

**Filozofická fakulta Univerzity Palackého**

**Evaluace a srovnání nástrojů pro strojový překlad**

(Bakalářská práce)

**2014**

**Veronika Běhálková**

**Filozofická fakulta Univerzity Palackého**

Katedra anglistiky a amerikanistiky

**EVALUACE A SROVNÁNÍ NÁSTROJŮ  
PRO STROJOVÝ PŘEKLAD**

**EVALUATION AND COMPARISON  
OF MACHINE TRANSLATION TOOLS**

(Bakalářská práce)

Autor: Veronika Běhálková

Studijní obor: Angličtina se zaměřením na komunitní tlumočení a překlad

Vedoucí práce: PhDr. Pavel Král

Olomouc 2014

*Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a uvedla úplný seznam citované a použité literatury.*

*V Olomouci dne 6. 5. 2014*

.....

*Děkuji PhDr. Pavlu Královi za odborné vedení a cenné rady při psaní této bakalářské práce.*

## SEZNAM POUŽITÝCH ZKRATEK

MT – machine translation (strojový překlad)

SMT – statistical machine translation (statistický strojový překlad)

ZJ – zdrojový jazyk

CJ – cílový jazyk

ZT – zdrojový text

CT – cílový text

V tabulkách:

OT – originální text

GMT – Google machine translation (strojový překlad Googlu Translate)

BMT – Bing machine translation (strojový překlad Bing Translatoru)

# OBSAH

ÚVOD.....	7
<b>OBECNÝ TEORETICKÝ PŘEHLED .....</b>	<b>9</b>
<b>1 VZNIK A VÝVOJ STROJOVÉHO PŘEKLADU .....</b>	<b>9</b>
<b>2 STROJOVÝ PŘEKLAD ZALOŽENÝ NA LINGVISTICKÝCH PRAVIDLECH...</b>	<b>11</b>
2.1 MODEL PŘÍMÉHO PŘEKLADU (DIRECT APPROACH) .....	11
2.2 MODEL NEPŘÍMÉHO PŘEKLADU (INDIRECT APPROACH).....	12
2.2.1 <i>Interlingvální přístup (interlingua approach)</i> .....	12
2.2.2 <i>Transferový přístup (transfer approach)</i> .....	13
<b>3 STATISTICKÝ STROJOVÝ PŘEKLAD .....</b>	<b>15</b>
3.1 PŘÍSTUP ZALOŽENÝ NA KORPUSECH (CORPUS-BASED APPROACH).....	15
3.1.1 <i>Statistické výpočty</i> .....	15
3.1.1.1 Model překladu.....	15
3.1.1.2 Model jazyka .....	17
3.2 HYBRIDNÍ STROJOVÝ PŘEKLAD .....	17
<b>4 BEZPLATNÉ ONLINE PŘEKLADAČE .....</b>	<b>18</b>
4.1 HISTORIE A VÝVOJ ONLINE PŘEKLADAČŮ .....	18
4.2 MICROSOFT BING TRANSLATOR.....	19
4.3 GOOGLE TRANSLATE.....	19
<b>PŘÍPADOVÁ STUDIE.....</b>	<b>21</b>
<b>5 METODOLOGIE.....</b>	<b>21</b>
<b>6 EVALUACE STROJOVÉHO PŘEKLADU .....</b>	<b>25</b>
6.1 PŘEHLED A ANALÝZA JEVŮ, K NIMŽ U MT DOCHÁZÍ.....	25
6.1.1 <i>Rovina gramatická na úrovni slova (morfologie)</i> .....	25
6.1.2 <i>Rovina gramatická na úrovni věty (syntax)</i> .....	26
6.1.3 <i>Rovina lexikální</i> .....	29
6.1.4 <i>Rovina sémantická</i> .....	32
6.1.5 <i>Rovina textová</i> .....	36
6.1.6 <i>Rovina pragmatická</i> .....	37
6.2 VÝSLEDKY ANALÝZY.....	40

6.2.1	<i>Výsledky analýzy z hlediska cílového čtenáře .....</i>	40
6.2.2	<i>Výsledky analýzy z hlediska profesionálního překladatele.....</i>	42
<b>DISKUZE.....</b>		<b>45</b>
<b>ZÁVĚR.....</b>		<b>46</b>
<b>SUMMARY .....</b>		<b>49</b>
<b>BIBLIOGRAFIE.....</b>		<b>51</b>
<b>ANOTACE.....</b>		<b>53</b>
<b>ABSTRACT .....</b>		<b>54</b>

## **EXTERNÍ PŘÍLOHY NA CD**

### **Příloha č. 1 – 10 pokusných textů přeložených Googlem Translate**

### **Příloha č. 2 – 10 pokusných textů přeložených MS Bing Translatorem**

Pozn.: Všechny použité ukázky strojového překladu jsou součástí textu této práce. Pro vyhledání konkrétních vět v přílohách na CD (více než 300 stran) doporučuji příkaz Najít (Ctrl+F).

# ÚVOD

Během 70 let své existence prošel strojový překlad rapidním vývojem. Od systémů založených na lingvistických pravidlech se poměrně zlomově přesunul ke statistické analýze již přeloženého materiálu, přičemž současný, moderní strojový překlad tyto i jiné metody kombinuje.

Se stoupající kvalitou automatického překladu se změnilo a rozšířilo i jeho využití. Od výzkumných projektů na univerzitách přes zakázky pro stát a veřejné instituce si strojový překlad našel cestu na trh a do firem, a posléze s všeobecným rozšířením internetu a přístupu k němu prakticky ke každému, kdo o něj má zájem, a to zdarma.

V současné době se výstup strojového překladu dá označit za natolik kvalitní, aby pomáhal i profesionálním překladatelům. Strojový překlad má potenciál ušetřit překladateli určitou část času, práce i energie v běžné praxi. Avšak aby se MT (tj. machine translation, strojový překlad) stal překladatelovým užitečným pomocníkem a ne obtěžujícím a zdržovacím faktorem, je nasnadě, že před tím, než ho překladatel začne používat, by měl vědět, co od něj může očekávat. Na tom se odborníci shodují. Například Gaspari (2007) přiznává, že hlavním problémem co se týče využití strojového překladu online je fakt, že většina uživatelů nemá povědomí o tom, kde leží jeho limity. Zdůrazňují to i Church a Hovy (1993), kteří uvádějí, že by uživatelům mělo být jasné, co takový systém umí a neumí.

Cílem této bakalářské práce je zjistit, co může moderní strojový překlad profesionálnímu překladateli nabídnout, a to prostřednictvím případové studie.

Práce je rozdělena do dvou částí. První část tvoří obecný teoretický přehled, který je shrnutím podstatných informací, ze kterých dále vychází praktický výzkum. Teoretické kapitoly jsou věnovány vzniku a vývoji strojového překladu, původním přístupům k MT založeným na lingvistických pravidlech, statistickému strojovému překladu, který vznikl v 90. letech v IBM, a v současnosti používaným hybridním systémům, které staré i nové metody slučují. Součástí teoretického přehledu jsou i dostupné informace o dvou v Česku nejběžnějších bezplatných online překladačích: Googlu Translate a Microsoft Bing Translatoru, protože právě tyto dva překladače tvoří základ praktického výzkumu.



Metodou výzkumu bude případová studie provedená na 10 různorodých pokusných textech. Tyto texty budou automaticky přeloženy Googlem Translate a Bing Translatorem. Strojově přeložený výstup poslouží k analýze a z této analýzy budou vyhodnoceny nejběžnější jevy, na které může uživatel MT, tedy i profesionální překladatel narazit. Dále bude shrnuto, na co by si profesionální překladatel při zařazení MT do své práce měl dát pozor, čím by se neměl nechat ovlivnit a na co by se naopak mohl relativně spolehnout.

Případová studie je zaměřena na jazykovou dvojici čeština – angličtina, konkrétně na směr překladu z češtiny do angličtiny. Čeština patří k „malým“ jazykům s nižším počtem mluvčích a praxe ukazuje, že značný objem překladů mezi těmito dvěma jazyky probíhá ve směru do angličtiny. Česko-anglický překladatel, většinou rodilý mluvčí češtiny, by tedy právě při překladu do jazyka mohl pomoc strojového překladu nejvíce ocenit.

