úkonů pro jazykovou kombinaci angličtina – čeština.

1.2.1 Definice a základní pojmy

Podle mezinárodní normy ISO 18587:2017 („Překladatelské služby – Posteditace výstupů strojového překladu – Požadavky“) znamená posteditovat „upravit a opravit výstup strojového překladu“. Norma dále rozlišuje dva základní stupně post-editace: a) úplnou posteditaci (*full post-editing*) a b) základní posteditaci (*light post-editing*). Úplná posteditace je definována jako proces, kterým se má „získat produkt srovnatelný s produktem získaným překladem pořízeným člověkem“⁹, zatímco definicí základní posteditace je „získat pouze srozumitelný text bez jakékoliv snahy o vytvoření produktu srovnatelného s produktem získaným překladem pořízeným člověkem“.¹⁰

Posteditace strojového překladu (MTPE) je činnost, která se v mnoha ohledech liší od překladu a dokonce i od revize (definice revize viz kapitola 1.1.2). Jak uvádí O'Brien (2002), při překladu obvykle pracujeme se zdrojovým textem za účelem vytvoření jednoho cílového textu na vysoké kvalitativní úrovni. Naproti tomu u MTPE pracujeme prakticky vzato se dvěma zdrojovými texty (jedním originálem a jedním strojovým překladem) a pomocí těch vytváříme výslednou verzi v cílovém jazyce, a to v kvalitě, která se může v závislosti na zadání lišit. Revize a MTPE

⁹ISO 18587:2017.

¹⁰ISO 18587:2017.

jsou odlišné v tom, jaké typy chyb se při vykonávání té které činnosti opravují, protože chyby způsobené člověkem jsou typově odlišné od chyb z překladače (O'Brien, 2002, s. 101). V čem se ale tyto tři činnosti naopak neliší, je záměr – vyhovět očekáváním, která od výsledného textu má zadavatel, potažmo přijímací kultura.

Výsledné produkty posteditace se mohou dělit do dvou skupin podle toho, za jakým účelem jsou vytvářeny. Liší se tedy zejména v tom, jaká je očekávaná kvalita výsledných textů (TAUS eLearning, „Post-Editing/Reviewing“). První skupinou jsou tzv. *inbound* překlady. Jsou to texty přeložené „nahrubo“ za účelem poskytnout informace těm, kdo nerozumí zdrojovému jazyku, ale není pro ně zásadní gramatická správnost a plynulost textu. Inbound překlady tedy vznikají buď většinou čistě pomocí strojového překladače bez jakéhokoliv následného zásahu člověka, nebo jsou to strojové překlady, které projdou pouze základní posteditací (opraví se pouze závažné chyby, viz dále). Tyto texty mají většinou využití např. pro interní komunikaci uvnitř mezinárodní firmy nebo pro překlad zpětné vazby od uživatelů. Druhou skupinou jsou tzv. *outbound* překlady. To jsou texty, které jsou určené k veřejné publikaci a šíření. Od takových překladů se tedy očekává vyšší kvalita, to znamená, že se jedná o plně posteditované strojové překlady, které se po zásahu člověka mají blížit kvalitě lidského překladu.

Rozdíly mezi plnou a základní posteditací MT je často diskutované téma. Existují různé přístupy, jak popsat, které kategorie úprav se vykonávají při plné a které při základní posteditaci. Pro představu zde vycházejme opět z elektronického kurzu posteditace od organizace TAUS (TAUS eLearning, „Post-Editing/Reviewing“, překlad KK), který obecné požadavky pro obě úrovně posteditace udává takto:

Základní posteditace:

- menší gramatické (a pravopisné) chyby jsou přijatelné
- odchylky/chyby v interpunkci jsou přijatelné
- pravopisné odchylky jsou přijatelné
- terminologie je srozumitelná a použitelná
- styl a vyznění nejsou urážlivé
- stylistické odchylky jsou přijatelné
- formátování je nepodstatné

Plná posteditace:

- gramatika a pravopis jsou správné
- interpunkce je správná a konzistentní
- pravopis je konzistentní (např. dělení slov)
- terminologie je přesná a konzistentní
- formulace a styl jsou vhodné pro daný obsah
- typografie je konzistentní (nadpisy a položky výčtů)
- dodrženy požadavky na formátování

Další důležité pojmy, které s posteditací souvisí, byť se jí netýkají přímo, jsou *řízený přirozený jazyk* (CNL) a *preeditace*. CNL je přirozený jazyk upravený tak, aby ho systémy strojového překladu uměly snadno zpracovat a vytvořily tak kvalitní překlad. Mezi tyto úpravy patří (TAUS eLearning, „Post-Editing/Reviewing“, překlad KK): eliminace gramatických a pravopisných chyb, zjednodušení lexika (a nejlépe omezení se na předdefinovanou terminologii), zjednodušení syntaktické struktury vět a omezení délky na max. 20-25 slov, psaní v činném rodě, vyhýbání se užití elips a zástupných zájmen (pokud možno) a další.

Preeditace je proces, při kterém profesionálové na daný jazyk vylepšují kvalitu zdrojového textu. U překladových projektů, ve kterých figuruje strojový překlad, je kvalita zdrojových textů zásadní. Zatímco lidský překladatel dokáže nedostatky zdrojového textu (ať už stylistické, gramatické či jiné) při překládání vykompenzovat, strojový překladač takto pracovat se zdrojovým textem nedokáže. Při preeditaci tedy lidé upravují zdrojový text tak, aby byla uplatněna obecná pravidla CNL a zároveň i pravidla specifická pro určitou doménu. Cílem je dosáhnout stylově, gramaticky, pravopisně i lexikálně konzistentního zdrojového textu tak, aby byl snadno přeložitelný překladačem. Některé úpravy lze i automatizovat pomocí podpůrných nástrojů, jako např. Acrolinx, které dokáží provést kontrolu pravopisu, gramatiky, stylistiky, terminologie i obratnosti jednotlivých formulací (Arenas, 2020).

1.2.2 Kdo jsou profesionální posteditoři

Jak bylo uvedeno v předchozí kapitole, činnost posteditace a překládání se od sebe z praktického hlediska liší v tom, že při překládání se pracuje s jedním zdrojovým textem, kdežto při posteditaci se dvěma (se zdrojovým textem a s referenčním materiálem z překladače). O'Brien (2002) ale dále také uvádí, že se liší i z hlediska kognitivního. Při posteditaci a při překládání se liší kognitivní procesy spojené s porozuměním zdrojovému textu. Proto si O'Brien položila otázku, zda by se posteditaci měli věnovat (a měli se jí učit) stejní lidé, kteří jsou vyučeni překladačské činnosti. Proti tomuto návrhu hovoří právě rozdílné

kognitivní i pracovní procesy probíhající při těchto dvou rozdílných činnostech. Nakonec i cíle obou činností se mohou významně lišit, např. v kýžené kvalitě cílového textu. Na druhou stranu se ale právě překladatelé zdají být ideálními kandidáty pro vykonávání posteditace, jelikož jsou vzhledem ke svým dovednostem schopni perfektně identifikovat jazykové chyby, mají potřebnou dovednost mezijazykově převádět koncepty a často mají i potřebné technologické zázemí na to, aby dokázali tuto činnost vykonávat efektivně. Na základě těchto argumentů O'Brien vyvodila závěr, že překladatelé by se měli během translatických programů mít možnost kromě překladatelské průpravy věnovat i přípravě na práci posteditorů. Zdůrazňuje ale, že by účast na prakticky zaměřené výuce posteditace měla být volitelná a měla by vycházet z osobnostních a schopnostních předpokladů jednotlivých studentů.

Jelikož se u překladu a posteditace liší kognitivní i pracovní procesy a postupy, je zřejmé, že každá z činností může vyžadovat i částečně rozdílné schopnosti těch, kdo je vykonávají. Popisu potřebných kompetencí se již věnovalo několik vědeckých článků. Gene (2020) uvádí, že pro překlad i pro posteditaci je potřeba ve stejné míře ovládat 9 shodných kompetencí, byť pro jejich využití je u každé činnosti potřeba různá míra úsilí. Jsou to: kulturní kompetence, schopnost dělat rozhodnutí, schopnost řešit problémy, schopnost dělat rešerše, a dále kompetence technické, oborové, textové, lingvistické a kompetence mezijazykového převodu (Gene, 2020, s. 5). V článku se dále uvádí, které z těchto kompetencí jsou pro posteditory stěžejní. Jsou to tři kategorie klíčových dovedností (Gene, 2020, s. 5-6):

1. Lingvistické dovednosti – posteditor musí dokonale ovládat zdrojový i cílový jazyk, musí ovládat mezikulturní komunikaci a textovou kompetenci, a dále je zapotřebí mít odbornost v daném oboru, kterého se text týká.
2. Dovednosti v oblasti využívání nástrojů – do této skupiny patří znalost technologické stránky posteditačního procesu, jako jsou systémy MT, CNL, preeditace, základní programování (umět používat makra), ovládání funkce „najít a nahradit“ (*find and replace*).
3. Ústřední kompetence – do této kategorie patří psycho-fyziologické předpoklady člověka, které umožňují uplatňovat všechny výše zmíněné kompetence v praxi, a strategické myšlení, díky kterému posteditoři dělají informovaná rozhodnutí mezi alternativami tak, aby ve výsledném textu zabránili stylistickým problémům.

Vedle těchto dovedností potřebných konkrétně pro práci s textem ještě Gene (2020) uvádí další procesní schopnosti, které posteditoři využijí při své práci, např. umět určit pořadí jednotlivých úkolů, poskytovat zpětnou vazbu pro další trénování MT nebo pro optimalizaci

zdrojových textů, případně ještě sbírat vzorky různých posteditačních problémů za účelem školení nových posteditorů. V neposlední řadě se posteditoři mají snažit držet krok s nejnovějšími technologiemi MT a nástroji pro posteditaci.

Podle O'Brien (2002, s. 102-103) jsou na posteditory kladeny oproti překladatelům navíc tyto požadavky: základní znalost MT systémů, schopnost programovat slovníky pro MT a spravovat terminologickou databázi, ovládat CNL a preediting a umět programovat makra na automatické opravování opakujících se chyb. Kromě jiného ale O'Brien také zdůrazňuje, že nelze překladače k vykonávání posteditace nutit; pro efektivní a kvalitní posteditaci je určitě zapotřebí dobrovolnost a také celkový pozitivní přístup k MT.

Závěrem lze rozdíly mezi podstatou práce překladatele a posteditora vyjádřit pomocí schopností, které jsou pro tyto činnosti stěžejní nad rámec těch, které jsou společné. Pro překladatele je to kreativita, pro posteditory jsou to řešení problémů a rozhodování se (Gene, 2020). Na základě zde uvedených rozdílů lze vyvodit, že činnost posteditace strojového překladu opravdu vyžaduje rozdílné přístupy k práci a rozdílné dovednosti oproti překládání, a tedy vyžaduje i speciální metodu výuky a praktického nácviku pro budoucí profesionály, kteří mají tuto činnost vykonávat. Jak by taková prakticky zaměřená výuka měla vypadat, jako první navrhla právě O'Brien (2002) a v této práci problematiku popíšeme v kapitole (1.3.2).

1.3 Výuka překladových technologií na univerzitách a vysokých školách

Překladové technologie jsou nezbytnou součástí překladatelské profese dnešní doby. Na tento stav přirozeně musely zareagovat i instituce, které se věnují přípravě budoucích profesionálních překladatelů a ostatních rolí na překladatelském trhu. V této kapitole poskytneme přehled toho, jak jsou v současné době technologie zařazovány do kurzů na vysokých školách a univerzitách u nás i v zahraničí. Kromě nástrojů CAT se budeme věnovat i tématům strojového překladu a posteditace právě s ohledem na odbornou přípravu studentů.

1.3.1 Výuka nástrojů počítačem podporovaného překladu

1.3.1.2 Mezinárodní pohled

První učebnice týkající se překladových technologií vycházely v angličtině po roce 2000 (Kenny, 2020), konkrétně se jedná o publikace *Electronic Tools for Translators* (Frank Austermaehl, 2001) a *Computer-Aided Translation Technology. A Practical Introduction* (Lynne Bowker, 2002). V těchto učebnicích se překladové technologie vnímají v širším slova smyslu (viz úvod kapitoly 1.1), takže obsahují například témata jako úvod do internetu nebo zdroje dostupné na CD-ROM. Zkrátka jejich cílem bylo pomoci překladatelům přesunout způsob práce z papíru do digitálního prostředí. Další podobnou publikací byla v kniha *Translation and Technology* (Quah, 2006). Všechny tyto a podobné další publikace popisují, jak fungují různé technologie a jak mohou překladatelům pomoci. Jako takové tedy předešlé téma toho, co by se v oblasti technologií mělo na překladatelských oborech vyučovat, nikoli ale jak.

První diskuze o systematické výuce překladové technologie probíhaly mezi vyučujícími překladu sdruženými v asociacích EAMT (European Association for Machine Translation) a AMTA (Association for Machine Translation in the Americas) v letech 2001 až 2003 (Kenny, 2020). Mezi témata patřily návrhy sylabů, výuková metodika, praktická cvičení i metody hodnocení. I jiné instituce začaly navrhovat kurikula výuky překladových technologií a vytvářet výukové materiály; velký vliv v této oblasti měla i Evropská unie.

Nejvýznamnější vliv na vyučování překladových technologií v evropském prostoru mělo pravděpodobně založení sítě European Master's in Translation (EMT) v roce 2009. Síť EMT je zaštitována Generálním ředitelstvím pro překlady Evropské komise (DGT) a sdružuje evropské magisterské univerzitní studijní programy zaměřené na překlad, které splňují požadované podmínky¹¹. Cílem EMT je zlepšovat kvalitu překladatelských oborů tak, aby jejich absolventi byli odpovídajícím způsobem připraveni uplatnit se na současném pracovním trhu. Jádrem projektu EMT je takzvaný kompetenční rámec, který byl poprvé vydaný v roce 2009, jeho aktualizovaná verze pak v roce 2017 (EMT Competence Framework, 2017). Kompetenční rámec EMT definuje 5 kategorií kompetencí, které by si měli absolventi členského programu studiem osvojit: jazykové a kulturní kompetence, překladatelské kompetence, technologické kompetence, lidské a mezilidské kompetence a kompetence pro poskytování služby.

¹¹ Více na https://ec.europa.eu/info/resources-partners/european-masters-translation-emt/emt-membership-criteria_en

V této práci považujeme za důležité blíže okomentovat zmíněné překladatelské a technologické kompetence. Mezi překladatelskými kompetencemi se kromě analytické práce s textem, znalostí pracovních jazyků a schopností překládat nejrůznější formáty objevuje i schopnost preeditovat zdrojový text za účelem lepší kvality výstupu z MT a schopnost posteditace na jakékoli úrovni podle zadání (EMT Competence Framework, 2017, s. 8). Mezi technologické kompetence spadá schopnost používat relevantní IT aplikace, adaptace na nové nástroje a databáze; práce s vyhledávači, korpusovými nástroji, nástroji textové analýzy a systémy CAT; práce s multimediálními soubory a zdroji, práce s webovými technologiemi; práce se strojovým překladem, znalost jeho role v procesu překladu; zhodnocení relevance využití MT pro překlad; využívání dalších nástrojů jako například software na správu workflow (European Master's in Translation, 2017, s. 9; Hrách, 2021, s. 34). Právě poskytování technologických kompetencí studentům bylo hlavním kritériem pro přijetí programů do sítě EMT. Kompetenční rámec EMT se stal často citovaným zdrojem v textech o překladatelské pedagogice a používá se i jako vzor při tvoření sylabů na překladatelských oborech v Evropě i mimo ni (Kenny, 2020, s. 12).

V roce 2012 udělali Rothwell a Svoboda v rámci evropského programu OPTIMALE průzkum mezi 50 evropskými postgraduálními překladatelskými programy týkající se toho, které technologické kompetence se jim daří studentům opravdu poskytovat a jak toho dosahují. V roce 2017 byl průzkum zopakován a EMT tak mohlo zhodnotit vývoj v této oblasti za pětileté období (Rothwell, Svoboda, 2017). Během těchto pěti let, tedy mezi lety 2012 a 2017, se členská základna sítě EMT rozrostla z 34 na 62 členy, ale pro výzkum byly osloveny i instituce mimo síť EMT. V roce 2017 dotazník vyplnilo 55 respondentů, z nichž 17 se průzkumu zúčastnilo i v roce 2012. Z celkového počtu 55 respondentů bylo 10 mimo členství v EMT. Průzkum ukázal, že ve 100 % programů se vyučuje teorie i praxe překladových nástrojů (nárůst o 7 procentních bodů od r. 2012), přičemž ještě větší nárůst byl zaznamenán v časové dotaci na výuku nástrojů. 9 % programů věnuje výuce nástrojů více než 50 % času a dalších 50 % programů jí věnuje více než 25 % času (v roce 2012 to bylo jen 7 % programů). Všechny programy vyučují využití nástrojů z pohledu překladatele, zatímco „jen“ 84 % programů se věnuje nástrojům i z pohledu projektového manažera (nárůst o 35 p.b) a 91 % programů při výuce pracuje na vícejazyčných projektech (nárůst o 23 p.b.). Zajímavý, mírně klesající trend se projevil ve výuce nástrojů z pohledu LSP. V roce 2012 se jí zabývalo 47 %, v roce 2017 už pouze 42 %. V další části výzkumu byly dotazy směřovány na způsoby využití nástrojů při výuce. V porovnání s průzkumem z r. 2012 se v r. 2017 objevilo několik nových konceptů, např. cloudové nástroje, posteditace nebo samostatné nástroje TMS. V roce 2012 jeden z respondentů uvedl, že bohužel kolegové nejsou ochotni využívat nástroje CAT ve svých

hodinách, tzn. využívat nástroje CAT nad rámec předmětů dedikovaných přímo výuce nástrojů CAT. V roce 2017 už se podobných komentářů objevilo méně a dokonce se objevily i protichůdné zkušenosti: „Schopností využívat překladové nástroje se využívá ve všech praktických překladových předmětech.“¹² Průzkum se dotazoval i na konkrétní vyučované nástroje. V obou průzkumech se jako nejčastěji vyučované nástroje ukázaly být SDL Trados Studio a SDL MultiTerm. Nově se v r. 2017 mezi odpověďmi objevily nástroje, které využívají cloudové technologie, jako jsou český Memsource nebo Wordfast Anywhere. Výsledkům dotazů týkajících se MTPE se bude věnovat kapitola 1.3.2.

Celkově tento průzkum ukázal, že v evropském prostoru jsou překladové nástroje vnímány jako důležitý aspekt překladatelské profese, a důsledkem toho jsou v překladatelských programech už úspěšně integrovány, dokonce se na ně klade čím dál větší důraz. Kromě toho se i pozitivně vyvíjí přístup překladatelů k překladovým technologiím. V komerčním i akademickém prostředí se stále ještě objevují úplné zákazy užívání MT. Naopak jinde se na technologie dívají velmi nekriticky. Je ale důležité vnímat technologie tak, že mají lidem v jejich práci pomáhat. Překladové nástroje mají být překladatelům k dispozici pro zefektivnění a zrychlení jejich práce. Průzkum EMT ukázal, že nejpokročilejší programy jsou schopné kvalitní výuku technologií nabídnout a neustále se v ní zlepšují (Rothwell, Svoboda, 2017, s. 54).

Výukové programy se adaptovaly a zařadily současné nástroje do svých kurzů, nicméně svět technologií se neustále vyvíjí. Nepřetržitý vývoj a změny znamenají pro výukové programy určité výzvy. V současné době je jako největší výzva označována otázka, jak do odborné přípravy překladatelů zařadit problematiku strojového učení (Kenny, 2020, s. 26). Nejen, že vědci přicházejí s novými způsoby, jak vůbec můžou technologie zastat lidskou práci, ale také vzniká velké množství nových komerčních nástrojů. Ve výuce překladových nástrojů není jednoduché nacházet rovnováhu mezi dvěma přístupy: a) poskytovat studentům přístup k nejmodernějším a nejpoužívanějším překladovým nástrojům na trhu (které se ale za pár let mohou pravděpodobně opět proměnit) tak, aby byli konkurenceschopní na současném trhu; a za b) předat studentům přenositelné dovednosti jako je řešení problémů a kritické myšlení, aby byli připraveni vstoupit na trh, který je proměnlivý a rychle se vyvíjí (Bowker, 2015, s. 95). Bowker navrhuje rovnováhy mezi těmito přístupy dosáhnout tak, že cílem předmětů na výuku překladových nástrojů bude učit studenty jak tyto nástroje a jejich základní funkce fungují, ale dále také proč a kdy daný nástroj použít. Obsahem předmětu by tedy neměly být jen instrukce

¹² Rothwell, Svoboda, 2017, s. 38, překlad KK.

pro použití konkrétních nástrojů, ale také teoretické porozumění překladových procesů a jakou roli v nich nástroje hrají.

Analyzováním, jak se v současnosti vybrané univerzity snaží uzpůsobit své kurzy tak, aby zvýšily zaměstnatelnost svých absolventů, se ve své diplomové práci zabýval Šanca (2018). Nástroje CAT jsou součástí některých z popisovaných kurzů. V praktické části diplomové práce Šanca navrhuje svůj vlastní kurz, který by mohly využít univerzity se zájmem o implementaci výuky CAT nástrojů do svých osnov jako úvod do těchto nástrojů. Kurz je určen primárně pro práci s nástrojem Memsource, protože autor práce vycházel při výzkumu ze zkušeností univerzit zapsaných v akademickém programu Memsource, ale podle autora je kurz aplikovatelný i na jiné cloudové nástroje. Šanca navrhuje, že ve středu technologického kurzu by měli být studenti sami, měli by se dokonce podílet na obsahu kurzu. Proaktivní přístup by měl studenty připravit na profesní dráhu, ve které budou muset sami sledovat nově vznikající nástroje a technologické pokroky a sami se s nimi učit pracovat. Další doporučení je uspořádat práci v kurzu do projektu, jelikož i v praxi se většinou pracuje v rámci překladových projektů. Při práci na celkovém překladovém projektu se studenti mají možnost seznámit se všemi kroky potřebnými pro správu projektů. Podrobný popis 13týdenního kurzu je uveden v praktické části zmíněné diplomové práce (Šanca, 2018).

1.3.1.3 Přístup českých univerzit a vysokých škol

V této kapitole shrneme současnou podobu výuky nástrojů CAT a případně práce s MT na českých univerzitách a vysokých školách, konkrétně na Ústavu translatologie Univerzity Karlovy, překladatelských oborech Masarykovy univerzity, Jihočeské univerzitě v Českých Budějovicích, Univerzity Palackého v Olomouci, Západočeské univerzity v Plzni, Univerzitě Pardubice a Ostravské univerzitě.

Ústav translatologie Univerzity Karlovy

Ústav translatologie Filozofické fakulty Univerzity Karlovy je jediným pracovištěm v ČR, které bylo přijato do sítě European Master's in Translation.¹³ Tento ústav nabízí bakalářské studijní obory Mezikulturní komunikace: překlad a tlumočení pro jazyky angličtina, němčina, španělština, francouzština a ruština; a navazující magisterské studijní obory Překladatelství a Tlumočnictví pro stejný výčet jazyků v kombinaci s češtinou. V bakalářských i magisterských

¹³ <https://utrl.ff.cuni.cz/cs/veda-a-vyzkum/european-masters-in-translation-emt/> [cit. 6.7.2021]

překladačských programech je předmět věnovaný nástrojům CAT součástí studijního plánu všech uvedených programů.

Bakalářský studijní plán¹⁴ zahrnuje pro všech pět jazykových kombinací (viz výše) jednosemestrální povinný předmět Počítačem podporovaný překlad. Anotace tohoto kurzu:

- Informace k organizaci kurzu (doporučená literatura, podmínky atestace atd.); uvedení do problematiky (strojový překlad, počítačem podporovaný překlad, integrace a využití).
- Počítačem podporovaný překlad – podrobnosti k fungování nástrojů CAT, typy textů vhodné pro zpracování pomocí nástrojů CAT, problematika překladových pamětí.
- Strojový překlad – úvod: kategorie systémů, nejnovější vývoj.
- Nejrozšířenější systémy CAT: SDL Trados Studio.
- Nejrozšířenější systémy CAT: Memsource.
- Intratextová a intertextová repetitivnost, nácvik zpracování textů v nástrojích.
- Správa terminologie, zapojení strojového překladu v nástroji CAT.
- Export souboru z prostředí CAT.¹⁵

Magisterský studijní plán¹⁶ zahrnuje pro všechny jazykové kombinace jednosemestrální povinný předmět Nástroje CAT, strojový překlad a lokalizace. Anotace tohoto kurzu:

- Informace k organizaci kurzu (předepsaná literatura, podmínky atestace atd.); uvedení do problematiky (rozdíl mezi strojovým překladem, počítačem podporovaným překladem, člověkem vyhotoveným překladem (HT), současné trendy; integrace a míra využití).
- Počítačem podporovaný překlad – podrobnosti k fungování nástrojů CAT, specifikace typů textů vhodných pro zpracování pomocí nástrojů CAT, problematika překladových pamětí.
- Počítačem podporovaný překlad – souvislosti: paradigmatická změna překladačského procesu, authoring, důsledky pro komunikáty; integrace CAT a MT. Širší pojetí počítačem podporovaného překladu: překladačský proces a elektronické nástroje.

¹⁴ Studijní plán sdruženého studia bakalářského programu Mezikulturní komunikace: překlad a tlumočení. Dostupné z: https://www.ff.cuni.cz/wp-content/themes/ffuk/data/B_2_APT_2019.doc [cit. 6.7.2021]

¹⁵ Dostupné z: <https://is.cuni.cz/studium/predmety/index.php?do=predmet&kod=ASPPP0001> [cit. 6.7.2021]

¹⁶ Studijní plán samostatného studia navazujícího magisterského programu Překladačství. Dostupné z: <https://www.ff.cuni.cz/studium/studijni-obory-plany/studijni-plany/> [cit. 6.7.2021]

- Strojový překlad - úvod: historický přehled, různé přístupy k problematice, tradice v různých jazykových prostředích, principy využití korpusů, kategorie systémů, nejnovější vývoj.
- Zpracování překladatelského projektu – příprava všech složek projektu. Požadavky na účastníky komunikačního procesu vyplývající z jeho automatizace: pre-editing, pre-treatment; tzv. alignment; převod needitovatelných formátů na editovatelné (OCR: grafické soubory, pdf).
- Nejrozšířenější systémy CAT: SDL Trados (opakování oproti bakalářské úrovni výuky počítačem podporovaného překladu a rozšiřující cvičení).
- Nejrozšířenější systémy CAT: Memsource.
- Informace o dalších systémech CAT (Wordfast, Memsource Cloud, MemoQ, Across, DéjàVu a další).
- Nástroje CAT: paměti, vzdálené paměti, konkordanční funkce; typy databází, autorská práva ve vztahu k databázovým souborům, crowd-sourcing. Zpracování různých formátů souborů (docx, xml, pdf, pptx, xlsx) v systémech CAT.
- Nástroje pro řízení projektu. Systémy pro kontrolu kvality překladu.
- Integrace různých nástrojů v rámci nejnovějších systémů CAT: rešeršní zdroje, správa terminologie, paměť, nástroje kontroly kvality. Správa terminologie: způsoby zpracování terminologického materiálu v paralelních textech apod.
- Export souboru z prostředí CAT, podrobnosti týkající se kontroly kvality; problematika revizí při využití automatizovaných systémů.
- Lokalizace, rozdíl mezi jazykovou lokalizací a lokalizací softwaru, her apod. Různé přístupy a nejrozšířenější programy.
- Výstupy ze strojového překladu. Posteditace strojového překladu. Praktický nácvik.¹⁷

Z anotací kurzů v bakalářském i magisterském programu vyplývá, že cílem předmětů je poskytnout přehled o fungování systémů počítačem podporovaného překladu (CAT), a to jak na teoretické, tak na praktické rovině. Magisterský předmět má kromě práce s nástroji CAT i širší záběr do témat lokalizace, strojového překladu, práce se soubory a dokonce obsahuje i praktický nácvik posteditace, byť půjde pravděpodobně jen o okrajovou záležitost, vezmeme-li v potaz fakt, že kurz trvá pouze 1 semestr, pokrývá poměrně velké množství témat a vyučuje se společně pro studenty napříč jazykovými kombinacemi. Posteditaci strojového překladu v prostředí nástroje CAT se dále v magisterském programu podle zkušeností autorky práce

¹⁷ Dostupné z: <https://is.cuni.cz/studium/predmety/index.php?do=predmet&kod=AMP000003> [cit. 6.7.2021]

okrajově věnuje ještě předmět Translatologická analýza. Hlavním cílem tohoto předmětu je připravit studenty na aplikaci modelů translatologické analýzy při samostatné kriticko-hodnotící a redakční práci. V části časové dotace se tento předmět věnuje i porovnání zdrojového textu a jeho překladu vyhotovenému strojovým překladačem.¹⁸ Nejedná se ale o centrální zaměření kurzu.

Výuka samostatného kurzu zaměřeného na MTPE se již na Ústavu translatologie plánuje. V červnu 2021 zde proběhla diskuze o harmonogramu zavádění takového kurzu v magisterském programu všech vyučovaných jazyků. Již v rozhovoru v diplomové práci z loňského roku (Hrách, 2020, s. lxiii) PhDr. Bc. Tomáš Svoboda, Ph.D. uvedl, že za účelem zavedení takového kurzu „bude vhodné zajistit pro vyučující doplňující školení“.

Masarykova univerzita

Na Masarykově univerzitě se překladatelství věnují některé magisterské obory na tamější Filozofické fakultě. V této práci se zaměříme na kurzy týkající se nástrojů CAT a strojového překladu na navazujícím magisterském oboru Překladatelství anglického jazyka.

Prvním relevantním předmětem je Překlad podporovaný počítačem, který se vyučuje blokově. Osnova předmětu:

- Úvod do nástrojů CAT (TM, terminologické databáze, základní funkce textových editorů a překladových formátů)
- Informace o překládání pro Evropskou komisi
- Memsources
- SDL Trados Studio 2011¹⁹

¹⁸ Dostupné z: <https://is.cuni.cz/studium/predmety/index.php?do=predmet&kod=AMPA00001> [cit. 6.7.2021]

¹⁹ Dostupné z: <https://is.muni.cz/predmet/phil/podzim2021/AJP69049> [cit. 6.7.2021], překlad KK.

Dalším souvisejícím předmětem je Lokalizace softwaru v praxi. Cílem předmětu je poskytnout vhled do principů lokalizace a celého projektového workflow v mezinárodní lokalizační firmě. Osnova tohoto kurzu:

Úvod, základy lokalizace, nejazykové aktivity a úkoly, praktická cvičení, správa terminologie a praktická cvičení, lingvistické zásady a praktická cvičení, překladové nástroje a praktická cvičení, strojový překlad a posteditace a praktická cvičení, management jazykové kvality a praktická cvičení, překlady v marketingu, transkrece a praktická cvičení.²⁰

Podle dostupných informací o osnově tohoto kurzu se můžeme domnívat, že tento kurz je zaměřen na praktickou ukázkou toho, jak probíhají různorodé lokalizační projekty v mezinárodní firmě. Součástí je i posteditace strojového překladu. Bohužel na základě veřejně dostupných informací o kurzu nelze zjistit, jaké konkrétní nástroje se v kurzu používají a jaká časová dotace je věnována konkrétním tématům.

Okrajově se tématu strojového překladu věnuje i předmět Překlad v prostředí internetu, který má v osnově uvedeno, že 1 týden se věnuje „stručnému přehledu automatických překladačů a jejich možností a omezení.“²¹

Z toho předmětu vyplývá, že ani zde není posteditace strojového překladu centrálním zaměřením žádného konkrétního kurzu.

Jihočeská univerzita v Českých Budějovicích

Na Jihočeské univerzitě v Českých Budějovicích se překladu v kombinaci angličtina – čeština věnuje navazující magisterský obor Překladatelství anglického jazyka na Ústavu anglistiky Filozofické fakulty Jihočeské univerzity. Jediný předmět související s překladovými technologiemi je zřejmě předmět Korpusová a počítačová lingvistika pro překladatele.²² Předmět se zabývá zejména korpusovou lingvistikou, ale jedním z témat v osnově je i „strojový překlad – systémy založené na pravidlech (rule-based) a na statistice (statistical machine translation), hybridní systémy. Principy fungování, možnosti a meze strojového překladu,

²⁰ Dostupné z: <https://is.muni.cz/predmet/phil/podzim2021/AJP69078> [cit. 6.7.2021], překlad KK.

²¹ Dostupné z: <https://is.muni.cz/predmet/phil/podzim2021/AJP69043> [cit. 6.7.2021].

²² Studijní plán Překladatelství anglického jazyka FF JU. Dostupné z: <http://www.ff.jcu.cz/studium/studijni-plany-ii/studijni-plany-pro-ar-2020-2021-nmgr/nmgr-neucitelske-programy/prekladatelstvi-paj-pfj.pdf> [cit. 6.7.2021].

evaluace chyb“ a „Řešení konkrétních problémů s pomocí korpusu a nástrojů strojového překladu.“²³

Na Jihočeské univerzitě není konkrétní předmět věnovaný práci s nástroji CAT, na druhou stranu má ale oproti ostatním programům předmět zaměřený na práci s korpusy, jehož součástí je i základní vhlad do problematiky strojového překladu.

Univerzita Palackého v Olomouci

Na Univerzitě Palackého v Olomouci lze překladatelství angličtiny studovat na Katedře anglistiky a amerikanistiky na oboru Angličtina se zaměřením na komunitní tlumočení a překlad v bakalářském i navazujícím magisterském programu.