# OBECNÝ TEORETICKÝ PŘEHLED

## 1 VZNIK A VÝVOJ STROJOVÉHO PŘEKLADU

Následující kapitola vychází z článku *Machine Translation: a concise history* od W. J. Hutchinse (2010).

První počítače byly vyrobeny ve 40. letech 20. století a jejich hlavním a původním úkolem bylo zpracovávat numerické operace. V pořadí druhým úkolem, kterým se nové přístroje měly zabývat a který měly usnadnit, byl mezijazykový překlad. Záhy vznikají první publikace o strojovém překladu (Weaverovo memorandum z roku 1949) a je uspořádána první konference.

K popularizaci strojového překladu přispěl veřejný experiment (1954), při němž bylo několik desítek anglických vět přeloženo do ruštiny. Strojový překlad se díky tomu stal předmětem dalšího výzkumu (ačkoli kvůli nedostatečnému technickému vybavení to mnohdy znamenalo jen studium teoretické) a veřejná ukázka také oslovila různé sponzory.

První překladové softwary z 50. let sestávaly ze speciálního rozšířeného dvojjazyčného slovníku a souboru pokynů, do něhož byla zakomponována morfologická, gramatická, syntaktická a další pravidla, tj. lidé se snažili počítače naučit jazyk v podstatě tak, jak se ho učí sami. První překladače lze rozdělit na 3 základní modely podle konceptu, na jehož základě fungovaly: **model přímého překladu** (direct translation), **transferový model** (transfer approach) a **interlingvální model** (interlingua approach). Detailnějším popisem těchto systémů se zabývá kapitola 2.

Mnozí tehdy očekávali plně automatický překlad s výsledky srovnatelnými s výkonem profesionálního překladatele (fully automatic high quality translation, FAHQ) během několika následujících let. Avšak výzkumníci narazili na tzv. sémantickou bariéru, tedy na to, že počítačům schází tzv. znalost skutečného světa.

Poradní výbor pro automatické zpracování jazyka (Automatic Language Processing Advisory Committee, ALPAC) vydal v roce 1966 zprávu, jejíž závěry ohodnotily strojový překlad jako pomalejší, nepřesnější a dvakrát tak drahý než překlad lidský (ALPAC, 1966). Výzkum automatického překladu se po vydání této zprávy přibližně na 10 let výrazně omezil (především v USA, ale i jinde).

V roce 1970 vznikl v Kanadě transferový systém pro překlad předpovědi počasí Météo, přizpůsobený omezené slovní zásobě a syntaxi meteorologických zpráv, který po dalších úpravách funguje až do současnosti.

Na přelomu 70. a 80. let se zájem vědců o strojový překlad znovu prohloubil. Ještě pořád mu vévodily dvě základní, ač propracovanější linie výzkumu. Vědci se přiklonili buď k tří-fázovému transferovému přístupu (2. generace transferových systémů), nebo k univerzálnímu umělému jazyku (interlingvě).

Do konce 80. let se výzkumníci soustředili na dosažení automatického překladu pomocí lingvistických (morfologických, syntaktických, lexikálních, sémantických a analyzačních, transferových a generačních) **pravidel**, která implementovali v překladových softwarech. S přelomem desetiletí skupina výzkumníků z IBM vytvořila nový překladový systém založený na empirických poznatcích, konkrétně na výsledcích statistické analýzy, která byla provedena na bilingválním korpusu. Vznikl tak přístup založený na korpusech, tedy na člověkem přeloženém vzoru (corpus-based approach), během něhož **nebyla uplatněna žádná lingvistická pravidla**. Přestože byl statistický strojový překlad na počátku svého vývoje, jeho úspěšnost se rovnala téměř 50 %. Statistickým strojovým překladem se detailněji zabývá kapitola 3.

Rozsáhlejší korpusy poskytovaly časem stále přesnější statistickou analýzu, a díky tomu i kvalitnější překlad. Od 90. let se strojový překlad přesunul i do komerční a podnikatelské sféry a stal se součástí trhu.

V současnosti probíhá výzkum strojového překladu především v rámci velkých internetových korporací, které zpřístupnily automatický překlad široké veřejnosti jako jednu ze svých služeb (první bezplatný online překladač se na internetu objevil v roce 1997).

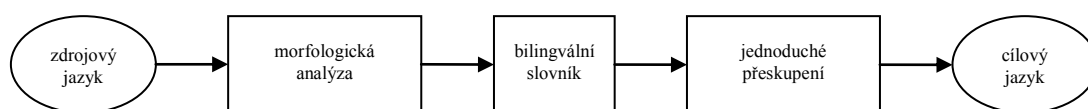
Současným nejnovějším trendem je hybridní strojový překlad, který čerpá ze statistického i lingvistického přístupu a dalších metod a to nejlepší z nich kombinuje.

## 2 STROJOVÝ PŘEKLAD ZALOŽENÝ NA LINGVISTICKÝCH PRAVIDLECH

Tato kapitola se týká postupů, jak dosáhnout automatického překladu pomocí informací, které nabízí jazykověda. V této prvotní fázi vývoje MT se vědci snažili počítači přiblížit jednotky (slova), z nichž se jazyk skládá, jak se tyto jednotky řadí do celku (věty), strukturu tohoto celku a jeho význam. To vše na základě morfologických, lexikálních, syntaktických a sémantických pravidel. Historicky starší strojový překlad založený na lingvistice se dělí na tři základní modely (v chronologickém sledu): model přímého překladu, interlingvální model a transferový model. Jednotlivé modely popisuje Hutchins a Somers (1992, s. 71-77).

### 2.1 MODEL PŘÍMÉHO PŘEKLADU (DIRECT APPROACH)

Model přímého překladu je nejstarším přístupem vůbec. Využívaly ho prvotní překladové programy, které vznikaly na přelomu 50. a 60. let minulého století. Období přímého překladu se označuje jako 1. generace MT systémů. Hutchins a Somers (1992, s. 72) definují tento přístup jako strategii, při níž se během překladového procesu neuplatňují žádné mezičlánky, tedy od zpracování vstupu ve zdrojovém jazyce vede přímá cesta k výstupu v jazyce cílovém.



Obr. 1 Model přímého překladu (převzato z Hutchins a Somers 1992, s. 72)

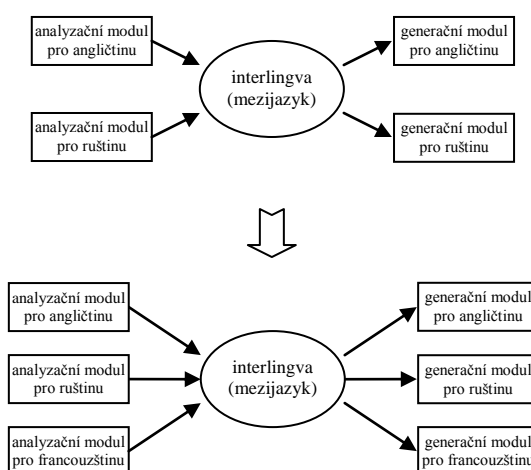
Obecný model přímého překladu znázorňuje obr. 1. Nejdříve systém provedl analýzu zdrojového textu z hlediska morfologie (např. rozpoznání koncovek a lemmat) a výsledky této analýzy pak zpracoval ve slovníku, kde došlo k překladu. Nakonec proběhlo menší přeskupení slov na základě větné stavby cílového jazyka (např. nominativ a adjektiv). Model přímého překladu se výrazně podobal překladu doslovnému.

Příkladem systému, který funguje na přímém modelu je Systran, který vznikl v 60. letech.<sup>1</sup>

## 2.2 MODEL NEPŘÍMÉHO PŘEKladU (INDIRECT APPROACH)

Přímému překladu bylo vytýkáno, že neobsahuje žádnou syntaktickou a sémantickou analýzu struktury vět a jejich významu. V systémech 2. generace tak přibyl mezičlánek, který tento abstraktnější rozbor prováděl. Vznikly metody nepřímého překladu, které dělíme na interlingvální a transferové podle toho, na jak abstraktní úrovni rozbor významu vět probíhá.