Podle studijního plánu²⁴ jsou předměty Počítačem podporovaný překlad 1 a 2 jsou povinné na bakalářském programu. Anotace obou kurzů:

„Kurz představí teoretické poznatky týkající se počítačového překládání obecně (základní odborná terminologie, principiální problémy strojového překladu, konfrontace "machine translation" a "human translation", oblasti využití a způsoby aplikace MT a CAT) a umožní studentům získat základní dovednosti v práci s různými softwarovými nástroji pro počítačem podporovaný překlad - CAT (Wordfast, Kilgray's MemoQ, Atril's Dej-vu, SDL Trados TWB), tj. tvorba a správa překladových pamětí, tvorba a správa elektronických glosářů, práce s překládaným dokumentem (import, export, čištění) apod.“²⁵

Na magisterském programu by měl být Počítačem podporovaný překlad 3 řazen mezi povinně volitelné předměty, bohužel v čase vzniku této práce není k dispozici anotace. Z rozhovoru (Hrách, 2020, s.lix) s vyučujícím tohoto kurzu PhDr. Pavlem Králem nicméně víme, že posteditaci vyučují:

„V rámci semináře CAT3 se post-editací zabýváme teoreticky i prakticky v rámci konkrétních překladových projektů. Je jí věnováno asi 40 % semináře.“

²³ Dostupné z:

https://wstag.jcu.cz/portal/studium/prohlizeni.html?pc_pagenavigationalstate=AAAAAQAGMjl3NjUzEwEAAAABAAhzdGF0ZUtleQAAAAEAFc05MjZMzcyMDM2ODU0NzczODM4AAAAA**#prohlizeniSearchResult [cit. 6.7.2021].

²⁴ Studijní plán Angličtina se zaměřením na překlad a tlumočení UPOL. Dostupné z:

https://anglistika.upol.cz/fileadmin/userdata/FF/katedry/kaa/docs/studijni_obory_brozura_ATP_2012.pdf [cit. 6.7.2021]

²⁵ Dostupné z: <https://stag.upol.cz/ects/predmet/KSR/7PPP1?lang=cs;>
<https://stag.upol.cz/ects/predmet/KSR/7PPP2/?lang=cs> [cit. 6.7.2021]

Z anotace bakalářských kurzů a z výše uvedené informace od Krále můžeme vyvodit, že na bakalářském programu na Univerzitě Palackého v Olomouci jsou překladové technologie pojaty zejména z hlediska nástrojů CAT; na magisterském programu se předmět věnuje z nadpoloviční většiny taktéž technologiím a jen částečně MTPE. Taktéž zde není vyučován předmět dedikovaný posteditaci strojového překladu.

Ostravská univerzita

Na navazujícím magisterském programu Angličtina pro překlad Ostravské univerzity se mezi povinně volitelnými předměty v odborném bloku vyučuje předmět Počítačem podporovaný překlad, jehož obsahem je:

„Základní přehled CAT nástroje a jeho komponentů. Rozbalení projektu, přípravná fáze. Zpracování překladu - jazykové páry, segmenty, překladová paměť. Práce s terminologií, databáze. Zabalení projektu, dokončovací fáze. Cvičné překlady a úkoly.“²⁶

Po prostudování několika dalších studijních programů jazykových oborů Slezské univerzity, Západočeské univerzity v Plzni a Univerzity Pardubice můžeme konstatovat, že na českých univerzitách a vysokých školách se momentálně nevyučuje samostatný předmět věnovaný posteditaci strojového překladu.

Slovensko

Velmi zajímavá vědecká práce vznikla v roce 2018 na Univerzitě Konštantína Filozofa v Nitře (Absolon, 2018). Zabývá se zkoumáním současného stavu překladatelské praxe v souvislosti se strojovým překladem a posteditací na Slovensku. Jedním z výsledků této dizertační práce je vytvoření metodiky pro prakticky zaměřenou výuku posteditace. Kurz navržený v této práci by mělo být možné využívat jak při univerzitním vzdělávání translatologů, tak i pro potřeby komerčních LSP (Absolon, 2018, s. 88).

Kurz se zaměřuje na konkrétní chyby, které se vyskytují ve strojovém překladu v jazykové kombinaci angličtina – slovenština, dále pak na schopnost rychle se rozhodnout, zda výstup z MT použít či nikoliv. Absolventi kurzu by měli být dále schopni identifikovat chyby strojového překladu (pravopis, gramatika, slovosled, význam, terminologie, typografie) a seznámit se s editační činností v softwarovém prostředí (Absolon, 2018, s. 88).

²⁶ Dostupné z:

https://studistag.osu.cz/portal/studium/prohlizeni.html?pc_pagenavigationalstate=AAAAQAENDczORMBAAAAQAic3RhdGVZLZXkAAAABABQtOTlyMzM3MjAzNjg1NDc3Mzk4MAAAAAA*#prohlizeniDetail [cit. 6.7.2021]

V zahraničí již kurzy MTPE proběhly a o zkušenostech z vedení takových kurzů byly publikovány vědecké články. Shrnutí některých z nich bude obsahem následující kapitoly.

1.3.2 Zkušenosti s kurzy zaměřenými na posteditaci

V českém univerzitním prostředí momentálně neprobíhají ucelené kurzy věnované výhradně výuce posteditace strojového překladu. V některých výše uvedených programech výuka posteditace probíhá okrajově v rámci předmětů soustředěných na nástroje CAT nebo korpusovou lingvistiku. V této kapitole uvedeme hlavní argumenty pro výuku posteditace a návrhy na obsah kurzů. Dále poskytneme výtah ze zahraničních zkušeností s kurzy, konkrétně na Helsinské univerzitě a Autonomní univerzitě v Barceloně (Universitat Autònoma de Barcelona).

Na rostoucí objem textů zadávaných k překladu reaguje překladový průmysl automatizací překladového procesu a využíváním technologií, včetně strojového překladu. Průzkum z loňského roku (EUATC, 2020, s. 44) ukázal, že 78 % dotazovaných poskytovatelů překladových služeb chce buďto začít tuto službu poskytovat nebo ji začít poskytovat častěji než dosud. Činnost posteditace se ale liší od překladu (viz kapitola 1.2.1) a tím pádem i lidé, kteří posteditaci vykonávají, potřebují k této činnosti odlišný soubor dovedností (viz kapitola 1.2.2). Proto se akademičtí pracovníci z translatických studijních programů začali věnovat tématu výuky posteditace v rámci studia překladatelství. Jeden z prvních článků popisující jak by takový ucelený kurz posteditace mohl být sestaven, vydala v roce 2002 již výše zmíněná Sharon O'Brien (O'Brien, 2002). V jejím návrhu se kurz posteditace skládá ze dvou hlavních částí – teoretické a praktické.

V teoretické části navrhuje pokrýt 6 témat: 1. úvod do posteditace (témata jako proč posteditovat, jak se posteditace liší od překladu, jak klasifikovat typické posteditační chyby apod.); 2. úvod do technologie strojového překladu (historický vývoj MT, druhy MT systémů, současný stav, výhledy do budoucna atd.); 3. úvod do CNL (historický vývoj, dostupné nástroje, současná situace atd.); 4. pokročilá správa terminologie (dostupné nástroje, programování slovníků k MT, formáty pro stažení a přesunu terminologické databáze XLT, OLIF, TBX); 5. pokročilá textová lingvistika (určování textových typů, korpusová lingvistika, dostupné nástroje); 6. základy programování (počítačové zpracovávání přirozeného jazyka, makra).

Větší část kurzu by měla být věnována praktickému nácviku posteditace. Obratnost v posteditační činnosti se podle zdrojů pro článek O'Brienové získává postupně s přibývajícím praxí. Aby se posteditor začal při posteditování cítit komfortně, měl by mít za sebou praxi

100 000 slov, což odpovídá zhruba 1 měsíci posteditování na plný úvazek (O'Brien, 2002, s.99). Kurz si nedává za cíl takový objem se studenty pokrýt (studenti by si měli posteditaci sami trénovat i mimo kurz). Namísto toho se kurz zaměří na pokrytí následujících aspektů: práce s různými překladači (alespoň se 2 komerčně dostupnými), různé úrovně posteditace a posteditace výstupů ve všech pracovních jazycích. Dále by studenti měli analyzovat a porovnávat výstupy z různých systémů a při různém nastavení, s použitím specifické terminologie, a v neposlední řadě by se měli seznámit s výhodami a nevýhodami s integrací MT do TMS.

Na závěr autorka uvažuje o vhodném zařazení takového kurzu v sylabu a rozložení rozsahu v rámci semestrů. S vědomím faktu, že každý studijní program je uspořádán odlišně, autorka navrhuje pouze pár obecných předpokladů, které by měli studenti splňovat v čase, kdy začínají tento kurz. Měli by mít vynikající jazykové a základní lingvistické znalosti, překladatelské dovednosti, základní dovednosti pro správu terminologie, IT dovednosti a základy překladové technologie. Na základně těchto předpokladů autorka navrhuje, aby se tento předmět vyučovat buďto na konci bakalářského nebo nejlépe až v navazujícím magisterském programu (O'Brien, 2002, s. 105).

Universitat Autònoma de Barcelona

V roce 2019 vyšel v časopise *The Journal of Specialised Translation* článek (Arenas, Moorkens, 2019) pojednávající o kurzu MT a posteditace v magisterském translatologickém programu (Master de Tradumàtica) na univerzitě v Barceloně. Tento program nabízí studentům výuku zaměřenou na překladové technologie (SDL Trados, MemoQ, WordFast, Catalyst), lokalizační inženýrství, lokalizaci softwaru a obrázků, makra, kontrolu kvality, analýzu chyb nebo projektový management. V roce 2009 byl poprvé součástí programu i modul posteditace. Při plánování tohoto modelu bylo vycházeno z výše zmíněného návrhu O'Brienové. Hlavním cílem modulu bylo seznámit studenty se základy posteditace, naučit je porozumět souvisejícím konceptům (kategorizace chyb MT, metriky produktivity, hodnocení kvality apod.) a hlavně poskytnout praktický nácvik posteditační činnosti tak, aby se cítili připraveni přijmout posteditační zakázku v profesním životě. Obsahem tohoto 8hodinového modulu jsou:

1. Základní definice posteditace (koncepty, posteditace vs. revize, profil posteditora, CNL a preeditace)

2. Kvalita (kvalita MT výstupu, hodnotící metriky a nástroje, očekávaná kvalita v posteditační zakázce a úroveň posteditace, metriky pro hodnocení posteditovaného obsahu)
3. Typy posteditace (základní a plná)
4. Obecná pravidla posteditace (pokyny pro základní a plnou posteditaci)
5. Běžné chyby MT systémů (příklady terminologických, gramatických, pravopisných, syntaktických, interpunkčních a stylistických chyb)
6. Úsilí a produktivita při posteditaci (časové, kognitivní a technologické úsilí)
7. Posteditace a cenotvorba (různé modely odměňování a vyjednávání)
8. Praktická cvičení (překlad technického textu bez jakékoliv pomoci, porovnání textu s výstupy z různých překladačů a určování kategorií chyb, cvičení na CNL, jednojazyčná posteditace, posteditace s použitím dvou různých TMS nástrojů podle pokynů)

Takto koncipovaný kurz je dále doplněn 8hodinovým kurzem o MT technologiích od pravidlových systémů až po aktuální NMT systémy. Autoři článku jako nejdůležitější aspekty kurzu hodnotí vysvětlení základních pojmů, kombinaci individuálních i skupinových praktických cvičení a hlavně představení reálných příkladů zadání z praxe. V neposlední řadě autoři zmiňují i rychlý vývoj v oblasti technologií a tedy i nutnost flexibility ze strany univerzit. Aby byli studenti co nejlépe připraveni pro vstup na pracovní trh, je nutné syllabus neustále upravovat vzhledem k aktuálnímu stavu technologií.

Helsinská univerzita

V článku z roku 2015 (Koponen, 2015) popisuje autorka plánování, průběh i hodnocení kurzu posteditace, který se vyučoval na Helsinské univerzitě v roce 2014. I zde se při sestavování kurzu vycházelo z návrhu O'Brienové (2002). Kurz, který byl na této univerzitě připraven, je určen jak pro bakalářské, tak i magisterské programy, a sestává ze 7 dvouhodinových přednášek, 5 domácích úkolů, praktických cvičení v hodině a reflexivní eseje po absolvování kurzu. Cílem kurzu bylo předat úvod do používání MT a posteditace v praxi a kurz se skládal z teoretické i praktické části, konkrétně z těchto témat:

1. Teorie a historie MT a posteditace
2. Praktické využití MT a posteditace
3. CNL a preeditace MT
4. Posteditace bez zdrojového textu
5. Výzkum v oblasti posteditační činnosti
6. Úrovně kvality a pokyny k posteditaci
7. Hodnocení kvality MT a úsilí při posteditaci
8. Posteditační kompetence

Z hodnocení kurzu vyplývá, že studenti si osvojili základní principy systémů MT, dokázali určit typy chyb a vyzkoušeli si, zda je pro ně posteditace strojového překladu rychlejší než překládat bez strojové pomoci. Negativně studenti hodnotili jednojazyčnou posteditaci, protože měli problémy překladačům důvěřovat bez možnosti porovnání se zdrojovým textem. Dále studenti nepovažovali za příliš přínosné preeditovat text před přeložením. Z hodnocení praktických cvičení vyplynulo, že hodnocení chyb MT studentům pomohlo si více uvědomit problémy MT systémů. Za problematické však studenti považovali rozlišování mezi požadovanými úrovněmi posteditace a většina studentů uvedla, že měli tendenci opravovat více, než bylo nutné, zejména u základní úrovně posteditace. Zároveň ale tito studenti uvedli, že považují za důležité naučit se přizpůsobovat svou práci danému zadání. Obecně autoři článku vyzdvihují jeden konkrétní aspekt, který lze považovat za přínosný, a sice změnu přístupu studentů k MT a k překladových technologiím od často zcela odmítavého k neutrálnímu nebo vyloženě pozitivnímu přístupu. Někteří ze studentů dokonce uvedli, že je činnost posteditace láká jako možné budoucí zaměstnání.

1.4 Shrnutí teoretické části

V teoretické části jsme shrnuli vývoj i současný stav překladových technologií. Představili jsme způsob používání nástrojů CAT a TMS, a zároveň byly popsány technologie, které stojí za fungováním strojových překladačů. Z neustále postupujícího vývoje v technologické oblasti vyplývá i neustále se zvyšující kvalita strojového překladu. V současné chvíli trhu dominuje NMT, díky kterému se kvalita výstupu překladačů skokově zlepšila. Dokonce se toto téma dostalo i do povědomí široké veřejnosti, a specificky v českém kontextu byly např. o překladačích DeepL a CUBBITT napsány populárně naučné články. S postupující globalizací ve světě stále narůstá i objem textů zadávaných k překladu. Stále se zvyšující kvalita strojových překladačů se tak nabízí jako odpověď na rostoucí poptávku. V současné chvíli však existuje

jen omezené množství kontextů, ve kterých by se dal strojový překlad jako takový bezpečně uplatnit jako koncový produkt, proto je velmi aktuální téma posteditace.

V teoretické části bylo uvedeno, jak se v překladatelském oboru činnost posteditace chápe, jak se liší od překládání, a jaká očekávání se kladou od lidí, kteří tuto činnost vykonávají. Na základě analýzy vědeckých článků můžeme konstatovat, že posteditace je v několika aspektech odlišná od překládání, a dokonce podle některých autorů vyžaduje i určité osobnostní předpoklady. Na základě těchto skutečností – tedy zvyšujícího se zájmu o činnost posteditace a zároveň odlišností mezi posteditací a překládáním, považujeme za důležité, aby se činnost posteditace vyučovala v rámci translatologických programů na vysokých školách a univerzitách jako ucelený kurz, aby byli absolventi těchto programů kvalifikovaně seznámeni s touto oblastí činnosti a v případě zájmu byli připraveni vstoupit na pracovní trh.

Při analýze vyučovaných kurzů v českém akademickém prostředí se ukázalo, že se programy zaměřují na výuku práce s překladovými nástroji a případně okrajově na strojový překlad a posteditaci. Nikde ale neprobíhá kurz věnovaný výhradně práci se strojovým překladem a posteditací. Následně bylo při prostudování akademických článků o proběhlých posteditačních kurzech v zahraničí zjištěno, že nejdůležitější aspekt posteditační činnosti je praktická zkušenost. To znamená, že studenti by si měli prakticky vyzkoušet analýzu výstupů různých překladačů, aby se seznámili s typy chyb, které překladače produkují, a následně by si měli vyzkoušet posteditaci jako takovou. Cílem takového kurzu podle zkušeností zmíněných zahraničních kurzu je, aby se studenti seznámili s důležitými aspekty práce se strojovým překladem (ať už se jedná o hodnocení kvality MT, metriky pro hodnocení posteditace, začlenění do nástrojů CAT atd.) a aby měli možnost praktického nácviku posteditace, který by jim dodal pocit připravenosti vykonávat tuto činnost na pracovním trhu, a připravil je pracovat s různými typy zadání.