### 2.2.1 Interlingvální přístup (interlingua approach)



Obr. 2 Model interlingválního přístupu (převzato z Hutchins a Somers 1992, s. 74)

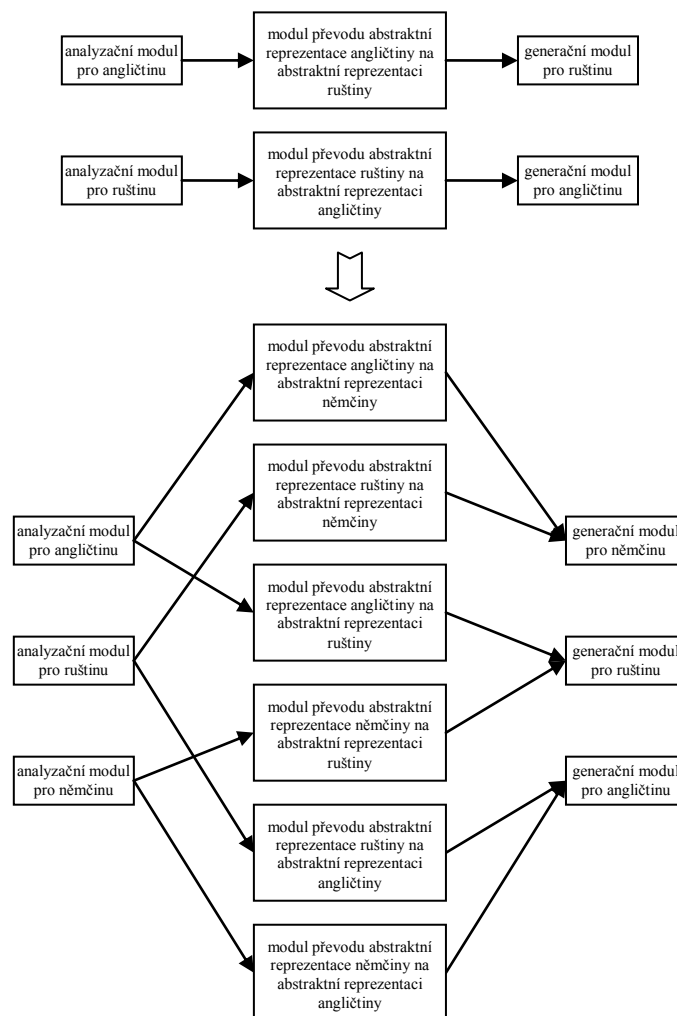
Starší interlingvální přístup se snažil z počáteční analýzy vytvořit až dokonalý abstraktní model podobný umělému jazyku (mezijazyku), který představoval samostatný článek překladového procesu a který byl zcela nezávislý na původním ZJ. Z této abstraktní reprezentace se potom vygeneroval cílový text. Cílem vědců v tomto

---

<sup>1</sup> Google Translate i Microsoft Bing Translator systém Systran částečně využívaly na počátku své existence, a to například, v případě Bing Translatoru, pro jazykové dvojice a tematické okruhy, které nepatřily k prioritám výzkumu v Microsoftu (který se zabýval především IT a technickými texty) (Schwartz 2007, Microsoft Translator Official Team Blog 2007).

období bylo vytvořit univerzální podobu této abstraktní reprezentace, kterou by bylo možné použít pro strojový překlad všech přirozených jazyků. To je ale prakticky nemožné. Přesto v tomto přístupu vědci spatřovali výhodu, protože pokud chtěli k interlingválnímu systému přidat další jazyk, stačilo přidat jen další analyzační a generační modul (obr. 2), na rozdíl od metody přímého překladu, kde bylo nutné pro každý jazykový směr napsat nový program.

### 2.2.2 Transferový přístup (transfer approach)



Obr. 3 Model transferového přístupu (převzato z Hutchins a Somers 1992, s. 75-76)

Mladší variantou nepřímého překladu je transferový přístup, který představoval „zlatou střední cestu“ mezi dvěma výše zmíněnými metodami. Překladový proces probíhal rovněž za pomoci abstraktního mezičlánku. Analýzou zdrojového textu vznikla jeho abstraktní podoba v mezičlánku a v něm poté došlo k převodu na abstraktní podobu cílového textu, kterou dále zpracoval generační modul (obr. 3). Avšak s přidáním

dalšího jazyka bylo potřeba napsat poměrně velké množství dalších programů. Kromě analyzačního a generačního modulu pro nový jazyk to byly ještě transferové moduly v obou směrech všech jazykových kombinací. Přesto vědci při konstrukci překladových systémů upřednostňovali právě transferový přístup, jelikož díky své propracovanější struktuře dosahoval nejlepších výsledků.

Strojový překlad založený principiálně na tom, že člověk učí počítač jazykovědě tak, jak ji ovládá sám, však dosáhl svého stropu. Vylepšovat se dala maximálně rychlost, kompatibilita s textovým editorem nebo upravitelnost slovníků, ale kvalita výstupu se dál nezlepšovala. Směr, kterým se v této fázi strojový překlad ubíral, vystihl Knight (1997, s. 82), který uvádí, že z odborného úhlu pohledu zůstával strojový překlad klasickým testem toho, jak dobře rozumíme přirozenému jazyku.

## 3 STATISTICKÝ STROJOVÝ PŘEKLAD

Následující kapitola se zabývá statistickými metodami strojového překladu a vychází z velké části z knižní publikace *Statistical Machine Translation* od Philippa Koehna (2010) a dále také z článku *Machine Translation: a concise history* od W. J. Hutchinse (2010). Principy statistického strojového překladu jsou v ní jen nastíněny, jelikož detailní popis těchto vysoce odborných procesů není cílem této práce.

S novým nápadem, statistickou analýzou korpusů, přišel v 90. letech tým výzkumníků z IBM. Vznikl tak nový přístup založený na člověkem přeloženém vzoru (corpus-based approach).

### 3.1 PŘÍSTUP ZALOŽENÝ NA KORPUSECH (CORPUS-BASED APPROACH)

Princip strojového překladu, který vychází z korpusu, spočívá v tom, že překladový program pracuje s bilingválním korpusem, v němž byly jednotlivá slova, ustálená spojení a věty jednoho jazyka spárovány (alignment) se svými protějšky v druhém jazyce. Překladač obdrží úsek textu k přeložení. Zaměří se na jednu překladovou jednotku zdrojového jazyka a v korpusu vyhledá všechny její překlady. Z nich vypočítá míru výskytu jednotlivých překladatelských ekvivalentů, tj. s jakou pravděpodobností se tato překladová jednotka překládá jako varianta A, varianta B, varianta C atd. (tzv. model překladu, kapitola 3.1.1.1). Na základě frekvence cílových překladových řešení v korpusu je přeložen nový text, který je upraven do uzuální podoby cílového jazyka pomocí dalších statistických výpočtů, tentokrát jen z korpusu cílového jazyka (tzv. model jazyka, kapitola 3.1.1.2).

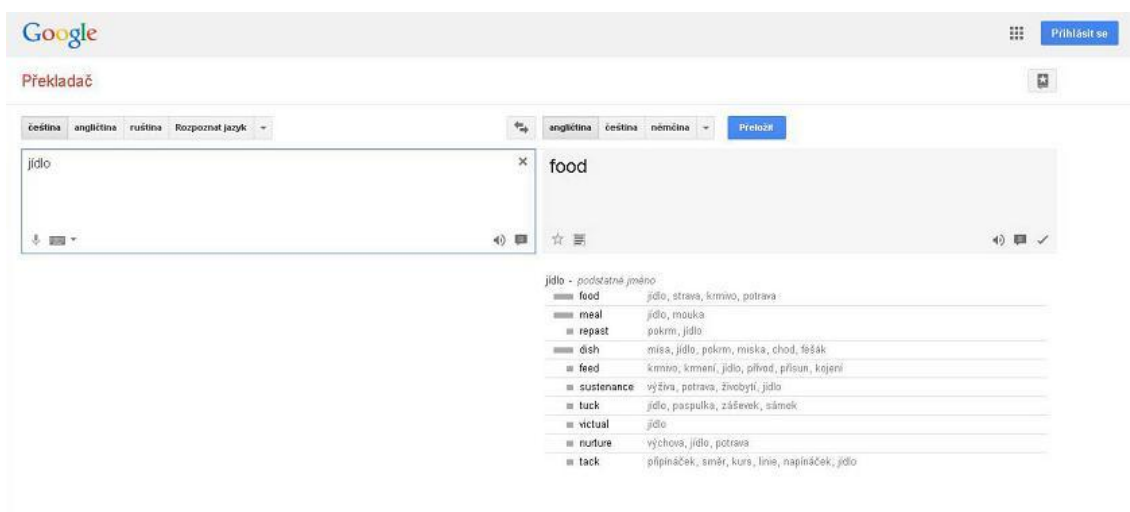
#### 3.1.1 Statistické výpočty

##### 3.1.1.1 Model překladu

Model překladu představuje první fázi SMT (statistical machine translation, statistický strojový překlad). Překlad slova do jiného jazyka je událost s několika možnými