2 PRAKTICKÁ ČÁST

V praktické části budeme pracovat s textovým datasetem a analyzovat, jaké lingvistické kategorie úprav se objevovaly během posteditace strojového překladu různých textů ve směru z angličtiny do češtiny, a dále také jaké kategorie úprav dělali revizoři v dalším kroku revize ještě po této posteditaci. Cílem bude seznámit se s typy chyb, které posteditoři a následně revizoři opravovali při práci na jazykovém materiálu přeloženém strojovými překladači z angličtiny do češtiny. Výsledné kategorie úprav by mohly posloužit jako základ pro instrukce případného posteditačního kurzu pro překladatele s jazykovou kombinací angličtina – čeština.

2.1 Analyzovaný dataset

2.1.1 Vznik datasetu

Dataset zkoumaný v empirické části nevznikl primárně za účelem této práce. Vznikl na základě zadání společnosti Memsource s cílem zkoumat závislost rychlosti a kvality profesionálního překladu formou posteditace na kvalitě výstupu strojového překladače. Tento výzkum, iniciovaný společností Memsource, probíhá v době vzniku této práce na ÚFAL MFF UK pod vedením doc. RNDr. Ondřeje Bojara, Ph.D. Řešitelka této diplomové práce je zaměstnaná ve společnosti Memsource a spolu se zástupcem společnosti se domluvili, že Memsource dataset poskytne řešitelce pro účely této práce.

Dataset se skládá ze souboru zdrojových textů (originálně v anglickém jazyce), jejich strojových překladů do češtiny vytvořených několika překladači, dále jejich posteditací od skupiny profesionálních překladatelů (v této práci budou nazýváni *posteditoři*) a následných revizí další skupinou profesionálních překladatelů (v této práci budou nazýváni *revizoři*). Ve fázi posteditace obdrželi někteří posteditoři k některým textům naslepo místo výstupu z MT referenční překlad pořízený člověkem, někteří také některé obdrželi pouze zdrojový text, a tak vznikly nové referenční překlady pro fázi revize. Fáze posteditace i revize probíhaly v účtu Memsource spravovaném firmou České překlady s.r.o.

2.1.3 Zdrojové texty

Zdrojové texty byly zadavatelem vybrány tak, aby byly zastoupeny různorodé obsahové oblasti, jelikož se předpokládá, že s různými typy textů mohou odlišně fungovat různé strojové překladače a následně s nimi mohou i odlišně pracovat posteditoři. Celkem bylo použito 8 různých textů. Průměrný počet segmentů na text je 12,4 segmentu.

Výčet zdrojových textů

Při určování charakteristiky textů vychází autorka z publikace *Stylistika češtiny* (Chloupek, 1991).

Název	Charakteristika	Obsahové téma
audit_r	Publicistický styl, funkce informativní, komuniké.	Doporučení vyplývající z auditu nejvyšších kontrolních institucí České republiky (NKÚ) a Polska (NIK).
audit_i	Publicistický styl, funkce informativní, komuniké.	Audit nejvyšších kontrolních institucí České republiky (NKÚ) a Polska (NIK).
hole	Populárně naučný styl, funkce informativní a vzdělávací, reportáž.	Astronomický úkaz, při kterém hvězda zmizela v černé díře.
leap	Odborný styl, slohový postup popisný s praktickým cílem, návod.	Uživatelské pokyny pro instalaci aplikace.
china	Publicistický styl, funkce informativní, zpráva.	Čína a její vyjádření o vlastních vojenských aktivitách.
turner	Publicistický styl, funkce informativní, vzdělávací, oznámení.	Výstava výtvarných děl a vyhlášení cen.
whistle	Publicistický styl, funkce informativní, zpráva.	Deník The Times zveřejnil informace o whistleblowerovi a sklídl za to kritiku.
lease	Odborný (právnícký) styl, funkce administrativní, smlouva.	Podnájemní smlouva.

2.1.3 Překladače

Všechny zdrojové texty byly přeloženy 13 různými strojovými překladači, konkrétně překladačem Google, Microsoft a 11 různě kvalitními variantami překladače CUBBITT vyvíjeného na ÚFAL (viz kapitola 1.1.1.3). Varianty překladače z ÚFAL jsou v datasetu označeny vzestupně od nejnižší po nejvyšší kvalitu kódy m01-m11. Kvalita jednotlivých verzí se liší na základě množství dat, které bylo použito pro natrénování dané verze, a byla změřena metrikou BLEU (viz kapitola 1.1.1.4). Dále je důležité zmínit, že verze m01-m10 byly trénované na úrovni jednotlivých vět, zatímco verze m11 byla natrénována na úrovni dokumentů. Podrobně jsou tyto dva způsoby trénování překladačů popsány v článku *CUNI English-Czech and English-Polish Systems in WMT20: Robust Document-Level Training* (Popel, 2020).

Různá kvalita výstupů překladačů byla pro primární výzkum zadaný firmou Memsource žádoucí, protože cílem je zkoumat závislost rychlosti a kvality profesionálního překladu formou posteditace na kvalitě výstupu strojového překladače. Ze stejného důvodu byl každému posteditorovi předložen soubor textů, z nichž každý byl přeložen jiným překladačem. Tento způsob kombinování výstupů MT měl zajistit, aby každý posteditor pracoval s různě kvalitními výstupy MT, a bylo tak možné pozorovat, zda kvalita MT ovlivní rychlost posteditace v rámci činnosti jednoho posteditora.

Zastoupení různě kvalitních překladačů je výhodné i pro výzkum této práce. Cílem práce je zmapovat kategorie úprav, se kterými se posteditoři MT mohou setkat. Věříme, že v praxi se posteditoři v současné době setkávají s různě kvalitními výstupy MT, proto očekáváme, že díky těmto datům bude výsledný výčet kategorií úprav reprezentativní pro širokou škálu reálných zakázek z praxe.

2.1.4 Posteditoři a revizoři

Posteditory a revizory zajistila společnost České překlady s.r.o. S kontaktní osobou na této straně byla řešitelka taktéž ve spojení, a tak získala zadání, které bylo posteditorům a revizorům posláno, a také dotazník o předchozích zkušenostech těchto pracovníků s překladem a posteditací strojového překladu.

Zadání

Posteditoři

Pro fázi posteditace byla vybrána skupina 15 posteditorů, kteří jsou v Českých překladech zaměstnaní, nebo jsou na volné noze a s Českými překlady spolupracují. Zadavatel na straně Českých překladů poskytl posteditorům přihlašovací údaje k účtu Memsource a odkaz na projekt, kde byl nahraný soubor k překladu. Soubor pro každého posteditora obsahoval všech 8 zdrojových textů. Pro každý segment zdrojového textu měl posteditor k dispozici 100% shodu z TM v cílovém jazyce. Ta pocházela většinou ze strojového překladu (každý text přeložen jiným) nebo z referenčního překladu (v datasetu označený zkratkou *ref*). Posteditoři neměli možnost zjistit, z jakého zdroje výstupy v cílovém jazyce pochází. V některých případech obdrželi pouze zdrojový text (v datasetu označený zkratkou *src*), přičemž jejich úkolem bylo text přeložit. S těmito podmínkami byli posteditoři seznámeni.

Dále zadání obsahovalo pokyny pro samotnou posteditační činnost. Instrukce neobsahovaly požadovanou úroveň posteditace nebo konkrétní chyby, na které se mají zaměřit. Jediným požadavkem bylo, aby posteditoři použili materiál v cílovém jazyce a upravili ho podle potřeby tak, jako by se jednalo o běžnou zakázku s připojenou TM. V případě, že by byl materiál v cílovém jazyce nekvalitní, mohli posteditoři návrh odmítnout a přeložit daný segment samostatně.

Poslední součástí zadání, důležitou pro primární výzkum zadaný firmou Memsource, byly instrukce o časovém rozvržení práce na tomto souboru. Jelikož pro výzkum je důležitý i čas, jaký posteditací jednotliví posteditoři stráví (automaticky měřeno v nástroji Memsource), dostali instrukci, aby případné delší přestávky ve zpracování překladu dělali vždy po dosažení konce dílčího dokumentu.

Po fázi posteditace všemi 15 posteditory vzniklo celkem 1483 segmentů dat. Rozsah této práce, jejímž cílem je vytvořit výčet kategorií úprav při posteditaci, nevyžaduje pracovat se všemi segmenty. Rozhodli jsme se tedy využít pouze výstupy práce 8 posteditorů (více v kapitole 2.1.5).

Revizoři

Pro následnou fázi revize bylo vybráno 17 revizorů taktéž z okruhu Českých překladů, kteří taktéž na revizi pracovali uvnitř projektu v Memsource. Každý revizor revidoval soubor všech 8 textů, přičemž každý z nich pocházel od jiného posteditora. Revizoři byli obeznámeni s tím, že jednotlivé dílčí dokumenty byly přeložené různým způsobem (neupravený výstup MT, MTPE nebo překlad pořízený člověkem).

Zadání k provádění úprav znělo tak, že v této fázi mají opravovat hlavně skutečné chyby - nesprávný překlad, nepřesnosti, gramatické/pravopisné chyby, stylistické problémy. Preferenční změny byly povolené, ale ne nezbytné. Pokud byl překlad v pořádku, byť by revizor třeba subjektivně formuloval něco jinak, neměl překlad opravovat.

Dotazník

Součástí spolupráce s posteditory a revizory na této studii byl také dotazník, který se skládal z několika otázek na preferovaný způsob práce nebo např. i na využívání MT v mimopracovním životě. Pro tuto práci považujeme tyto informace za nepodstatné, proto je nebudeme komentovat.

Pro lepší představu o profilu zúčastněných posteditorů a revizorů zde poskytneme pouze informace o počtu let předchozích zkušeností s překládáním a s činností MTPE (podrobněji v Příloze).

Posteditoři

Nejméně zkušený z 8 editorů, jejichž posteditace jsou předmětem zkoumání této práce, v dotazníku uvedl, že má 0 let zkušeností s překládáním a 0 let zkušeností s posteditací. Naopak mezi nejzkušenější posteditory se řadil respondent s 28 lety zkušeností s překladem a 5 lety s posteditací a respondent s 25 lety překladatelských 10 let posteditorských zkušeností.

Průměrná zkušenost těchto 8 posteditorů s překládáním je 12,4 roku a s posteditací 4 roky.

Revizoři

Mezi nejméně zkušené revizory zapojené do výzkumu patří respondent se 4 lety překladatelských a 3 lety posteditorských zkušeností nebo respondent s 12 lety překladatelských zkušeností a 0 let zkušeností s posteditací. Naopak nejzkušenějším revizor uvedl, že má 35 zkušeností s překladem a 15 let s posteditací, druhý nejzkušenější revizor uvedl 25 let zkušeností s překladem i posteditací.

Průměrná zkušenost všech 17 revizorů je 16,9 roku zkušeností s překládáním a 5,6 roku s posteditací.

Celkem se na studii podílela rozmanitá skupina osob, kteří momentálně pracují v překladovém průmyslu (minimálně vykonávají překladatelskou činnost pro České překlady), ale jejich zkušenosti se liší. Pro výzkum kategorií úprav, které posteditoři dělají při MTPE, je cenné, že máme k dispozici výsledky práce jak profesionálů, kteří už se činnosti věnují několik let a mají mnoho zkušeností, tak i začátečníků, kteří zatím v této fázi své kariéry nejsou ovlivněni svými vlastními ustálenými návyky při posteditaci.

2.1.5 Anotovaný výsek dat

Jak již bylo zmíněno výše, pro účely této práce nebylo nutné použít dostupná data v plném rozsahu, jelikož cílem bylo analyzovat typy úprav a shromáždit je do kategorií, jejichž počet je omezený. Po určitém množství analyzovaných segmentů se nové kategorie chyb již neobjevovaly a analyzování dalších segmentů by již patrně práci tohoto rozsahu nepřineslo žádné nové poznatky. Z výstupů 15 posteditorů jsme analyzovali práci 8 z nich. Těchto 8 bylo vybráno tak, aby byly zastoupeny osoby s různě dlouhými zkušenostmi a aby byly zastoupeny všechny překladače ve srovnatelném počtu.

Analyzovaný dataset se tedy skládal z výstupů od 8 posteditorů, kteří pracovali každý s 8 stejnými zdrojovými texty, nicméně každý z nich pracoval s jinými výstupy v cílovém jazyce. Odkaz na anotovaný dataset (včetně anotací) je uveden v Příloze. Zdroje výstupů v cílovém jazyce byly zastoupeny následovně:

	UFAL_CP13	UFAL_CP3	UFAL_CP4	UFAL_CP8	UFAL_CP9	UFAL_CP14	UFAL_CP10	UFAL_CP11
audit_r	m01	m05	m11	src	m03	m10	google	m08
audit_i	m10	m11	m02	m03	google	m04	m08	m07
hole	m09	m03	google	m04	m06	microsoft	m05	m11
leap	m06	m09	microsoft	m08	m07	m05	m01	m10
china	m11	ref	src	m01	m10	m02	m04	m06
turner	m05	microsoft	ref	m07	m01	m11	m10	m04
whistle	m02	src	m03	m10	m04	m09	m06	m05
lease	m04	m02	m09	google	m08	m06	m07	m01

Tabulka 1: řádky – jednotlivé texty, sloupce – jednotliví posteditoři (jejich identifikační čísla). Uvnitř jsou uvedeny zdroje překladu daného textu, jež posteditoval daný posteditor. Src = posteditor měl k dispozici pouze zdrojový text a vyhotovil vlastní překlad. Ref = posteditor prováděl úpravy referenčního překladu pořízeného člověkem.

Z uvedené tabulky je vidět, že u žádného textu nebyl použit dvakrát stejný zdroj výstupu v cílovém jazyce, tudíž nikdy dva různí posteditoři nepracovali se stejným výstupem a tedy nelze směrodatně porovnávat posteditory mezi sebou.

Předmětem zkoumání budou úpravy provedené:

1. posteditory na výstupech z MT systémů a na referenčních překladech. Překlady, které posteditoři vyhotovili samostatně od nuly, se nebudeme v kategorizaci úprav při fázi posteditace zabývat. Celkově tedy analyzujeme 766 segmentů (1 segment většinou odpovídá 1 větě nebo souvětí). Pro každý segment máme k dispozici zdrojový text v angličtině, výstup MT nebo referenční překlad v češtině a posteditaci tohoto výstupu.
2. revizory na výstupech posteditorů. Revizoři měli k dispozici zdrojový text v angličtině, výstup MT nebo referenční překlad v češtině a posteditaci tohoto výstupu. Jejich úkolem bylo revidovat tyto posteditace. Analyzujeme 766 revidovaných segmentů v porovnání s výstupy posteditace.

2.2 Kategorizace úprav

V této kapitole shrneme postup při tvoření kategorií úprav prováděných při posteditaci strojového překladu a při revizi posteditovaných výstupů v jazykové kombinaci angličtina – čeština a představíme výsledný výčet těchto kategorií s definicemi a vybranými příklady z analyzovaného datasetu.

2.2.1 Postup

Na počátku analýzy jsme si definovali základní skupiny úprav, ke kterým jsme očekávali, že bude docházet, např. úpravy slovosledu, syntaxe nebo terminologie. Poté jsme v první fázi pečlivě porovnávali výstup strojového překladače (případně referenční překlad) s konečným výstupem po zásahu posteditora, segment po segmentu. V případě potřeby jsme nahlíželi do zdrojového textu. Zde považujeme za důležité zmínit, že cílem analýzy nebylo hodnotit kvalitu výstupů daných překladačů. Cílem bylo zaměřit se na změny, které vzhledem k těmto výstupům provedli posteditoři. Ve druhé fázi jsme opět segment po segmentu porovnávali posteditované segmenty s výstupy od revizorů. Autorčin záměr není hodnotit výkon jednotlivých posteditorů nebo revizorů, kromě případů, kdy při analýze autorka identifikovala úpravu výstupu ze správného na chybné řešení (dle vyhodnocení autorky).

Každou provedenou úpravu jsme se snažili zařadit do jedné ze základních skupin úprav. Pokud se úprava nedala zařadit do žádné z těchto skupin, definovali jsme novou kategorii. V případě, že v jedné skupině byly úpravy, které bylo možné od sebe typově odlišit, vytvořili jsme zvlášť kategorie se svými vlastními definicemi. Po zhruba 200 zanalyzovaných segmentech se již dařilo všechny provedené úpravy zařadit do jedné z definovaných kategorií. Analýza probíhala tak, že jsme v tabulce s daty do řádku vedle analyzovaného segmentu přidávali zkratky

detekovaných kategorií, a tak byl každý segment označen zkratkami těch kategorií úprav, ke kterým v něm podle autorčina vyhodnocení došlo. Pokud v segmentu došlo k více výskytům stejné kategorie, snažili jsme se přidat jedno označení k jednomu významovému celku uvnitř segmentu.

2.2.2 Kategorie úprav ve fázi posteditace

Výsledný výčet kategorií úprav prováděných při posteditaci strojového překladu v jazykové kombinaci angličtina – čeština po analýze 766 segmentů se skládá ze 14 kategorií, které jsou uskupeny v 6 zastřešujících skupinách. V následujícím výčtu bude uveden název skupiny, název kategorie, zkratka kategorie používaná v datasetu, definice kategorie a několik příkladů segmentů obsahující tuto kategorii úpravy z datasetu. Uvnitř segmentu bude úprava dané kategorie zvýrazněna podtržením.