výsledky, z nichž každý má určitou relativní četnost, kterou můžeme vyjádřit v procentech. Po zpracování velkého počtu spárovaných překladových jednotek z korpusu (internetu) dojde překladač k výsledku, který může vypadat přibližně takto: Slovo „jídlo“ bylo v 44,8 % přeloženo jako „food“, v 23,6 % jako „meal“, v 19,3 % jako „dish“, v 6,4 % jako „feed“, v 2,4 % jako „nurture“, v 1,5 % jako „sustenance“, v 0,2 % jako „repast“, v 0,2 % jako „tuck“, v 0,1 % jako „victual“ atd. Google Translate uživatelům určitý nástin četnosti překladů ukazuje, Bing Translator zatím informuje jen o nejčetnější variantě (obr. 4 a 5).



Obr. 4 Webové rozhraní Google Translate

Překladová jednotka se ale většinou nerovná jen jednomu slovu, nýbrž sestává z více prvků, jak popisuje Koehn (2010):

V současnosti nejúspěšnější přístup ke strojovému překladu využívá na pozici nejmenších jednotek fráze. Těmito frázemi se rozumí jakékoli sousedící sekvence slov, které nemusí nutně představovat lingvistické jednotky.

Věta je tedy nejdříve rozdělena na velké množství překrývajících se víceslovných útržků, které představují výchozí materiál pro statistickou analýzu. Všechny možné útržky jsou pomocí statistického výpočtu, tzv. n-gramu<sup>2</sup> přeloženy a vyhodnoceny jako více či méně pravděpodobné. Tak překladač zjistí, že věta „V osm hodin jsem přišel

<sup>2</sup> N-gram je pravděpodobnostní jazykový model, metoda, která umožňuje rozdělit pravděpodobnost překladu delší věty na pravděpodobnost překladu jejích jednotlivých částí. Uplatní se především při překladu delších a dlouhých vět, které v korpusu obvykle nemají žádný protějšek (Koehn 2010).

z práce.“ sestává ze statisticky nejpřekládanějších útržků v+osm+hodin, jsem+přišel a z+práce. Na základě modelů pracujících jak s jednotlivými slovy, tak s víceslovnými útržky pak vznikají částečné překlady. Překladač software ohodnotí každou možnou variantu překladu dané věty a jako výsledek prezentuje překlad s nejlepším hodnocením.



Obr. 5 Webové rozhraní Bing Translatoru

### 3.1.1.2 Model jazyka

V druhé fázi SMT jsou výsledky z první fáze srovnávány tentokrát už jen s monolingválním korpusem cílového jazyka a překladač tak zjišťuje, s jakou četností se podoba přeložené věty, ke které dospěl, vyskytuje v textech jazyka, do kterého překládá. V této fázi dochází například ke změnám ve slovosledu.

## 3.2 HYBRIDNÍ STROJOVÝ PŘEKLAD

U některých specifických jazykových jevů, které vychází ze specifických např. gramatických pravidel, je však výhodné překladač ještě správně usměrnit, a proto se před fází statistických výpočtů přidává ještě fáze lingvistická. Lingvistický modul zajišťuje základní gramatické modifikace, např. převedení prepozičního adjektiva v angličtině na postpoziční adjektivum ve francouzštině nebo převod anglické větné struktury S-V-O na japonskou S-O-V (Quirk, Menezes, Cherry 2005).

Systemy vzniklé kombinací statistických, lingvistických a dalších metod se nazývají hybridní. Příkladem je Microsoft Bing Translator.

## 4 BEZPLATNÉ ONLINE PŘEKLADAČE

Tato kapitola se věnuje nejběžnějším a nejznámějším překladačům v České republice: Googlu Translate od společnosti Google Inc., který je volně dostupný na webové adrese <https://translate.google.cz/>, a Microsoft Bing Translatorem od firmy Microsoft Corporation, zdarma na <http://www.bing.com/translator>. Při popisu těchto současných systémů není možné jít do hloubky, protože detailní informace o tom, jak fungují, jsou součástí firemních knowhow.

### 4.1 HISTORIE A VÝVOJ ONLINE PŘEKLADAČŮ

Nejstarším bezplatným internetovým překladačem byl systém Babel Fish, který byl spuštěn 9. prosince 1997 jako součást dříve populárního a hojně využívaného vyhledávače AltaVista. V roce 2003 odkoupila vyhledávač AltaVista společnost Yahoo! Inc., internetoví uživatelé ale mohli překladač dále využívat pod názvem Yahoo! Babel Fish. V květnu 2012 se Yahoo! Inc. rozhodl Babel Fish stáhnout a přesměrovat uživatele na Microsoft Bing Translator, který funguje od roku 2007. Google Translate byl spuštěn v roce 2006. V článku z roku 2007 Gaspari uvádí, že uživatelé internetu mají k dispozici více než třicet různých bezplatných online překladačů.

Předtím než vědci z Microsoftu a Googlu vyvinuli vlastní systémy pro všechny poskytované jazykové páry, využívaly Microsoft Bing Translator i Google Translate na počátku své existence pro více či méně jazykových párů Systran (Microsoft Translator Official Team Blog 2007, Schwartz 2007).

## 4.2 MICROSOFT BING TRANSLATOR

Hlavním zdrojem informací o tomto překladači je oficiální blog divize strojového překladu firmy Microsoft<sup>3</sup>.

Translator vznikl původně pro vnitřní potřeby své mateřské společnosti, kde bylo potřeba překládat velké množství IT a technických textů. Vývojáři na něm před spuštěním pracovali více než tři roky. Pro veřejnost funguje zdarma **od roku 2007**. Jedná se o statistický strojový překladač s lingvistickými prvky (linguistically informed statistical machine translation).

Webové rozhraní tohoto překladače (obr. 5) umožňuje zadat maximální počet 5 000 znaků na jedno zadání (chceme-li nechat přeložit text delší než 5 000 znaků, musíme ho rozdělit na části a ty zadávat postupně, anebo např. otevřít další kartu vyhledávače). Služby tohoto překladače momentálně zahrnují také bezplatný plug-in strojového překladu do překladového studia, překladové nástroje (widgety) k webovým stránkám, překladové aplikace pro mobilní zařízení, překlad příspěvků na sociálních sítích Twitter a Facebook a další (Microsoft Translator Official Team Blog 2007-2014).

## 4.3 GOOGLE TRANSLATE

Zdrojem informací o tomto překladači je rovněž oficiální firemní blog<sup>4</sup>.

Google zprovoznil svůj bezplatný čistě statistický překladač Google Translate **v roce 2006**. Jeho hlavním tvůrcem je Franz Josef Och.

Kromě webového rozhraní (obr. 4), kam je možné zadat více než 300 000 znaků najednou, zahrnují momentálně služby tohoto překladače mimo jiné placený plug-in

---

<sup>3</sup> Microsoft Translator (and Bing Translator) Official Team Blog dostupný na <http://blogs.msdn.com/b/translation/>, který byl založen v roce 2007.

<sup>4</sup> The Official Google Translate Blog dostupný na <http://googletranslate.blogspot.cz/>, který funguje od roku 2009.

strojového překladu do překladatelského studia (20 USD/1 000 000 znaků<sup>5</sup>) a stejně jako u Bingu překladové widgety k webovým stránkám, překladové aplikace pro mobilní zařízení, překlad příspěvků na sociální síti Google+ a další (The Official Google Translate Blog 2009-2012).