1. Skupina PŘESNOST

a. Omise (v datasetu označeno zkratkou PŘESom)

Definice: Informace ze zdrojového textu je v navrhovaném překladu vypuštěna.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
To install a program just insert the CD with the program or run <u>the.Exe file</u> and follow the installation steps.	Pro instalaci stačí vložit CD s programem nebo <u>spustit</u> . <u>Spust'te</u> soubor a postupujte podle instalačních kroků.	Pro instalaci programu stačí vložit CD nebo spustit <u>soubor.exe</u> a projít jednotlivými kroky instalace.
China on Friday said it has not provoked a "single war or conflict" or "invaded a single square" of foreign land, <u>skirting any reference to the 1962 war with India.</u>	Čína v pátek tvrdila, že nevyvolala „jedinou válku nebo konflikt“ nebo „napadla jeden čtverec“ v cizí zemi, a to v roce 1962, a to v Indii.	Čína v pátek uvedla, že nevyvolala „jedinou válku nebo konflikt“ ani „nenapadla jediný čtvereční metr“ jiné země, <u>příčemž se důsledně vyhnula jakékoliv zmínce o válce s Indií z roku 1962.</u>
Karolína Černá, <u>Lessee</u>	Karolína Černá	Karolína Černá, <u>podnájemník</u>
This Supplement <u>No. 1</u> shall take effect on the date of signature by both Contracting parties.	Tento <u>dodatek 1</u> nabývá účinku dnem podpisu obou smluvních stran.	Dodatek <u>č. 1</u> nabývá účinku dnem podpisu obou smluvních stran.

b. Chybný překlad (v datasetu označeno zkratkou PŘESchyb)

Definice: Informace v navrhovaném překladu neodpovídá významově zdrojovému textu.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Make sure the game is compatible with Windows 8. <u>Check on the box of the game</u> or in the instruction manual.	Ujistěte se, že je hra kompatibilní se systémem Windows 8. <u>Zaškrtněte políčko hry</u> nebo v návodu k obsluze.	Ujistěte se, že je hra kompatibilní se systémem Windows 8. <u>Zkontrolujte krabici</u> nebo manuál hry.
China Says <u>It Didn't Fight Any War Nor Invaded Foreign Land</u>	Čína říká, že to není žádná válka, ani <u>cizí země</u> .	Čína říká, že <u>nevedla válku</u> ani <u>nenapadla jinou zemi</u>
In Art. III of the Sublease agreement, entitled "Term of the Lease," <u>the tenant and the lessee</u> agreed that the apartment in question would be rented to the tenant for a fixed period from 13th May 2016 to 31st December 2018.	V čl. III smlouvy o podnájmu s názvem „Doba nájmu“ se <u>nájemce a nájemce</u> dohodli, že dotyčný byt bude nájemci pronajat na dobu určitou od 13. května 2016 do 31. prosince 2018.	V čl. III smlouvy o podnájmu, „Doba nájmu“, se <u>nájemce a podnájemce</u> dohodli, že předmětný byt bude nájemci pronajat na dobu určitou od 13. května 2016 do 31. prosince 2018.
The support can be granted to eligible projects implemented in the regions and <u>voivodeships</u> adjacent to the joint border.	Podpora může být poskytnuta na způsobilé projekty prováděné v regionech a ve <u>vodách</u> sousedících se společnou hranicí.	Podpora může být poskytnuta na způsobilé projekty prováděné v krajích a ve <u>vojvodstvích</u> sousedících se společnou hranicí.
"China safeguards world peace through real actions. Over the past 70 years, China has not provoked a single war	„Čína chrání světový mír skutečnými činy. Za posledních 70 let nevyvolala Čína jedinou válku nebo	„Čína brání světový mír skutečnými činy. Za posledních 70 let nevyvolala Čína jedinou válku nebo

or conflict, nor invaded a single <u>square</u> of foreign land," the paper titled "China and the World in the New Era" said.	konflikt, ani nenapadla jediný <u>čtverec</u> cizí země," píše se v dokumentu nazvaném „Čína a svět v nové éře“.	konflikt a nenapadla ani jeden <u>metr</u> cizího území," píše se v dokumentu nazvaném „Čína a svět v nové éře“.
---	--	--

c. Nepřeloženo (v datasetu označeno zkratkou PŘESnepř)

Definice: V navrhovaném překladu zůstala část obsahu uvedena ve zdrojovém jazyce nepřeložena.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Karolína Černá, <u>Lessee</u>	Karolína Černá, <u>Lessee</u>	Karolína Černá, <u>podnájemce</u>
Why The Times Published Details of the <u>Whistle-Blower's Identity</u>	Proč The Times zveřejnila podrobnosti o <u>Whistle-Blower's Identity</u>	Proč The Times zveřejnily podrobnosti o <u>totožnosti oznamovatele</u>
Windows provides <u>Paint</u> by default. You can also use free programs like GIMP, or professional programs such as Photoshop, for which you must pay.	Windows ve výchozím nastavení poskytují program <u>Paint</u> . Můžete také využít bezplatné programy jako GIMP nebo profesionální programy jako Photoshop, za které musíte zaplatit.	Windows ve výchozím nastavení poskytuje program <u>Malování</u> . Lze využít také bezplatné programy jako GIMP nebo profesionální programy jako Photoshop, které jsou však zpoplatněné.
<u>The Supplement</u> No. 1 is bilingual.	<u>The Supplement</u> Číslo 1 je dvojjazyčné.	<u>Dodatek</u> č. 1 je dvojjazyčný.
The owner of the flat in question is a housing cooperative "Team of Owners of House Alfrédova	Majitel bytu je „Tým majitelů domu Alfrédova 13, Praha 4“, <u>ID</u> : 123 45 678, se sídlem v Praze 4, Alfrédova 13,	Vlastníkem předmětného bytu je bytové družstvo „Společenství vlastníků domu Alfrédova 13, Praha

13, Prague 4", <u>ID</u> : 123 45 678, with registered office at Prague 4, Alfrédova 13, postal code: 142 00 (hereinafter referred to as " <u>Housing Cooperative</u> ").	poštovní kód: 142 00 (dále jen „ <u>Housing Cooperative</u> “).	4“, <u>IČO</u> : 123 45 678, se sídlem Alfrédova 13, Praha 4, PSČ 142 00 (dále jen „ <u>bytové družstvo</u> “).
---	---	---

2. Skupina PLYNULOST

a. Interpunkce (v datasetu označeno zkratkou PLYNinterp)

Definice: Interpunkce v navrhovaném překladu je nesprávná vzhledem k místním zvyklostem přijímající kultury.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
The programmes cover a variety of border regions; some work with maritime areas, others with internal EU land borders or borders shared with candidate countries.	Programy se týkají různých příhraničních regionů; některé spolupracují s námořními oblastmi, jiné s vnitřními pozemními hranicemi EU nebo s hranicemi, které jsou sdíleny s kandidátskými zeměmi.	Programy se týkají různých typů příhraničních regionů – některé se věnují námořní oblasti a jiné vnitřním pozemním hranicím mezi zeměmi EU nebo hranicím s kandidátskými zeměmi.
The following are main <u>attributes</u> : <u>joint</u> selection of leading partner; joint preparation of project and applications; making of joint agreement of partnership signed by all partners; overtaking of complete responsibility for project	Hlavními atributy <u>jsou</u> : <u>společný</u> výběr vedoucího partnera, společná příprava projektů a aplikací, společná dohoda o partnerství podepsaná všemi partnery, převzetí plné odpovědnosti za realizaci projektu vedoucím partnerem, celkové platby	Hlavními atributy <u>jsou</u> <u>společný</u> výběr vedoucího partnera, společná příprava projektů a žádostí, společná dohoda o partnerství podepsaná všemi partnery, převzetí plné odpovědnosti za realizaci projektu vedoucím partnerem a to, že celkovou

implementation by leading partner; total support payments received by leading partner who completes their transfer to respective project partners.	podpory přijaté předním partnerem, který dokončí jejich převod na příslušné projektové partnery.	platbu podpory přijímá přední partner a následně ji převádí k příslušným projektovým partnerům.
By signing this Supplement No. 1, <u>the</u> tenant then declares that her husband has also given his consent to this Supplement No. 1.	Podpisem tohoto dodatku č. <u>1</u> , <u>pak</u> nájemce prohlašuje, že její manžel dal také souhlas s tímto dodatkem č. 1.	Podpisem tohoto dodatku č. <u>1</u> <u>pak</u> nájemkyně prohlašuje, že její manžel dal také souhlas s tímto dodatkem č. 1.

b. Pravopis (v datasetu označeno zkratkou PLYNprav)

Definice: Zápis slov neodpovídá normám cílového jazyka.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Many readers, including some who work in national security and intelligence, have criticized The Times's <u>decision</u> to publish the details, saying it potentially put the person's life in danger and may have a chilling effect on would-be whistle-blowers.	Mnoho čtenářů, včetně těch, kteří pracují v národní bezpečnosti a inteligenci, kritizovalo <u>Rozhodnutí</u> Timesů zveřejnit podrobnosti a říct, že by mohlo ohrozit život člověka a mohlo by to mít mrazivý vliv na rádobu oznamovatele.	Mnoho čtenářů, včetně těch, kteří pracují v oblasti národní bezpečnosti a zpravodajských služeb, kritizovalo <u>rozhodnutí</u> deníku zveřejnit podrobnosti a namítali, že by mohlo ohrozit život člověka a mohlo by v budoucnu odradit potenciální oznamovatele.
The star, which was about the same size as our <u>sun</u> , was	Hvězda, která měla přibližně stejnou velikost jako naše	Hvězda, která měla přibližně stejnou velikost jako naše

seen from 375 million light years away warping and spiralling into the gravitational pull of a supermassive black hole, researchers said in the <i>Astrophysical Journal</i> .	<u>slunce</u> , byla viděna ze vzdálenosti 375 milionů světelných let, která se krivila a spirálovitě se táhla do gravitačního působení supermasivní černé díry, uvedli vědci v <i>Astrophysical Journal</i> .	<u>Slunce</u> , byla spatřena ze vzdálenosti 375 milionů světelných let právě ve chvíli, kdy ji deformovalo gravitační pole supermasivní černé díry, uvedli vědci v <i>Astrophysical Journal</i> .
Subject of the <u>Supplement</u> No. 1	Předmět <u>dotatku</u> č. 1	Předmět <u>Dodatku</u> č. 1
Curator Fiona Parry explains that <u>Murillo's</u> work focuses on the current political and socio-economic situation in the UK. The group of papier-mâché figures represent a globalised workforce, while the black cloth represents the "darkness of the contemporary moment in the UK and globally."	Kurátorka Fiona Parryová vysvětluje, že <u>Murilloova</u> práce se zaměřuje na současnou politickou a socioekonomickou situaci ve Velké Británii. Skupina papír-mâché postavy představují globalizované pracovní síly, zatímco černá tkanina představuje "temnotu současného okamžiku ve Velké Británii i na celém světě."	Kurátorka Fiona Parryová vysvětluje, že se <u>Murillova</u> práce zaměřuje na současnou politickou a společensko-hospodářskou situaci ve Spojeném Království. Skupina postav vyrobených z papírové hmoty představuje globalizovanou pracovní sílu, zatímco černá látka představuje "temnotu současného okamžiku ve Spojeném království i na celém světě."

c. Slovosled (v datasetu označeno zkratkou PLYNslov)

Definice: Pořadí slov a větných členů neodpovídá zvyklostem cílového jazyka.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
<p><u>On Thursday, The Times</u> published exclusive details about the identity of the whistle-blower whose claims led <u>Democrats</u> to begin an impeachment inquiry against President Trump <u>this week</u>. (The article reported that the whistle-blower is a C.I.A. officer who was previously detailed to work at the White House and had expertise on Ukraine.)</p>	<p><u>Deník The Times ve čtvrtek</u> zveřejnil exkluzivní podrobnosti o totožnosti informátora, jehož tvrzení vedla <u>demokraty tento týden</u> k zahájení vyšetřování obvinění prezidenta Trumpa z velezrady. (Článek uváděl, že udavačem je důstojník CIA, který byl dříve přidělen k práci v Bílém domě a měl odborné znalosti o Ukrajině.)</p>	<p><u>Ve čtvrtek deník The Times</u> zveřejnil exkluzivní podrobnosti o totožnosti whistleblowera, jehož tvrzení vedla <u>tento týden demokratickou stranu</u> k zahájení vyšetřování prezidenta Trumpa. (Článek uváděl, že whistleblowerem je důstojník CIA, který byl dříve přidělen k práci v Bílém domě a který je odborníkem na Ukrajinu.)</p>
<p>The president and some of his supporters have attacked the credibility of the whistle-blower, who has presented information that <u>has touched off a landmark impeachment proceeding</u>. The president himself has called the whistle-blower's account a "political hack job."</p>	<p>Prezident a někteří jeho příznivci napadli důvěryhodnost informátora, který předložil informace, jež <u>odstartovaly přelomové řízení o obžalobě z velezrady</u>. Sám prezident označil výpověď informátora za „politickou loupež“.</p>	<p>Prezident a někteří jeho příznivci napadli důvěryhodnost whistleblowera, který předložil informace, jež <u>přelomové řízení o impeachmentu odstartovaly</u>. Sám prezident označil výpověď whistleblowera za „politickou fušeřinu“.</p>
<p>The funding is granted on condition that organisations from both sides of the border,</p>	<p>Financování se poskytuje za podmínky, <u>že organizace z</u> obou stran hranic, jako jsou</p>	<p>Financování se poskytuje za podmínky, <u>že se organizace z</u> obou stran hranic, jako jsou</p>

<p>such as regional authorities, universities and small and medium enterprises, <u>come together</u> to implement projects based on the needs of the border region.</p>	<p>regionální orgány, univerzity a malé a střední podniky, <u>se spojí</u> s prováděním projektů na základě potřeb pohraniční oblasti.</p>	<p>regionální orgány, univerzity a malé a střední podniky, <u>spojí</u> za účelem realizace projektů na základě potřeb dané pohraniční oblasti.</p>
<p><u>We decided to publish</u> limited information about the whistle-blower - including the fact that he works for a nonpolitical agency and that his complaint is based on an intimate knowledge and understanding of the White House - because we wanted to provide information to readers that allows them to make their own judgments about whether or not he is credible.</p>	<p><u>Rozhodli jsme se zveřejnit</u> omezené informace o oznamovateli – včetně skutečnosti, že pracuje pro nevládní agenturu a že jeho stížnost je založena na důvěrných znalostech a porozumění Bílého domu – protože jsme chtěli poskytnout informace čtenářům, kteří jim umožní, aby se sami rozhodli, zda je důvěryhodný.</p>	<p>Omezené údaje o informátorovi – včetně skutečnosti, že pracuje pro nevládní agenturu a že jeho stížnost je založena na důvěrných znalostech a porozumění Bílého domu – <u>jsme se rozhodli zveřejnit</u> proto, že jsme čtenářům chtěli poskytnout informace, které jim umožní, aby se sami rozhodli, zda je důvěryhodný.</p>

d. Gramatika (v datasetu označeno zkratkou PLYNgram)

Definice: Text neodpovídá gramatickým pravidlům cílového jazyka.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
The Sino-India border <u>dispute</u> involving 3,488-km-long Line of Actual Control (LAC) <u>remained unresolved</u> . China also claims Arunachal Pradesh as part of South Tibet, which India contests.	Čínsko-indický hraniční <u>spor</u> zahrnující 3,488 km Linie skutečného řízení (LAC) <u>zůstala nevyřešena</u> . Čína si také nárokuje Arunáčalpraděš jako součást Jižního Tibetu, což Indie zpochybňuje.	Čínsko-indický hraniční <u>spor</u> zahrnující 3,488 km linie aktuální kontroly (LAC) <u>zůstal doposud nevyřešen</u> . Čína si také nárokuje Arunáčalpraděš jako součást Jižního Tibetu, což Indie zpochybňuje.
<u>The Supplement</u> No. 1 is <u>bilingual</u> .	<u>Dodatek</u> Číslo 1 je <u>dvojazyčné</u> .	<u>Dodatek</u> č. 1 je <u>dvojazyčný</u> .
The Contracting parties to this Supplement No. 1 declare that <u>they have read</u> the Supplement No. 1, agree with its contents and that the Supplement No. 1 was concluded freely, seriously, not in distress, under considerably unfavorable conditions.	Smluvní <u>strany</u> tohoto dodatku č. 1 prohlašují, že <u>četli</u> <u>dodatek</u> č. 1, souhlasit s jeho obsahem a že <u>dodatek</u> Číslo 1 bylo uzavřeno svobodně, vážně, ne v tísní, za značně nepříznivých podmínek.	Smluvní <u>strany</u> tohoto dodatku č. 1 prohlašují, že <u>četly</u> <u>dodatek</u> č. 1, souhlasí s jeho obsahem a že <u>dodatek</u> č. 1 byl uzavřen svobodně, seriózně, nikoliv v tísní či za značně nepříznivých podmínek.
This <i>Supplement</i> No. 1 is <u>written and signed</u> in 2 (in words: <u>two</u>) <u>copies</u> , each of which is valid for the original.	Tento <u>dodatek</u> Ne. 1 je <u>napsáno a podepsáno</u> 2 (slovy: <u>dva</u>) <u>kopie</u> , z nichž každý je platný pro originál.	Tento <u>Dodatek</u> č. 1 byl <u>sepsán</u> ve 2 (slovy: <u>dvou</u>) <u>výtiscích</u> a <u>podepsán</u> . Každý výtisk má platnost originálu.

3. Skupina LEXIKUM

a. Lexikální interference (v datasetu označeno zkratkou LEXinterf)

Definice: V navrhovaném překladu se vyskytuje lexikální volba ovlivněná zdrojovým jazykem. Jedná se buďto o chybný překlad z důvodu použití tzv. zrádných slov (*false friends*) nebo o doslovné vyjádření určitých lexikálních prvků zdrojového jazyka ve formě jinak obecně přijatelného výroku v cílovém jazyce.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
"China safeguards world peace through real <u>actions</u> . Over the past 70 years, China has not provoked a single war or conflict, nor invaded a single square of foreign land," the paper titled "China and the World in the New Era" said.	„Čína chrání světový mír prostřednictvím skutečných <u>akcí</u> . Za posledních 70 let, Čína nezpůsobila jedinou válku, ani ani neporazila na jeden čtverec cizí země, a to v novinách „Čína a svět v New Era“ řekl „Čína a svět v New Era“.	„Čína chrání světový mír reálnými <u>kroky</u> . Za posledních 70 let Čína nevyvolala jedinou válku nebo konflikt a nenapadla jediný čtvereční metr jiné země,“ uvádí bílá kniha pojmenovaná „Čína a svět v nové éře“.
The owner of the flat in question is a housing cooperative " <u>Team</u> of Owners of House Alfrédova 13, Prague 4", ID: 123 45 678, with registered office at Prague 4, Alfrédova 13, postal code: 142 00 (hereinafter referred to as "Housing Cooperative").	Vlastníkem předmětného bytu je bytové družstvo „ <u>Tým</u> vlastníků domu Alfrédova 13, Praha 4“, IČ: 123 45 678, se sídlem Praha 4, Alfrédova 13, PSČ: 142 00 (dále jen „Bytové družstvo“).	Vlastníkem předmětného bytu je bytové družstvo „ <u>Sdružení</u> vlastníků Alfrédova 13, Praha 4“, IČ: 123 45 678, se sídlem Praha 4, Alfrédova 13, PSČ: 142 00 (dále jen „Bytové družstvo“).
Representatives of cooperating regions, <u>thematic</u> ministries and other local	Zástupci spolupracujících regionů, <u>tematických</u> ministerstev a dalších	Zástupci spolupracujících regionů, <u>příslušných</u> ministerstev a dalších

partners meet regularly and decide which projects are going to be supported.	místních partnerů se pravidelně scházejí a rozhodují o tom, které projekty budou podporovány.	místních partnerů se pravidelně scházejí a rozhodují o tom, které projekty budou podpořeny.
Named after 19th-century landscape painter JMW Turner, the award <u>made stars</u> of Grayson Perry, Damien Hirst and 12 Years a Slave director Steve McQueen.	Cena pojmenovaná po krajináři JMW Turnerovi z 19. století <u>vynesla hvězdy</u> Graysona Perryho, Damiena Hirsta a 12 let otrokem režiséra Steva McQueena.	Tato cena pojmenovaná po krajináři 19. století J. M. W. Turnerovi <u>zahájila hvězdnou kariéru</u> Graysona Perryho, Damiena Hirsta a režiséra Steva McQueena, který je autorem filmu 12 let v řetězech.
<u>The white paper</u> , while highlighting the CPS's "peaceful rise" made no reference of the bloody 1962 war with India and the vast tracts of land, especially in the Aksai Chin area, occupied by China.	<u>Bílý papír</u> , který zdůrazňoval „mírumilovný vzestup“ se v roce 1962 v Indii a v celé celé zemi, a to především v oblasti Aksai Chin, v Číně.	<u>Bílá kniha</u> zdůrazňuje „mírumilovný vzestup“ Číny, přičemž se ani slovem nezmiňuje o krvavé válce s Indií z roku 1962 ani o obrovských územích, která Čína okupuje, a to zejména v regionu Aksai Čin.
The following are main attributes: joint selection of leading partner; joint preparation of project and <u>applications</u> ; making of joint agreement of partnership signed by all partners; overtaking of complete responsibility for project implementation by leading partner; total support payments received by leading	Hlavními atributy jsou: společný výběr vedoucího partnera, společná příprava projektů a <u>aplikací</u> , společná dohoda o partnerství podepsaná všemi partnery, převzetí plné odpovědnosti za realizaci projektu vedoucím partnerem, celkové platby podpory přijaté předním partnerem, který dokončí	Hlavními atributy jsou společný výběr vedoucího partnera, společná příprava projektů a <u>žádostí</u> , společná dohoda o partnerství podepsaná všemi partnery, převzetí plné odpovědnosti za realizaci projektu vedoucím partnerem a to, že celkovou platbu podpory přijímá přední partner a následně ji

partner who completes their transfer to respective project partners.	jejich převod na příslušné projektové partnery.	převádí k příslušným projektovým partnerům.
--	---	---

b. Lexikální neidiomatická (v datasetu označeno zkratkou LEXneidiom)

Definice: V navrhovaném překladu se vyskytuje lexikální volba, která není v cílovém jazyce idiomatická nebo neodpovídá terminologii ustálené v daném kontextu či typu textu.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
The funding is granted on condition that organisations from both sides of the border, such as regional authorities, universities and small and medium enterprises, come together to implement projects based on the needs of the border region.	Financování se poskytuje za podmínky, že organizace z obou stran hranic, jako jsou regionální orgány, univerzity a malé a střední podniky, se spojí za účelem <u>provádění projektů</u> založených na potřebách pohraničního regionu.	Financování se poskytuje za podmínky, že organizace z obou stran hranice, jako jsou regionální orgány, univerzity a malé a střední podniky, se spojí za účelem <u>realizace projektů</u> založených na potřebách pohraničního regionu.
The president and some of his supporters have attacked the credibility of the <u>whistle-blower</u> , who has presented information that has touched off a landmark impeachment proceeding. The president himself has called the <u>whistle-blower's</u> account a "political hack job."	Prezident a někteří jeho příznivci napadli důvěryhodnost <u>informátora</u> , který předložil informace, jež odstartovaly přelomové řízení o obžalobě z velezrady. Sám prezident označil výpověď <u>informátora</u> za „politickou loupež“.	Prezident a někteří jeho příznivci napadli důvěryhodnost <u>whistleblowera</u> , který předložil informace, jež přelomové řízení o impeachmentu odstartovaly. Sám prezident označil výpověď <u>whistleblowera</u> za „politickou fušeřinu“.

Try to <u>access</u> the Play Store and install the application.	Zkuste <u>se dostat</u> do Obchodu Play a aplikaci nainstalovat.	Zkuste <u>se připojit</u> k Obchodu Play a aplikaci nainstalovat.
This Supplement No. 1 shall take <u>effect</u> on the date of signature by both Contracting parties.	Tento dodatek Číslo 1 nabývá <u>účinku</u> dnem podpisu oběma smluvními stranami.	Tento dodatek č. 1 nabývá <u>účinnosti</u> dnem podpisu oběma smluvními stranami.

c. Lexikální preference posteditora (v datasetu označeno zkratkou LEXpref)

Definice: V navrhovaném překladu se vyskytuje lexikální volba, u které se nejedná o neidiomaticnost, nevhodnou volbu z hlediska stylu a kontextu ani lexikální interferenci, která způsobila významový posun, přesto se jí ale posteditor rozhodne upravit. Jedná se tedy o úpravu založenou na osobní preferenci konkrétního člověka.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
The other provisions of the Sublease agreement <u>remain unchanged</u> .	Ostatní ustanovení smlouvy o podnájmu <u>zůstávají nezměněna</u> .	Ostatní ustanovení smlouvy o podnájmu <u>zůstávají beze změny</u> .
Each of the Contracting parties shall receive one <u>copy</u> .	Každá ze smluvních stran obdrží jeden <u>výtisk</u> .	Každá ze smluvních stran obdrží jedno <u>vyhotovení</u> .
These are <u>often</u> peripheral, underdeveloped or marginalised areas- sometimes the borders are historical scars.	<u>Často</u> se jedná o okrajové, málo rozvinuté nebo marginalizované oblasti - někdy jsou hranice historickými jizvami.	<u>Mnohdy</u> se jedná o okrajové, málo rozvinuté nebo marginalizované oblasti - někdy jsou hranice historickými jizvami.
Generally, it is also necessary to aim for designing monitoring indicators both on the programme and project	Obecně je také nutné usilovat o navržení monitorovacích ukazatelů jak na úrovni programu, tak na úrovni	Obecně je také nutné usilovat o navržení monitorovacích ukazatelů jak na úrovni programu, tak i projektu,

level which would enable <u>more operative</u> verification of the achievement of set effects.	projektu, které by umožnily <u>operativnější</u> ověřování dosažení stanovených účinků.	které by umožnily <u>pružnější</u> ověřování dosažení stanovených účinků.
“185 million EU citizens, or 37% of the <u>population</u> , live in border areas.	„185 milionů občanů EU, neboli 37 % <u>obyvatel</u> , žije v pohraničních oblastech.	„185 milionů občanů EU, neboli 37 % <u>populace</u> , žije v pohraničních oblastech.

4. Skupina STYL

a. Neobratnost (v datasetu označeno zkratkou STYLneobr)

Definice: V navrhovaném překladu se vyskytují stylistické neobratnosti, jako jsou například chybějící prostředky kontextové deixe nebo nevyhovující užití trpného rodu, které sice nebrání porozumění textu, ale způsobují, že text nepůsobí plynule. Tyto stylistické úpravy někdy nebývají izolované. Mohou být doprovázeny i úpravami na ostatních úrovních (lexikální, morfologické, syntaktické) a v takových případech se výstupy z MT z většiny odstraní a vytvoří se nový překlad.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Colombian artist Oscar Murillo's work "surge (social cataracts) " depicts people staring out to the North Sea, but <u>with their view of the water and the European continent on the other side of it obscured by a dark canvas, with only a sliver of what lies beyond revealed through a slash in the canvas.</u>	Dílo kolumbijského umělce Oscara Murilla „surge (sociální katarakta) „ zobrazuje lidi hledící k Severnímu moři, ale <u>s pohledem na vodu a evropský kontinent na jeho druhé straně zakrytým tmavým plátnem, s pouhým střípkem toho, co se skrývá za ním,</u>	Dílo kolumbijského umělce Oscara Murilla „surge (social cataracts)“ (vzepětí – sociální vodopád) zobrazuje lidi hledící k Severnímu moři, ale jejich pohled na vodu a <u>evropský kontinent na druhé straně je zakrytý tmavým plátnem; jenom malý kousek prosvítá přes průřez v plátně.</u>

	<u>odhaleným lomítkem v plátně.</u>	
<u>China also had a major military conflict with Vietnam in 1979.</u> China claims sovereignty over all of South China Sea. Vietnam, Malaysia, the Philippines, Brunei and Taiwan have counter claims.	<u>Čína měla v roce 1979 velký vojenský konflikt s Vietnamem.</u> Čína je v celém Jihočínském moři. Vietnam, Malajsie, Filipíny, Brunej a Tchaj-wan mají protinároky.	<u>V roce 1979 také probíhal velký vojenský konflikt mezi Čínou a Vietnamem.</u> Čína tvrdí, že celé Jihočínské moře je jejím svrchovaným územím, ovšem Vietnam, Malajsie, Filipíny, Brunej i Tchaj-wan toto tvrzení rozporují.
Thomas Holoiien, an astronomer for the Carnegie Institution for Science who led the research, said: "This was really a combination of both being good and being lucky, and sometimes that's what you need to push the science forward."	Thomas Holoiien, astronom Carnegie Institution pro vědu, který vedl výzkum, řekl: <u>„To byla opravdu kombinace jak být dobrý, tak být šťastný, a někdy je to to, co musíte posunout vědu kupředu.“</u>	Thomas Holoiien, astronom organizace Carnegie Institution for Science, který výzkum vedl, k tomu řekl: <u>„V tomto případě to byla kombinace umu i štěstí, což někdy ve vědě potřebujete.“</u>
So far, the two countries held 21 rounds of Special Representatives talks <u>to resolve the border dispute.</u>	Obě země zatím uspořádaly 21 kol jednání zvláštních zástupců, <u>aby vyřešily spor o hranicích.</u>	Obě země zatím uspořádaly 21 kol jednání zvláštních zástupců, <u>na nichž se snaží vyřešit spor o hranici.</u>
Astronomers used a worldwide network of telescopes to detect the phenomenon before looking at TESS, whose permanent viewing zones designed to find distant planets, caught	Astronomové použili celosvětovou síť teleskopů k detekci tohoto <u>jevu předtím, než se podívali na TESS,</u> jejíž trvalé pozorovací zóny určené k nalezení vzdálených planet zachytily začátek násilné události.	Vědci použili k detekci tohoto jevu nejprve celosvětovou síť dalekohledů <u>a až následně se přesunuli k dalekohledu TESS,</u> jehož zóny stálého pozorování určené k odhalení vzdálených

the beginning of the violent event.		planet zachytily počátek incidentu.
Remarks and recommendations resulting from the audits carried out on both sides of the border are included in the separate audit reports of NKU and NIK on CD.	Poznámky a doporučení vyplývající z auditů <u>provedených na obou stranách hranice jsou obsaženy v samostatných zprávách o auditu NKU a NIK na CD.</u>	Poznámky a doporučení vyplývající z auditů, <u>kteřé byly provedeny v obou zemích, jsou součástí samostatných revizních zpráv NKU a NIK na CD.</u>

Zvláštním typem zásahu do koheze textu je případ, kdy posteditor (kromě lexikálních a syntaktických úprav v rámci segmentů) provedl úpravu členění textu provedeného systémem Memsources a propojil dva původně samostatné segmenty do jednoho:

This Supplement No. 1 is written and signed in 2 (in words: two) copies, each of which is valid for the original.	Tento dodatek Ne. 1 je napsán a podepsán ve 2 (slovy: dvou) kopiích, z nichž každá je platná pro originál.	Tento Dodatek č. 1 je vypracován ve 2 (slovy: dvou) vyhotoveních s platností originálu, z nichž každá smluvní strana obdrží jedno.
Each of the Contracting parties shall receive one copy.	<i>Z důvodu propojení segmentů v datasetu chybí výstup MT pro tento původně oddělený segment.</i>	

b. Stylistická neidiomaticita (v datasetu označeno zkratkou STYLneidiom)

Definice: V navrhovaném překladu jsou použity slovní obraty, které nejsou adekvátní vzhledem k úzu pro daný funkční styl nebo typ textu. Oproti lexikální neidiomaticitě se liší tím, že úprava za účelem vhodnějšího vyjádření vyžaduje více změn, než jen zvolit jiné synonymum. Někdy může být diskutabilní, zda konkrétní úprava byla provedena z důvodu úzu pro daný funkční styl či typ textu, nebo zda byla úprava motivována spíše snahou o větší plynulost a přirozenost cílového textu (v tomto případě bychom takovou úpravu řadili do kategorie *neobratnost*). Rozlišování mezi těmito dvěma kategoriemi vnímáme jako vysoce subjektivní.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Stars get sucked in when they venture too close to a supermassive black hole, which lives at the centre of most galaxies, including the Earth's Milky Way.	Hvězdy <u>se vcucnou</u> , když <u>se odváží</u> příliš blízko supermasivní černé díry, která žije ve středu většiny galaxií včetně zemské Mléčné dráhy.	<u>K</u> <u>pohlčení hvězdy</u> <u>dojde</u> , pokud <u>se dostane</u> příliš blízko obří černé díry, která se nachází ve středu většiny galaxií včetně naší Mléčné dráhy.
The funding is granted on <u>condition</u> that organisations from both sides of the border, such as regional authorities, universities and small and medium enterprises, <u>come together to implement</u> projects based on the needs of the border region.	Financování je poskytováno <u>pod podmínkou</u> , že organizace z obou stran hranice, jako jsou regionální orgány, univerzity a malé a střední podniky, <u>se sejdou</u> , <u>aby provedly</u> projekty založené na potřebách příhraničního regionu.	Financování se poskytuje <u>za podmínky</u> , že organizace z obou stran hranice, tedy například regionální orgány, univerzity a malé a střední podniky, <u>společně zrealizují</u> projekty založené na potřebách příhraničního regionu.
In the event that any provision of this Supplement No. 1 is or it becomes invalid or ineffective, this shall not affect the validity or	V případě, že některé ustanovení tohoto dodatku č. 1 <u>je neplatné nebo neúčinné</u> , <u>nemá to vliv na platnost</u> nebo	V případě, že některé ustanovení tohoto dodatku č. 1 <u>bude neplatným nebo neúčinným</u> , <u>nebude mít tato skutečnost dopad</u> na platnost

effectiveness of the other provisions of this Supplement No. 1.	účinnost ostatních ustanovení tohoto dodatku č. 1.	nebo účinnost ostatních ustanovení tohoto dodatku č. 1.
---	--	---