---

<sup>5</sup> <https://developers.google.com/translate/v2/pricing>

# PŘÍPADOVÁ STUDIE

## 5 METODOLOGIE

Hlavním cílem této práce je pomocí případové studie ohodnotit výkon dvou nejdostupnějších internetových překladačů v ČR, tedy Googlu Translate a Microsoft Bing Translatoru, jako pomocníků profesionálního překladatele. Jaké má strojový překlad pro překladatele výhody, kdy se na něj lze relativně spolehnout a na co by si překladatel měl dát pozor?

Pro účely této studie jsem vybrala 10 různorodých věcně-sdělných textů v českém jazyce z oblastí ekonomie, gastronomie, medicíny, práva, přírodních věd, psychologie, publicistiky, sportu, techniky a tlumočnictví. Jedná se o různé slohové útvary: články z odborných i populárně-naučných časopisů, obchodní podmínky, seznámení zaměstnanců s BOZP a PO či vybrané věty z učebnice, diplomové práce nebo soutěžního řádu. Pokusné texty mají rozsah minimálně 2 500 slov, nejsou to překlady z jiných jazyků, jejich autory jsou čeští rodilí mluvčí a dohromady tvoří lingvisticky rozmanitý celek vhodný k experimentování. Pokusné texty jsou podrobně popsány v Tabulce 1 a ve formátu upraveném ke zpracování se nacházejí v Přílohách 1 a 2.

Překlepy a formální a gramatické chyby byly v pokusných textech ponechány, aby experiment co nejvíce simuloval reálnou situaci a aby byla zachována autentičnost.

Pokusné texty jsem nechala automaticky přeložit. V případě Googlu Translate jsem pokusné texty importovala do MemoQu a strojový překlad provedla v něm. V případě MS Bing Translatoru jsem překlad provedla přímo na webovém rozhraní pomocí základních funkcí Kopírovat a Vložit. Výstup strojového překladu Googlu Translate jsem exportovala v bilingválním formátu .rtf přímo z MemoQu a je součástí Přílohy 1. Cílový text přeložený Bing Translateorem jsem i s originálem upravila do sloupců ve Wordu a nachází se v Příloze 2.

Přeložený materiál jsem podrobila analýze (kapitola 6), která je zaměřena na nedostatky strojového překladu z hlediska běžného cílového čtenáře (kterým je i profesionální

překladač) a dále na srovnání výkonu překladačů a kvalitu strojového překladu podle tematických okruhů a náročnosti textu, a to z hlediska profesionálního překladatele, který se MT rozhodl využít v praxi. Výsledky analýzy jsou shrnuty v kapitole 6.2. Při provádění analýzy a hodnocení výsledků jsem využila vlastní překladatelské kompetence, které jsem získala studiem bakalářského oboru Angličtina se zaměřením na tlumočení a překlad a praxí jako středně pokročilá překladatelka. Při provádění analýzy a hodnocení výsledků jsem se snažila o maximální objektivnost. Nejedná se o analýzu komplexní. Z pohledu překladatele, který se rozhodl zařadit MT do své praxe, prezentuji poznatky, které jsem vyhodnotila jako praktické, užitečné a účelné.

Výstup strojového překladu jsem popsala na základě jednotlivých úrovní jazykového plánu. Postupovala jsem od gramatiky (na úrovni morfologie a syntaxe) přes lexikální, sémantickou a textovou rovinu až po rovinu pragmatickou.

<u>Tematický okruh</u>	<u>Typ textu</u>	<u>Rok publikace</u>	<u>Zdroj</u>	<u>Autor/ři</u>	<u>Celkový počet slov</u>	<u>Příjemce textu</u>	<u>Stručná charakteristika textu</u>
<b>Ekonomie</b>	3 články: -Lotyšsko se proškrtalo k Euru. Řecko dál tápe; -České firmy platí faktury nejpozději v Evropě; -Za pár let největší ekonomika světa	2013	týdeník Ekonom: -číslo 29/18. – 24. 7. 2013; -číslo 44/31. 10. – 6. 11. 2013; -magazín Hospodářských novin a týdeníku Ekonom Český exportér, číslo 05/červenec 2013	Jan Němec; Jan Němec; Petr Zapletal	2883	zajímavost veřejnost	o složitých ekonomických tématech cestou přístupnou pro veřejnost, jednodušší stavba vět, přesná fakta, přesné číselné údaje, ekonomické termíny
<b>Gastronomie</b>	vybrané celé věty z učebnice Potraviny a výživa I	2011	Střední škola hotelnictví, gastronomie a služeb Litoměřice	Bc. Lukáš Kovač	4736	studenti oboru	vysvětlující informace, přehlednost, chemické a gastronomické termíny, snadné porozumění
<b>Medicína</b>	odborný článek Dermatoskopické vyšetření v neobvyklých indikacích	2012	časopis Česko-slovenská dermatologie: číslo 2/květen 2012, 87. ročník	T. Fikrle, L. Drlík	3234	odborníci (lékaři)	odbornost, složité a dlouhé věty, latinské lékařské termíny, popisy, obrázky
<b>Právo</b>	Všeobecné obchodní podmínky stavebního spoření	1997	Hypo stavební spořitelna a. s.	není známo	3997	veřejnost	právní terminologie, floskule, složité členění textu, dlouhá komplikovaná souvětí, modální slovesa, kondicionály
<b>Přírodní vědy</b>	2 články: -Jak opeřenci svádějí samičky?; -Po stopách počasí značených křídou	2013	časopis 21. století: číslo 11/listopad 2013	Kristýna Slomková; Jan Petříčko	2596	zajímavost veřejnost	populárně naučný styl, poutavost, cílem je zaujmout, přesná fakta, informační bohatost, příznakovost
<b>Psychologie</b>	2 články: -Chvála doteků; -Smutek	2012	časopis Psychologie dnes: -číslo 1/2012, 18. ročník; -číslo 4/2012, 18. ročník	Kateřina Rodná; Dak	2741	zajímavost veřejnost	jednoduchý rozvolněný styl, delší věty s nijak složitou stavbou, terapeutický nádech



<u>Tematický okruh</u>	<u>Typ textu</u>	<u>Rok publikace</u>	<u>Zdroj</u>	<u>Autor/ři</u>	<u>Celkový počet slov</u>	<u>Příjemce textu</u>	<u>Stručná charakteristika textu</u>
<b>Publicistika</b>	článek Dvě hodiny nudy	2013	časopis Reflex: číslo 14/4. 4. 2013, ročník XXIV	Luděk Staněk, Petr Kolář	2721	široká veřejnost	ostrá kritika, sarkasmus, ironie, černý humor, řečnické obraty, příznakovost
<b>Sport</b>	Soutěžní řád	2012	Český nohejbalový svaz	Výkonný výbor ČNS	3426	hráči nohejbalu, trenéři, rozhodčí, management	direktivita, přehlednost, přesnost, jednoznačnost, kondicionály
<b>Technika</b>	Záznam o vstupním seznámení zaměstnanců s BOZP a PO	2006	Univerzita Palackého v Olomouci	není známo	2725	zaměstnanci UP	direktivita, imperativní infinitivy, přesnost, jednoznačnost, formálnost
<b>Tlumočnictví</b>	bakalářská diplomová práce Vliv stresu během konsektivního tlumočení	2013	Univerzita Palackého v Olomouci	Bui Anh Ngoc	2710	akademická obec	organizovanost, přehlednost, přesná fakta, odborná slovní zásoba, křížové odkazy, poznámky pod čarou, grafy

Tabulka 1 Podrobný popis pokusných textů

## 6 EVALUACE STROJOVÉHO PŘEKLADU

Tato kapitola prezentuje výsledky analýzy, která byla provedena na pokusných textech, a její vyhodnocení.

### 6.1 PŘEHLED A ANALÝZA JEVŮ, K NIMŽ U MT DOCHÁZÍ

#### 6.1.1 Rovina gramatická na úrovni slova (morfologie)

*Jev č. 1: Singulár a plurál*

Popis: U podstatných jmen dochází k záměně singuláru za plurál a plurálu za singulár, zejména u deklinačních tvarů, které jsou shodné pro více pádů a vyskytují se jak v jednotném, tak v množném čísle.