5. LOKÁLNÍ KONVENCE (v datasetu označeno zkratkou LOKAL)

Definice: Navrhovaný překlad neodpovídá lokálním konvencím přijímací kultury, ať už se jedná např. o zápis kalendářního data, lokalizaci místních názvů nebo o explicitaci (přidání vysvětlující informace) z důvodu odlišné kulturní presupozice.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Colombian artist Oscar Murillo's work " <u>surge (social cataracts)</u> " depicts people staring out to the North Sea, but with their view of the water and the European continent on the other side of it obscured by a dark canvas, with only a sliver of what lies beyond revealed through a slash in the canvas.	Dílo kolumbijského umělce Oscara Murilla „ <u>surge (sociální katarakta)</u> „ zobrazuje lidi hledící k Severnímu moři, ale s pohledem na vodu a evropský kontinent na jeho druhé straně zakrytým tmavým plátnem, s pouhým střípkem toho, co se skrývá za ním, odhaleným lomítkem v plátně.	Dílo kolumbijského umělce Oscara Murilla „ <u>surge (social cataracts)</u> “ (vzepětí – sociální <u>vodopád</u>) zobrazuje lidi hledící k Severnímu moři, ale jejich pohled na vodu a evropský kontinent na druhé straně je zakrytý tmavým plátnem; jenom malý kousek prosvítá přes průřez v plátně.
"China has always been dedicated to resolving territorial and maritime delimitation disputes through negotiation and consultation," stated an official white paper	„Čína se vždy věnovala řešení územních a námořních sporů prostřednictvím jednání a konzultací,“ uvedl oficiální bílý papír, který byl vydán, a to o čtyři dny před tím, aby	„Čína se vždy snažila řešit spory o pozemní i námořní hranice prostřednictvím jednání a konzultací,“ uvádí oficiální bílá kniha, která byla vydána 1. října tohoto

released, four days ahead of the country set to celebrate its 70th anniversary of the leadership of the ruling <u>Communist Party of China (CPC)</u> on October 1.	oslavil své 70. výročí vedení vládnoucí <u>komunistické strany Číny (CPC)</u> 1. října.	roku, tedy čtyři dny před zahájením oslav 70. výročí uchopení moci vládnoucí <u>Komunistickou stranou Číny.</u>
On the day, month and year written below Marta Burešová, <u>pers. no.</u> 695604/3017 Address: Radimova 8, Prague 6, 169 00 as the tenant on the one hand (Hereinafter referred to as "the tenant") and Karolína Černá, <u>pers. no.</u> 136205/891 Address: Alfrédova 13, Praha 4, 142 00 As a lessee on the other (Hereinafter referred to as "the lessee") collectively also referred to as "the Contracting parties" have agreed on this Supplement No. 1 to the Agreement on the sublease the apartment, of 13th May 2016 (hereinafter referred to as the "Supplement No. 1")	V den, měsíc a rok napsané níže Marta Burešová, <u>os. Ne.</u> 695604/3017 Adresa: Radimova 8, Praha 6, 169 00 jako nájemce na jedné straně (dále jen „nájemce“) a Karolína Černá, <u>os. Ne.</u> 136205/891 Adresa: Alfrédova 13, Praha 4, 142 00 Jako nájemce na straně druhé (dále jen „nájemce“) společně označované také jako „smluvní strany“ se dohodli na tomto dodatku č. 1 ke smlouvě o podnájmu bytu, ze dne 13. května 2016 (dále jen „dodatek č. 1“)	V den, měsíc a rok uvedené níže se Marta Burešová, <u>RČ</u> 695604/3017, bydliště: Radimova 8, Praha 6, 169 00 jako nájemce na jedné straně (dále jen „nájemce“) a Karolína Černá, <u>RČ</u> 136205/891, bydliště: Alfrédova 13, Praha 4, 142 00 jako podnájemce na straně druhé (dále jen „podnájemce“), společně označované také jako „smluvní strany“, dohodly na tomto dodatku č. 1 ke smlouvě o podnájmu bytu ze dne 13. května 2016 (dále jen „dodatek č. 1“)
Why The Times Published Details of the Whistle-Blower's Identity	Proč The Times zveřejnila podrobnosti o identitě Whistle-Blower	Proč <u>deník</u> The Times zveřejnil podrobnosti o identitě informátora
Remarks and recommendations resulting	Poznámky a doporučení vyplývající z auditů	Poznámky a doporučení vyplývající z auditů

from the audits carried out on both sides of the border are included in the separate audit reports of <u>NKU</u> and NIK on CD.	provedených na obou stranách hranice jsou obsaženy v samostatných auditních zprávách <u>NKU</u> a NIK na CD.	provedených na obou stranách hranice jsou obsaženy v samostatných auditních zprávách <u>NKÚ</u> a NIK na CD.
Prague, <u>31st December 2018</u> .	V Praze dne <u>31. prosince 2018</u> .	V Praze dne <u>31.12. 2018</u>
Named after 19th-century landscape painter JMW Turner, the award made stars of Grayson Perry, Damien Hirst and <u>12 Years a Slave</u> director Steve McQueen.	Ocenění pojmenované po krajináři 19. století JMW Turnerovi udělalo hvězdy Graysona Perryho, Damiena Hirsta a režiséra filmu <u>12 let otroka</u> Steva McQueena.	Toto ocenění získalo svůj název po krajináři J. M. W. Turnerovi, který žil v 19. století, a dostalo do povědomí lidí talenty jako Grayson Perry, Damien Hirst nebo Steve McQueen, režisér filmu <u>12 let v řetězech</u> .
The <u>Sino-India</u> border dispute involving 3,488-km-long <u>Line of Actual Control (LAC)</u> remained unresolved. China also claims <u>Arunachal Pradesh</u> as part of <u>South Tibet</u> , which India contests.	<u>Sino-indický</u> hraniční spor, který se týká 3 488 km <u>Řád skutečného řízení (LAC)</u> zůstal nevyřešen. Čína také tvrdí, že <u>Arunachal Pradesh</u> je součástí <u>jižního Tibetu</u> , který Indie napadá.	<u>Čínsko-indický</u> spor o hranici s 3 488 km dlouhou <u>Liníí aktuální kontroly (LAC)</u> se dosud nevyřešil. Čína si rovněž pod názvem <u>Jižní Tibet</u> nárokuje <u>Arunáčalpraděš</u> , který je pod správní nadvládou Indie.

6. CHYBNÁ ÚPRAVA (v datasetu označeno zkratkou CHYB)

Definice: Úpravy ze stran posteditora, které způsobily chyby v cílovém textu, ale ve výstupu MT původně nebyly, nebo úpravy, které byly provedeny posteditorem za účelem eliminovat existující chybu, ale nevedly ke správnému řešení.

Příklady:

<i>Zdrojový text</i>	<i>Výstup MT systému nebo referenčního překladu</i>	<i>Výstup posteditora</i>
Named after 19th-century landscape painter JMW Turner, the award made stars of Grayson Perry, Damien Hirst and <u>12 Years a Slave</u> director Steve McQueen.	Tato cena byla pojmenována po krajináři z 19. století JMW Turnerovi a stala se hvězdami filmů Grayson Perry, Damien Hirst a <u>12 let otrokem režiséra Steva McQueena</u> .	Tato cena byla pojmenována po krajináři z 19. století J.M.W. Turnerovi a proslavila Graysona Perryho, Damiena Hirsta a <u>film 12 let v otroctví</u> režiséra Steva McQueena.
It was finally diplomatically resolved <u>after which</u> both sides pulled back their troops.	Nakonec se to podařilo, <u>když</u> obě strany stáhly své vojáky.	Spor byl diplomaticky vyřešen <u>poté, co</u> obě strany stáhly z území své vojáky.
On 13th May 2016, the tenant and the lessee closed the Agreement on the sublease of the apartment, under which the tenant let the lessee use the apartment No. 4 (area 49 m ²) of size <u>1+1/L</u> in the ground floor of the house in Prague 4, Alfrédova 13, Post Code: 142 00, Kamýk (hereinafter referred to as the “apartment in question” and “Sublease agreement”), including the equipment and	Dne 13. května 2016 nájemce a nájemce uzavřeli Smlouvu o podnájmu bytu, na jejímž základě nájemce pronajal nájemci byt č. 4 (plocha 49 m ²) o velikosti <u>1 + 1 / L</u> v přízemí domu dům v Praze 4, Alfrédova 13, PSČ: 142 00, Kamýk (dále jen „předmětný byt“ a „smlouva o podnájmu“), včetně vybavení a vybavení uvedeného ve smlouvě o podnájmu.	Dne 13. května 2016 nájemce a podnájemce uzavřeli Smlouvu o podnájmu bytu, na jejímž základě nájemce dává podnájemci k užití byt č. 4 (plocha 49 m ²) o velikosti <u>1+1</u> v přízemí domu v Praze 4, Alfrédova 13, PSČ: 142 00, Kamýk (dále jen „předmětný byt“ a „smlouva o podnájmu“), včetně vybavení a zařízení uvedeného ve smlouvě o podnájmu.

amenities as specified in the Sublease agreement.		
The black hole's impressive gravitational forces tear the star apart, with some of its material thrown into space and the rest into the black hole, forming a disc of hot, bright gas as it is swallowed.	Působivé gravitační síly černé díry trhají hvězdu od sebe, přičemž <u>část</u> jejího materiálu byla <u>vyhozena</u> do prostoru a zbytek do <u>černé díry</u> a při pohlcování tvoří disk horkého jasného plynu.	Působivá gravitační síla černé díry roztrhá hvězdu na kusy, přičemž <u>část</u> jejího materiálu je <u>vyvržen</u> do okolního vesmíru a zbytek pohltí <u>černá díry</u> . Kolem ní se při tom tvoří disk horkého zářícího plynu.
Colombian artist Oscar Murillo's work "surge (social cataracts) " depicts people staring out to the North Sea, but with their view of the water and the European continent on the other side of it <u>obscured</u> by a dark canvas, with only a sliver of what lies beyond revealed through a slash in the canvas.	Kolumbijský umělec Oscar Murillo v díle „Přepětí (sociální katarakty)“ zobrazuje lidské postavy hledící ven na Severní moře, ale s pohledem na vodu a evropský kontinent na druhé straně moře <u>zakryté</u> tmavým plátnem, jen s malým náznakem toho, co leží za poodkrytým pruhem plátna.	Kolumbijský umělec Oscar Murillo v díle „Přepětí (sociální katarakty)“ zobrazuje lidské postavy hledící ven na Severní moře, ale s pohledem na vodu a evropský kontinent na druhé straně moře <u>zakrytým</u> tmavým plátnem, jen s malým náznakem toho, co leží za poodkrytým pruhem plátna.
<u>RECOMMENDATIONS</u>	<u>DOPORUČENÍ</u>	<u>RECOMMENDATIONS</u>
Try to access the <u>Play Store</u> and install the application.	Zkuste se dostat do <u>Obchodu Play</u> a aplikaci nainstalovat.	Zkuste se dostat do <u>aplikace Play Store</u> a tuto aplikaci nainstalovat.
It was finally diplomatically resolved <u>after which</u> both sides pulled back their troops.	Nakonec bylo vše vyřešeno diplomatickou cestou <u>a</u> obě strany stáhly své vojenské jednotky.	Nakonec bylo vše vyřešeno diplomatickou cestou <u>poté</u> , <u>kdy</u> obě strany stáhly své vojenské jednotky.

2.2.3 Kategorie úprav ve fázi revize

Při kategorizaci úprav provedených ve fázi revize jsme postupovali stejně jako ve fázi posteditace. Všechny kategorie úprav definované během analýzy fáze posteditace byly zastoupeny i ve fázi revize. Navíc byly identifikovány dva nové typy úprav, které ve fázi posteditace nebyly zastoupeny:

1. FORMÁTOVÁNÍ (v datasetu označeno zkratkou FORMAT)

Definice: Úpravy ve formátování textu, jako je např. tučné písmo, kurzíva, podtržení, barva nebo font písma.

Příklad:

<i>Zdrojový text</i>	<i>Výstup posteditora</i>	<i>Výstup revizora</i>
Named after 19th-century landscape painter JMW Turner, the award made stars of Grayson Perry, Damien Hirst and <u>12 Years a Slave</u> director Steve McQueen.	Tato cena pojmenovaná po krajináři 19. století J. M. W. Turnerovi zahájila hvězdnou kariéru Graysona Perryho, Damiena Hirsta a režiséra Steva McQueena, který je autorem filmu <u>12 let v řetězech</u> .	Tato cena pojmenovaná po krajináři 19. století J. M. W. Turnerovi zahájila hvězdnou kariéru Graysona Perryho, Damiena Hirsta a režiséra Steva McQueena, který je autorem filmu <i><u>12 let v řetězech</u></i> .

V příkladu vidíme, že revizor změnil formátování u názvu filmu. Oproti zdrojovému i posteditovanému textu ho zvýraznil kurzívou, což je v datasetu vyznačeno pomocí značek {i>; <i}.

2. Byl identifikován nový podtyp kategorie CHYBNÉ ÚPRAVY, a sice úprava, kdy revizor nechal v cílovém jazyce dvě různé verze překladu, zřejmě z nepozornosti.

<i>Zdrojový text</i>	<i>Výstup posteditora</i>	<i>Výstup revizora</i>
Try to access the Play Store and install the application.	Zkuste se dostat do aplikace Play Store a tuto aplikaci nainstalovat.	Zkuste se dostat do aplikace Play Store a tuto aplikaci nainstalovat. / Pokuste se nalézt aplikaci Play Store a nainstalovat ji.

Ani jednu z těchto kategorií, které byly definovány pouze ve fázi revize, nepovažujeme za specifikum revizní činnosti. V obou případech se v analyzovaném datasetu jedná o ojedinělý výskyt; jedná se tedy spíše o důsledek rozhodnutí konkrétních revizorů. Úpravy ve formátování i podtyp chybné úpravy, při které jedinec ponechá v cílovém textu více variant překladu, by se mohly vyskytnout i ve fázi posteditaci. Zároveň žádná kategorie zastoupena ve fázi posteditace nebyla kompletně eliminována ve fázi revize. Nebyly tedy zjištěny žádné rozdíly v rozmanitosti kategorií úprav prováděnými ve fázi posteditace a ve fázi revize.

2.3 Diskuze

Hodnota tohoto výzkumu spočívá v tom, že na posteditační činnost nebylo nahlíženo z pohledu kvality použitého MT systému. Cílem nebylo identifikovat chyby, kterých se dopouštěly MT systémy. V centru zájmu byl způsob, jakým s výstupy z MT systémů pracovali posteditoři a dále revizoři. Analyzováno bylo velké množství dat složené z různých zdrojových textů, které byly přeložené různými MT systémy a dále ve dvou fázích upraveny posteditory a revizory s různě dlouhými profesními zkušenostmi. Výsledkem analýzy dat s tolika proměnnými tedy nemohou být kauzality typu „pokud překládáme MT systémem X , posteditoři/revizoři budou vykonávat úpravy kategorie Y “ nebo „posteditor/revizor X má tendence se uchýlovat k úpravám kategorie Y “. Namísto toho byly vytvořeny kategorie lingvistických úprav, které jsou reprezentativní pro posteditační a revizní činnost ve směru z angličtiny do češtiny obecně – na základě výsledků analýzy při překladu 8 typů textů přeloženými strojovými překladači rozmanité kvality a upravenými různorodou skupinou posteditorů a revizorů. Takový výčet kategorií může tvořit základ pro výuku MTPE, při které se budou studenti učit zaměřovat se na

určitý typ úprav. Pro práci se strojovým překladem je vždy důležité znát požadavky od zadavatele a umět podle nich pracovat. Pokud zadavatel žádá např. *light MTPE* (viz kapitola 1.2.1), nemusí se dle definice tohoto typu posteditace posteditor zabývat stylistickými odchylkami a formátováním, ale měl by se zaměřit výhradně na chybně převedený význam nebo nesrozumitelné části. Pokud by posteditor neuměl oddělit nezbytné úpravy od úprav preferenčních, mohlo by se stát, že pro něj MTPE nebude efektivní, a v důsledku ani ekonomicky výhodná.

Uvědomujeme si, že způsob kategorizace, definice jednotlivých kategorií i zařazování jednotlivých úprav do lingvistických kategorií je do jisté míry subjektivní, a proto by se pravděpodobně výstupy jiných badatelů při práci se stejným datasetem lišily. Jazyk je živý, proměnlivý a každý jedinec ho pravděpodobně vnímá a užívá částečně rozdílným způsobem. I překlad je do určité míry subjektivní záležitost; ve většině případů neexistuje jen jedno jediné správné překladatelské řešení a ne vždy se více jedinců shodne na jednom nejlepším řešení. Stejně tak to bude pravděpodobně také s posteditační činností jako takovou i s hodnocením a popisováním vykonaných úprav během posteditace, zejména u kategorií, které např. nepodléhají pevně stanoveným pravidlům, jako jsou gramatická pravidla nebo pravidla pravopisu. Co pro jednoho může být úprava na základě úzu pro daný typ textu, může být druhým vnímáno jako úprava preferenční. Jelikož v tomto výzkumu bylo na dataset nahlíženo pouze pohledem řešitelky této práce, je zde značný prostor pro subjektivitu.