<b>OT:</b>	<i>Euro má být logickým pokračováním lotyšské <u>cesty</u>.</i>
<b>GMT:</b>	<i>Euro has to be a logical continuation of the Latvian <u>roads</u>.</i>
<b>BMT:</b>	<i>The Euro has to be a logical continuation of the Latvian <u>way</u>.</i>

Výskyt 1

<b>OT:</b>	<i>Lotyšský <u>lat</u> je navázán na euro od roku 2008...</i>
<b>GMT:</b>	<i>Latvian <u>lat</u> is pegged to the euro since 2008, ...</i>
<b>BMT:</b>	<i>The Latvian <u>lats</u> is pegged to the euro since 2008...</i>

Výskyt 2

<b>OT:</b>	<i>...objevují se literární zmínky o ento(mo)dermatoskopii jako nové diagnostické <u>možnosti</u> [32].</i>
<b>GMT:</b>	<i>...there are literary references to the ento (mo) Dermatoscopy as new diagnostic <u>possibilities</u> [32].</i>
<b>BMT:</b>	<i>...literary mentions of ento (mo) dermatoskopii as the new diagnostic <u>options</u> [32].</i>

Výskyt 3

<b>OT:</b>	<i>Virové <u>infekce</u></i>
<b>GMT:</b>	<i>Viral <u>infection</u></i>
<b>BMT:</b>	<i><u>A</u> viral <u>infection</u></i>

Výskyt 4

<b>OT:</b>	<i><u>Změny smluv</u> o stavebním spoření</i>
<b>GMT:</b>	<i><u>Changes</u> building savings <u>contracts</u></i>
<b>BMT:</b>	<i><u>Change</u> building savings <u>contract</u></i>

Výskyt 5

<b>OT:</b>	<i>Je zakázáno otvírat <b>rozvaděče</b> elektrické energie, zapínat vypadlé <b>jističe, pojistky</b>, apod.</i>
<b>GMT:</b>	<i>Do not open the <b>cabinet</b> of electricity, turn Tripped <b>circuit breaker, fuse</b>, etc.</i>
<b>BMT:</b>	<i>It is forbidden to open the <b>distributor</b> of electrical energy, falling on <b>circuit breakers, fuses</b>, etc.</i>

Výskyt 6

Komentář: Často. Četněji záměna singuláru za plurál.

*Jev č. 2: Přechylování ženského rodu*

Popis: S touto typicky českou záležitostí se strojový překlad vypořádává pomocí opakování.

<b>OT:</b>	<i><b>Hráči a hráčky</b> kategorie veteránů mohou startovat i v kategorii dospělých.</i>
<b>GMT:</b>	<i><b>Players veterans and players</b> can also take part in the adult category.</i>
<b>BMT:</b>	<i><b>The players and the players</b> category veterans start even in the adult category.</i>

Výskyt 1

Komentář: V pokusných textech se vyskytlo jen několik případů, ale vždy došlo k překladu opakováním.

*Jev č. 3: Členy*

Popis: Dochází k vynechání členu.

<b>OT:</b>	<i>Rozdílné přístupy ke <b>krizi</b></i>
<b>GMT:</b>	<i>Different approaches to crisis [<b>the crisis of 2009</b>]</i>
<b>BMT:</b>	<i>Different approaches to <b>the</b> crisis</i>

Výskyt 1

<b>OT:</b>	<i>Tři čtvrtiny <b>tamní fiskální konsolidace</b>...</i>
<b>GMT:</b>	<i>Three-quarters of local fiscal consolidation... [<b>the Latvian fiscal consolidation</b>]</i>
<b>BMT:</b>	<i>Three quarters of <b>its</b> fiscal consolidation... [<b>Latvia's</b>]</i>

Výskyt 2

Komentář: Úspěšnost v překladu členů je poměrně variabilní. Správně bývají přeloženy především zavedené členy (the Czech Republic, the IMF). Pokud se užití členu liší podle konkrétní situace a daného kontextu, úspěšnost správného strojového překladu je dosti nahodilá a není možné říci, že by jeden překladač vynikal nad druhým.

### 6.1.2 Rovina gramatická na úrovni věty (syntax)

*Jev č. 4: Nominální konstrukce*

Popis: U vícečetných spojení substantiv a adjektiv, tj. u různých substantiv rozvinutých dalšími nominálními přívlasky shodnými i neshodnými, není zachován jejich vzájemný syntaktický vztah a jednotlivá slova jsou přeložena samostatně.

<b>OT:</b>	<i>Rychlost řešení</i>
<b>GMT:</b>	<i>Speed solutions</i>
<b>BMT:</b>	<i>The speed <u>of</u> the solution</i>

Výskyt 1

<b>OT:</b>	<i>...na <u>pracovištích zaměstnavatele</u>...</i>
<b>GMT:</b>	<i>...at the workplace <u>and</u> the employer...</i>
<b>BMT:</b>	<i>...at the workplace <u>of</u> the employer...</i>

Výskyt 2

<b>OT:</b>	<i>...po <u>předání postiženého kvalifikovaným záchranářům</u>.</i>
<b>GMT:</b>	<i>...the <u>victim handover qualified responders</u>.</i>
<b>BMT:</b>	<i>...after handing over the affected <u>or</u> qualified first responders.</i>

Výskyt 3

Komentář: Někdy. Tyto nominální konstrukce působí těžkosti spíše Googlu. U Bingů se dá předpokládat zásah lingvistického modulu, který může dopomoci ke správnému překladu.

*Jev č. 5: Podmět nevyjádřený*

Popis: Jelikož v angličtině je podmět obligátně vyjádřen, hledají i překladatelé v české větě s podmětem nevyjádřeným nějaký větný člen, který by podmětu odpovídal.

<b>OT:</b>	<i>Stále je sice televizní ikonou, ... [Moravec]</i>
<b>GMT:</b>	<i>Although <u>television</u> is still an icon, ...</i>
<b>BMT:</b>	<i>Although <u>television</u> is still an icon, ...</i>

Výskyt 1

Komentář: Jen občas, ale posun významu je pak nevyhnutelný.

*Jev č. 6: Slovesný čas*

Popis: Nesprávně zvolený slovesný čas a chyby v jeho gramatice.

<b>OT:</b>	<i>Závažnost konkrétních případů <u>posuzuje</u> řídicí orgán soutěže.</i>
<b>GMT:</b>	<i>The severity of individual cases <u>assessed</u> governing body of the competition.</i>
<b>BMT:</b>	<i>The severity of the cases, the managing authority <u>shall examine</u> the competition.</i>

Výskyt 1

<b>OT:</b>	<i>„<u>Viděl jsi</u> XY, jak byl včera u Krause?“</i>
<b>GMT:</b>	<i>“<u>You saw</u> XY, how was yesterday at Krause?“</i>
<b>BMT:</b>	<i>“<u>have you seen</u> the XY, how was last night for Krause?“</i>

Výskyt 2

Komentář: Relativně často. V odhalení, kterým z mnoha anglických slovesných časů se český přítomný, budoucí a minulý čas v dané situaci překládá, podává o málo lepší výkon Bing.

<b>OT:</b>	<i>A Moravec si zjevně <b>zahrál</b> o udržení vlastní „nezávislosti“.</i>
<b>GMT:</b>	<i>A Moravec <b>have</b> obviously <b>played</b> a sustain their “independence”.</i>
<b>BMT:</b>	<i>And Moravec <b>played</b> apparently to maintain their own “independence”.</i>

Výskyt 2

Komentář: Občas se u sloves setkáváme i se základními gramatickými chybami, například chybějící či naopak nadbytečnou příponou -s 3. osoby singuláru, anebo záměnou jednoho konjugačního tvaru za druhý, jak ukazuje Výskyt 2.

*Jev č. 7: Pasivum pomoci zvratného se*

Popis: České pasivum, které bylo vytvořeno pomocí zvratného zájmena *se*, překladače jako pasivum vůbec nerozpoznají.

<b>OT:</b>	<i><b>Sklíží se</b> postupně.</i>
<b>GMT:</b>	<i><b>Harvesting</b> gradually.</i>
<b>BMT:</b>	<i><b>Reaps</b> gradually.</i>

Výskyt 1

<b>OT:</b>	<i>O úvěru ze stavebního spoření <b>se sepíše</b> úvěrová smlouva.</i>
<b>GMT:</b>	<i>The loan of savings <b>to draw up</b> a loan contract.</i>
<b>BMT:</b>	<i>on loan from building savings with a credit agreement. [přísudek zcela chybí]</i>

Výskyt 2

Komentář: Někdy. V překladu se místo pasiva vyskytuje jeden z přítomných časů v aktivu, infinitiv, gerundium, anebo přísudek zcela chybí.