Pokud by se měl podobný výzkum opakovat s cílem eliminovat subjektivní pohled konkrétního badatele, doporučovali bychom sbírat data o motivaci posteditorů pro jednotlivé úpravy už během posteditační činnosti a následně bychom navrhovali nechat vykonané úpravy ohodnotit dalšími osobami. Z takových dat by se dalo lépe zobecnit chápání úprav na základě odůvodnění každé úpravy. Dále by bylo zajímavé ohodnotit vnímání jednotlivých úprav z hlediska nezbytnosti, čili zda byla konkrétní úprava provedena z důvodu naprosto nevhodného překladu vzhledem k účelu textu v přijímací kultuře, nebo zda by byl navrhovaný překlad v cílové kultuře přijatelný i bez zásahu člověka. Pokud by se ukázalo, že existuje souvislost mezi určitými kategoriemi úprav a jejich vnímání jako nezbytné či spíše preferenční, mohl by takový výzkum přispět k zadávání přesnějších instrukcí a požadavků směrem od zadavatele k posteditorovi.

ZÁVĚR

Tato práce se zaměřila na činnost posteditace strojového překladu v jazykové kombinaci angličtina – čeština. Cílem bylo vytvořit výčet kategorií úprav, ke kterým při této činnosti dochází, s cílem poskytnout základní přehled pro případné kurzy posteditace strojového překladu.

Posteditace strojového překladu je činnost, jejíž průběh se odvíjí od několika faktorů. Teoretická část této práce se proto zabývá aspekty, které jsou pro posteditaci strojového překladu podstatné. Nejprve jsme pojednali o tématu překladových technologií z hlediska historického vývoje, který je důležitý pro celkové pochopení problematiky práce se strojovým překladem. Dále jsme se zaměřili nejen na nástroje CAT, ale hlavně také na technologie strojového překladu a strojové překladače. Zvláštní důraz byl kladen na strojové překladače specializované na překládání v kombinacích s českým jazykem. Na závěr kapitoly o překladových technologiích jsme uvedli několik současných trendů v oboru strojového překladu, které by mohly v blízké budoucnosti ovlivnit směr technologického vývoje a potažmo i způsob práce překladatelů a posteditorů.

V dalších kapitolách teoretické části jsme se zaměřili na činnost posteditace. Uvedli jsme základní pojmy a definice, které s činností posteditace souvisejí. Tím jsme představili kontext problematiky rozdílů mezi činností překládání a posteditace, mezi požadavky na překladatele a na posteditory, ale také jsme představili současný stav výuky posteditace v Česku i v zahraničí.

Poskytli jsme podrobný přehled kurzů, které se na českých univerzitách a vysokých školách posteditaci strojového překladu věnují nebo s ní alespoň souvisí. Po tomto průzkumu současné situace jsme konstatovali, že v současné době v České republice neprobíhá samostatný kurz věnovaný posteditaci strojového překladu, byť se do budoucna alespoň v některých studijních programech plánuje zařadit. V zahraničí už takové kurzy v minulosti proběhly a na základně nabytých zkušeností vznikly akademické články. V teoretické části nabízíme přehled klíčových poznatků z těchto článků. Nejdůležitějším aspektem při výuce posteditace se ukázal být důraz na praktický nácvik. V závěru teoretické práce jsme tedy potvrdili přesvědčení, že by ani v českých studijních programech neměly chybět kurzy věnované výhradně posteditační činnosti.

V praktické části jsme analyzovali texty, které byly přeložené 13 různými strojovými překladači a následně posteditované profesionálními překladateli a posteditory. Cílem takové analýzy bylo definovat lingvistické kategorie úprav, ke kterým během posteditační činnosti

dochází. Navíc jsme dále analyzovali ještě úpravy, které prováděla další skupina profesionálních překladatelů v následné fázi revize těchto posteditací. Celkově jsme tedy porovnávali čtyři verze jednoho textu – zdrojový text v angličtině, výstup ze strojového překladače v češtině, posteditovaný výstup a revidovanou posteditaci. Po této analýze vznikl výčet kategorií lingvistických úprav, které posteditoři dělají při práci se strojovým překladem ve směru z anglického do českého jazyka. Součástí výsledku bylo také zjištění, že kategorie úprav se ve fázi posteditace a v následné fázi revize neliší.

V této práci byly doloženy důvody pro potřebnost samostatných kurzů posteditace strojového překladu na českých vysokých školách a univerzitách. Zároveň je výsledkem této práce výčet lingvistických kategorií úprav prováděných při posteditaci strojového překladu z angličtiny do češtiny. Věříme, že tento výčet může být nápomocen při snaze posteditační činnosti v této jazykové kombinaci systematicky vyučovat, a připravovat tak absolventy translatologických studijních programů na současný trh, kde dovednost posteditovat strojový překlad hojně využijí.

BIBLIOGRAFIE

ABSOLON, J. (2018). *Strojový preklad a posteditovanie*. [Dizertačná práca, Univerzita Konštantína Filozofa v Nitre] [online] [cit. 26.7.2021] Dostupné z:

<http://opac.crzp.sk/?fn=detailBiblioForm&sid=B91108598CC4B84B242519F148AB&seo=CRZP-detail-kniha>

ARENAS, G. A. (2020). „Pre-editing and post-editing“. *The Bloomsbury Companion to Language Industry Studies*. [online] [cit. 2.7.2021] Dostupné z:

https://www.researchgate.net/publication/338310852_Pre-editing_and_post-editing

ARENAS, G.A., MOORKENS, J. (2019). „Machine Translation and Post-editing Training as Part of a Master's Programme“. *The Journal of Specialised Translation*. 2019 (31), 217-238.

BAR-HILLEL, Y. (1960). „The Present State of Automatic Translation of Languages“. *Advances in Computers*, Vol. 1, New York: Academic Press.

„Best Machine Translation Software“. *G2*. [online] [cit. 18.6.2021] Dostupné z:

<https://www.g2.com/categories/machine-translation>

BOJAR, O. (2012). *Čeština a strojový preklad: Strojový preklad našincům, našiinci strojovému překlada*. Ústav formální a aplikované lingvistiky.

BOWKER, L. (2015). „Computer-aided translation: translator training“. In *Routledge Encyclopedia of Translation Technology* (s. 88-104). Routledge.

DEPALMA, A. D. (2017). „Augmented Translation Powers up Language Services“. *CSA Research Blog*. [online] 15.2.2017. [cit.20.6.2021] Dostupné z: <https://csa-research.com/Blog/ArticleID/140>

DEPALMA, A. D. (2021). „Augmenting Human Translator Performance“. *CSA Research Blog*. [online] 10.2.2021. [cit.20.6.2021] Dostupné z: <https://csa-research.com/Blog/Augmenting-Human-Translator-Performance>

EUATC. (2020). „European Language Industry Survey: 2020: Before & After Covid-19“. [online] [cit. 15.6.2021] Dostupné z:

https://ec.europa.eu/info/sites/default/files/2020_language_industry_survey_report.pdf

European Master's in Translation. (2017). *EMT Competence Framework – 2017*. Evropská komise. [online] [cit. 5.7.2021] Dostupné z:

https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf

„G2 Research Scoring Methodologies“. *G2*. [online] [cit. 18.6.2021] Dostupné z:

<https://research.g2.com/g2-scoring-methodologies>

GARCIA, I. (2015). „Computer-aided translation: systems“. In *Routledge Encyclopedia of Translation Technology* (s. 3-31). Routledge.

GENE, V. (2020). „Changing Roles: From Translation to Post-Editing“. *Intertranslations.com*. [online] [cit. 02.04.2021] Dostupné z: <https://www.intertranslations.com/downloads/post-editing-whitepaper/>

HAIJČ, J., PANEVOVÁ, J., HLADKÁ, B. (2012). „Lingvistika na Matematicko-fyzikální fakultě?“. *Vesmír*. 2012(9). [online] [cit. 14.6.2021] Dostupné z: <https://vesmir.cz/cz/casopis/archiv-casopisu/2012/cislo-9/lingvistika-matematicko-fyzikalni-fakulte.html>

HOUSER, P. (2020). „Cubbitt: Automatický překladač, který už opravdu nahradí překladatele“. In: *SCIENCEmag.cz* [online]. 26.10.2020 [cit. 19.6.2021]. Dostupné z: <https://sciencemag.cz/cubbitt-automaticky-prekladac-ktery-uz-opravdu-nahradi-prekladatele/>

HRÁCH, O. (2020). *Vyhledky překladatelské profese v éře moderních technologií: interdisciplinární pohled*. [Diplomová práce, Univerzita Karlova] [online] [cit. 15.6.2021] Dostupné z: <https://dspace.cuni.cz/handle/20.500.11956/122864>

CHAN, S. (2015). „The development of translation technology: 1967 – 2013“. In *Routledge Encyclopedia of Translation Technology* (s. 3-31). Routledge.

CHLOUPEK, J. a kol. (1991). *Stylistika češtiny*. Praha: Státní pedagogické nakladatelství.

„Introducing Memsource Translate“. *MultiLingual*, 2020. MultiLingual Computing, Inc., 31 (5). ISSN 1523-0309. [online] [cit. 20.6.2021]. Dostupné z: <https://multilingual.com/issues/sept-oct-2020/>

ISO, *ISO 17100:2015: Translation services – Requirements for translation services*. Dostupné z: http://www.iso.org/iso/catalogue_detail.htm?csnumber=59149

JON, J. (2019). *Exploring Contextual Information in Neural Machine Translation* [online]. [Diplomová práce. Vysoké učení technické v Brně.] [online] [cit. 13.7.2021]. Dostupné z: <http://hdl.handle.net/11012/180410>

KASÍK, P. (2017). „Google skokově vylepšil překlad. Umělá inteligence se zakousla do češtiny“. *iDnes*. [online] [cit. 15.6.2021] Dostupné z: https://www.idnes.cz/technet/internet/google-translate-neural-networks.A170418_224649_sw_internet_pk_a

KASÍK, P. (2021). „,Nečekal jsem, že se toho dožiju!“ Profíci žasnou, jak překládá stroj“. *Seznam Zprávy*. [online] [cit. 18.6.2021] Dostupné z: <https://www.seznamzpravy.cz/clanek/necekal-jsem-ze-se-toho-do-ziju-prekladatele-zasnou-co-umi-stroj-156128>

KENNY, D. (2020). „Technology in Translator Training“. In *The Routledge Handbook of Translation Technology* (s. 498-515). Routledge.

- KLABAL, O. (2014). *Peter Newmark a jeho přínos translatologii* [Diplomová práce, Univerzita Karlova] [online] [cit. 14.6.2021] Dostupné z: https://dspace.cuni.cz/bitstream/handle/20.500.11956/71213/DPTX_2012_2_11210_0_343250_0_126247.pdf?sequence=1
- KOPONEN, M. (2015). *How to teach machine translation post-editing? Experiences from a post-editing course*. University of Eastern Finland. [online] [cit. 15.6.2021] Dostupné z: <https://www.researchgate.net/publication/299346656>
- LARSONNEUR, C. (2021) „Neural Machine Translation: From Commodity to Commons?“. *In: When Translation Goes Digital* (s. 257-280). Palgrave Macmillan. [online] [cit. 20.6.2021] Dostupné z: <https://doi.org/10.1007/978-3-030-51761-8>
- LEVÝ, J., ČERVENKA, M. (1971), ed. *Bude literární věda exaktní vědou?: výbor studií*. Praha: Československý spisovatel.
- LEVÝ, J. (2012). *Umění překladačů*. Praha: Miroslav Pošta – Apostrof.
- MARKING, M. „Survey Reveals Impact of Covid-19 on Europe’s Language Industry“. *Slator*. [online] [cit. 11.05.2021] Dostupné z: <https://slator.com/industry-news/survey-reveals-impact-of-covid-19-on-europes-language-industry/>
- MIHALACHE, I. (2021) „Human and Non-Human Crossover: Translators Partnering with Digital Tools“. *When Translation Goes Digital* (s. 19-44). Palgrave Macmillan. [online] [cit. 20.6.2021] Dostupné z: <https://doi.org/10.1007/978-3-030-51761-8>
- NOVÁK, O. (2021). „Český překladač CUBBITT dokáže v přesnosti překonat člověka. Podívejte se na srovnání s konkurencí“. *Lupa.cz* [online] 22.6.2021 [cit. 23.6.2021] Dostupné z: <https://www.lupa.cz/clanky/cesky-prekladac-cubbitt-dokaze-v-presnosti-prekonat-cloveka-podivejte-se-na-srovnani-s-konkurenci/>
- „Nový překladač Lingea“. *Lingea*. [online] [cit.18.6.2021] Dostupné z: <https://www.lingea.cz/aktuality/novy-prekladac-lingea>
- O'BRIEN, S. (2002). „Teaching Post-Editing: A Proposal for Course Content“. *6th EAMT Workshop Teaching Machine Translation*. [online] [cit. 15.6.2021] Dostupné z: https://www.academia.edu/1160352/Teaching_Post_Editing_A_Proposal_for_Course_Content
- POPEL, M. (2020). *CUNI English-Czech and English-Polish Systems in WMT20: Robust Document-Level Training*. [online] [cit. 23.7.2021]. Dostupné z: <https://aclanthology.org/2020.wmt-1.28.pdf>
- POPEL, M., TOMKOVÁ, M., TOMEK, J. *et al.* (2020). „Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals“. *Nature Communications*. 11 (4381) [online] [cit. 18.6.2021]. Dostupné z: <https://rdcu.be/cmQdK>

„Překladač z Matfyzu dohání v kvalitě běžné překladače“. In: *SCIENCEmag.cz* [online]. 17.9.2020 [cit. 19.6.2021]. Dostupné z: <https://sciencemag.cz/prekladac-z-matfyzu-dohani-v-kvalite-bezne-prekladatele/>

QIAN, D. (2013). „The Technological Turn in Translation Studies“. *Foreign Languages and Translation*, 2013(3), 39–45. [online] [cit. 13.6.2021] Dostupné z: https://www.academia.edu/14821174/The_Technological_Turn_in_Translation_Studies

QUN, L., XIAOJUN, Z. (2015). „Machine translation: general“. In *Routledge Encyclopedia of Translation Technology* (s. 105-119). Routledge.

RIVERA-TRIGUEROS, I. (2021). *Machine translation systems and quality assessment: a systematic review*. Lang Resources & Evaluation. [online] [cit. 20.6.2021] Dostupné z: <https://doi.org/10.1007/s10579-021-09537-5>

ROTHWELL, A., SVOBODA, T. (2019). „Tracking Translator Training in Tools and Technologies: Findings of the EMT Survey 2017“. *The Journal of Specialised Translation*. 32, s. 26-60.

SHANNON, C., E., WEAVER, W. (1964). *The Mathematical Theory of Communication*. The University of Illinois Press. Urbana.

SNELL-HORNBY, M. (2006). *The Turns of Translation Studies: New paradigms or shifting viewpoints?*. Amsterdam: John Benjamins Publishing Company. Benjamins translation library.

SVOBODA, T. (2012). *Kapitoly z překladačské praxe: Odborný překlad mezi němčinou a češtinou*. Univerzita Karlova, Filozofická fakulta.

ŠANCA, F. (2018). *The Use of CAT Tools in University Translation Courses: A Case Study Based on Teaching with Memsource*. [Diplomová práce, Univerzita Karlova] [online] [cit. 6.7.2021] Dostupné z: <http://hdl.handle.net/20.500.11956/100244>

TAUS eLearning Platform, *Post-Editing/Reviewing*. [elektronický kurz] [online] [cit. 30.6.2021] Dostupné z: <https://elearning.taus.net/course/index.php#pe-overview>

TAMCHYNA, A. (2021). „MT Metrics Explained: Feeling BLEU?“ *Medium*. [online] [cit. 30.6.2021] Dostupné z: <https://medium.com/memsource-engineering/mt-metrics-explained-feeling-bleu-f62b3ab4500>

„What is machine translation?“. *Microsoft*. [online] [cit. 15.6.2021] Dostupné z: <https://www.microsoft.com/en-us/translator/business/machine-translation/>

ZEHNALOVÁ, J. (2015). *Kvalita a hodnocení překladu: Modely a aplikace*. Olomouc. Univerzita Palackého v Olomouci.

ZEHNALOVÁ, J., MOLNÁR, O., KUBÁNEK, M. (2012). *Teaching Translation and Interpreting Skills in the 21st Century*. Olomouc. Palacký University in Olomouc.

PŘÍLOHA

KOPECKA, K. (2021): *Kopecka_Kategorizace_2021_Priloha.xlsx*. figshare. Dataset. Dostupné z: <https://doi.org/10.6084/m9.figshare.15087894.v1>