*Jev č. 8: Podmiňovací způsob*

Popis: Strojový překlad podmiňovacího způsobu.

<b>OT:</b>	<i>„<b>Kdybych jen udělal</b> to a to, <b>dopadlo by</b> to jinak.“</i>
<b>GMT:</b>	<i>“<b>If</b> only <b>I had done</b> this and that , it <b>would have been</b> otherwise.”</i>
<b>BMT:</b>	<i>“I just <b>did</b> it and it <b>turned out</b> it <b>would be</b> otherwise.” [chybí if]</i>

Výskyt 1

<b>OT:</b>	<i><b>Drží-li vás partner</b> ve stresové situaci budící strach za ruku, <b>nebudete se tolik bát.</b></i>
<b>GMT:</b>	<i><b>Holds if you partner</b> in a stress situation for fear-inspiring hand, <b>you will not worry</b> so much.</i>
<b>BMT:</b>	<i><b>If you partner</b> in a stress situation, formidable, <b>not so much to worry about.</b> [chybí slovesa]</i>

Výskyt 2

<b>OT:</b>	A <i><b>pokud to pravda je, jsme</b> na tom možná ještě hůře, než si myslíme.</i>
<b>GMT:</b>	And <i><b>if it is true</b>, we <b>were</b> probably even worse than you think.</i>
<b>BMT:</b>	And <i><b>if it is true</b>, <b>are we</b> perhaps even worse than we think.</i>

Výskyt 3

Komentář: Relativně často. Dochází k významovému posunu, vynechávkám či matoucím zpřeházení větných členů.

### 6.1.3 Rovina lexikální

*Jev č. 9: Slovní zásoba*

Popis: Pokud překladáč v dostupných materiálech nenalezne žádný ekvivalent daného slova, nechává ho v anglickém cílovém textu v původní podobě, tedy nepřeložené.

Výrazy, se kterými si neporadí ani jeden z překladáčů, jsou například *proškrtało, euroizovaná, podlážkách, zasyrova, sladkovišně, tungiázu, akuminátní, trsovité, vykřičníkovité, hypopigmentovaných, pseudosít, postaršení, přespoření, tokaniště, zabydlující, čtyřsetnásobek, egoističtější, konsektiva* aj. Bing Translator si navíc na rozdíl od Googlu nevěděl rady se slovy jako *nepěstěné, moučnatostí, ploskách, krvácivým, červenočerné, nepřeceňovat, propletou, podmaněnými, žactvo, konsektivně, vyškubnutá, polopouštích, šušotáž* aj. Výrazů, které přeložil pouze Bing a Google nikoli, bylo zanedbatelné množství a daly by se „spočítat na prstech jedné ruky“.

Na textu č. 3 z oblasti medicíny je rozdíl v širší slovní zásoby obou překladáčů obzvláště patrný. Google zvládá i pokročilou terminologii jako *dermatoskop, dermatoskopické, imerzní* či *anxiózních (dermatoscope, dermatoscopic, immersion, anxious)*, zatímco v cílovém textu Bingu zůstávají tato a další slova nepřeložená.

Komentář: Větší slovní zásobou disponuje jednoznačně Google Translate. Bing Translator nechává násobně více výrazů nepřeložených, tzn. v Bingem přeloženém anglickém cílovém textu se česká slova vyskytují poměrně často.

*Jev č. 10: Vynechání slova*

Popis: Slovo není přeloženo ani zkopírováno ze zdrojového textu, v cílovém textu zcela chybí.

<b>OT:</b>	Účel <i><b>smlouvy</b></i> o stavebním spoření
<b>GMT:</b>	The purpose of the building savings [ <b>protějšek zcela chybí</b> ]
<b>BMT:</b>	The purpose of <i><b>the contract</b></i> on building savings

Výskyt 1

Komentář: Relativně často. V některých takových případech má vynechání slova na předání invariantu informace minimální vliv, naopak v jiných, pokud se jedná o slovo, které je neopomenutelným nositelem významu, dojde k jeho posunu.

<b>OT:</b>	<i>V ostatních případech <b>platí</b> příslušná zákonná ustanovení.</i>
<b>GMT:</b>	<i>In other cases, the relevant statutory provisions. [protějšek zcela chybí]</i>
<b>BMT:</b>	<i>in other cases, the relevant statutory provisions. [protějšek zcela chybí]</i>

Výskyt 2

Komentář: Vynechaným slovem často bývá předložka či sloveso.

*Jev č. 11: Přebytečný výraz*

Popis: Přidání nového výrazu, který se v originálu vůbec nevyskytuje.

<b>OT:</b>	<i>Všeobecné obchodní podmínky <b>stavebního spoření</b></i>
<b>GMT:</b>	<i>Terms and conditions of <b>building savings</b></i>
<b>BMT:</b>	<i>General business conditions for <b>building savings loan</b> [úvěr ze stavebního spoření]</i>

Výskyt 1

<b>OT:</b>	<i>...vypracoval studii, v níž poprvé nastiňuje paralelu mezi globálním ochlazováním planety před miliardou let a kolapsem vodního ekosystému v pozdních <b>druhohorách.</b></i>
<b>GMT:</b>	<i>...a study in which first outlines the parallels between global cooling of the planet billions of years ago and the collapse of the aquatic ecosystem in the late <b>Mesozoic.</b></i>
<b>BMT:</b>	<i>...drafted the study, in which the first outlines the parallel between the global cooling of the planet before a billion years and the collapse of the aquatic ecosystem in the late <b>"said my uncle.</b></i>

Výskyt 2

Komentář: Je-li nadbytečné slovo zároveň naprosto nesmyslné, dochází i k narušení orientace v cílovém textu. S tímto problémem se setkáváme spíše u Bing Translatoru.

*Jev č. 12: Zkratky*

Popis: Strojový překlad zkratk

Překladače zvládají bez potíží zkratky *např., resp., aj., apod., atd.* a *Kč.* Zkratku *event.* kopírují v původní podobě, ta ale funguje i v cílovém textu. Google na rozdíl od Bingu správně překládá i *tj.* a *mj.* Oba překladače pak mají problém u zkratky *př.*, kterou buď vynechávají, anebo ji zaměňují s určením letopočtu a překládají jako *BC.*

<b>OT:</b>	<i>Ohlásit požár (nebo zabezpečit jeho ohlášení) na <b>HZS</b> Olomouckého kraje... [Hasičský záchranný sbor]</i>
<b>GMT:</b>	<i>Report a fire (or secure its announcement) to <b>FRS</b> Olomouc Region... [Fire (&amp;) Rescue Service]</i>
<b>BMT:</b>	<i>Report a fire (or secure its announcement) on <b>BOTH</b> of the Olomouc region... [?]</i>

Výskyt 1

<b>OT:</b>	<i>Pokud se oba oddíly nedohodnou jinak, je odstupné při přestupu stanoveno dle ustanovení tohoto <b>SŘ</b>. [soutěžního řádu]</i>
<b>GMT:</b>	<i>If both sections agree otherwise, the transfer of severance provided under the provisions of <b>the Slovak Republic</b>.</i>
<b>BMT:</b>	<i>If both sections have agreed otherwise, the severance pay provided for under the transfer provisions of this <b>SŘ</b>.</i>

Výskyt 2

Komentář: U zkratk názvů záleží na konkrétní situaci, jejich strojový překlad je variabilní.

### *Jev č. 13: Terminologie*

Popis: Strojový překlad termínů

V rámci pokusných textů byly oběma překladači správně přeloženy termíny *sodium chloride, sodium carbonate, hydrochloric acid, higher fatty acids, the IMF, Council for Radio and Television Broadcasting, the Ministry of Finance, State aid, gross domestic product (GDP), labour code, occupational safety and health (OSH), fire protection, personal protective equipment (PPE), First World War* aj.

<b>OT:</b>	<i><b>jeřába popelavého (Grus grus)</b></i>
<b>GMT:</b>	<i><b>common crane (Grus grus)</b></i>
<b>BMT:</b>	<i><b>the crane (Grus Grus) Ashy</b></i>

Výskyt 1

<b>OT:</b>	<i><b>vestfálského míru</b></i>
<b>GMT:</b>	<i><b>the Peace of Westphalia</b></i>
<b>BMT:</b>	<i><b>the Treaty of peace</b></i>

Výskyt 2

Komentář: Oba překladače se umějí vypořádat se zavedenými termíny. Těmi jsou například názvy chemických sloučenin, názvy institucí, ekonomické termíny atd. Google podává v tomto aspektu o málo lepší výkon než Bing, jak ukazují Výskyty 1 a 2.

### *Jev č. 14: Prvky cílového jazyka v originálu*

Popis: Angličtina v českém zdrojovém textu

<b>OT:</b>	<i>Časným dermatoskopickým znakem androgenní alopecie je rozmanitost průměru jednotlivých vlasů („<b>hair diameter diversity</b>“)...</i>
<b>GMT:</b>	<i>Dermatoskopickým early sign of androgenic alopecia is the variety the diameter of each hair („<b>hair diameter diversity</b>“)...</i>
<b>BMT:</b>	<i>Early dermatoskopickým with androgenic alopecia is the diversity of the diameter of each hair („<b>hair diameter diversity</b>“)...</i>

Výskyt 1



Komentář: Překladače slova z jiného než ZJ kopírují do CT.

<b>OT:</b>	<i>Příkladem takového ovlivnění je nedávný výzkum dvou psychologů z univerzity v Yale, <b>Johna Bargha</b> a <b>Lawrence Williamse</b>.</i>
<b>GMT:</b>	<i>An example of such an effect is the recent research of two psychologists from the University of Yale, <b>John Bargh</b> and <b>Lawrence Williams</b>.</i>
<b>BMT:</b>	<i>An example of such influence is a recent research of two psychologists from the University of Yale, <b>John Bargha</b> and <b>Lawrence Williams</b>.</i>

Výskyt 2

Komentář: Překladače někdy zvládají i převod anglických jmen zpět do jejich původní nesklonné podoby.

<b>OT:</b>	<i>Řecko dál <b>tápe</b></i>
<b>GMT:</b>	<i>Greece go on <b>tape</b></i>
<b>BMT:</b>	<i>Greece continued <b>to grope</b></i>

Výskyt 3

Komentář: S výskytem cílového jazyka ve zdrojovém textu vzdáleně souvisí také to, že občas je české slovo velmi podobné úplně jinému anglickému slovu, a v důsledku to pak překladač může zmást.

*Jev č. 15: Jednopísmenná slova*

Popis: Strojový překlad *a, i, u* apod.

<b>OT:</b>	<i><b>A</b> Moravec si zjevně zahrál...</i>
<b>GMT:</b>	<i><b>A</b> Moravec have obviously played...</i>
<b>BMT:</b>	<i><b>And</b> Moravec played...</i>

Výskyt 1

<b>OT:</b>	<i>To si začal brzy uvědomovat <b>i</b> Moravec.</i>
<b>GMT:</b>	<i>It soon began to realize <b>and</b> Moravec.</i>
<b>BMT:</b>	<i>It began early <b>even</b> Moravec.</i>

Výskyt 2

Komentář: U Googlu často dochází k nepřeložení spojky *a* nebo předložky *u* apod., vyskytují-li se tato jednopísmenná slova na začátku věty. Překladače také často nezvládají spojku *i*. Bývá přeložena jako *and* a stupňovací, popř. odporovací složka jejího významu tak úplně zaniká.

#### 6.1.4 Rovina sémantická

*Jev č. 16: Jemné sémantické posuny*

Popis: Sémantický posun

<b>OT:</b>	...které si námluvy <b>zpestřují</b> nádhernými duety.
<b>GMT:</b>	...that have <b>punctuated</b> by spectacular courtship duets.
<b>BMT:</b>	...courtship <b>punctuated</b> by wonderful duets.

Výskyt 1

Komentář: Velmi často. Drobné sémantické posuny jsou u strojového překladu běžnou věcí a porozumění cílovému textu komplikují většinou jen minimálně.

Jev č. 17: Ztráta negace

Popis: Překlad je opakem originálu, jelikož došlo ke ztrátě negace

<b>OT:</b>	Počínat si při práci <b>tak, aby neohrozili</b> zdraví své ani svých spolupracovníků, ...
<b>GMT:</b>	Behave at work <b>so as to endanger</b> their health or their employees, ...
<b>BMT:</b>	Do you work <b>so as to ensure the health</b> of its employees, ...

Výskyt 1

<b>OT:</b>	S Moravcem se prostě něco stalo a <b>výsledek pro něj nevyznívá právě lichotivě.</b>
<b>GMT:</b>	And Moravec is just something happened to him, and <b>the result is rather positive just flattering.</b>
<b>BMT:</b>	With Moravcem is just something happened and <b>it just doesn't result flatteringly.</b>

Výskyt 2

<b>OT:</b>	<b>Pokud tyto náklady neuhradí,</b> považuje se to za porušení povinnosti družstva stejně jako neuhrzení pokut.
<b>GMT:</b>	<b>If the costs are not paid,</b> shall be deemed an infringement cooperatives and non-payment of fines.
<b>BMT:</b>	<b>If you pay these costs</b> shall be considered a breach of the obligations of cooperatives as well as the non-payment of fines.

Výskyt 3

<b>OT:</b>	Jespák bojovný ( <i>Philomachus pugnax</i> ) <b>nezískal</b> své jméno pro nic za nic.
<b>GMT:</b>	Ruff ( <i>Philomachus pugnax</i> ) <b>got</b> its name for nothing.
<b>BMT:</b>	Ruff ( <i>Philomachus pugnax</i> ) <b>gained</b> its name for nothing.

Výskyt 4

Komentář: Relativně často. Bývá úsměvné.

Jev č. 18: Homonymie, polysémie a neschopnost monosémantizovat kontextem

Popis: Počítačový program není schopen monosémantizovat podle kontextu (ani v rámci jedné věty).

<b>OT:</b>	<b>Plod</b> má hruškovitý tvar...
<b>GMT:</b>	<b>The fruit</b> is pear-shaped, ...
<b>BMT:</b>	<b>The fetus</b> has a pear-shaped...

Výskyt 1

<b>OT:</b>	<i>Jsou dobře stravitelné, nenáročné na přípravu a mají všestranné využití (např. jako zavářka, k přípravě <b>příloh</b> či hlavních jídel)</i>
<b>GMT:</b>	<i>They are easy to digest, easy to prepare and have the versatility (as a garnish, or <b>attachments</b> to prepare main dishes)</i>
<b>BMT:</b>	<i>They are easily digestible, easy to prepare and versatile applications (e.g. as a garnish, to prepare the <b>annexes</b> or main dishes)</i>

Výskyt 2

<b>OT:</b>	<i>Dermatoskopické vyšetření je jednoduché, neinvazivní, rychlé, umožňuje bezproblémové vyšetření většího počtu <b>projevů</b> nebo větší části kožního povrchu.</i>
<b>GMT:</b>	<i>Dermatoscopic examination is simple, noninvasive, rapid, makes a major number of <b>speeches</b> or more of the skin surface.</i>
<b>BMT:</b>	<i>Dermatoskopické examination is a simple, non-invasive, fast, allows seamless testing a large number of <b>speeches</b> or the greater part of the surface.</i>

Výskyt 3

<b>OT:</b>	<i><b>Poklepáním</b> na rameno dodáváme odvalu tomu, kdo jde na zkoušku nebo na start závodu.</i>
<b>GMT:</b>	<i><b>Double click</b> on the arm give courage to the person who goes to trial or to start the race.</i>
<b>BMT:</b>	<i><b>Double-click</b> on the shoulder deliver the courage, who goes on trial or at the start of the race.</i>

Výskyt 4

Komentář: V pokusných textech se vyskytly například výrazy *plod*, *příloha*, *projev* či *poklepání*, které překladače nejsou schopny monosémantizovat kontextem a dochází tak často k nesprávným, poměrně úsměvným překladům.

Jev č. 19: Idiomy

Popis: Strojový překlad idiomů

<b>OT:</b>	<i>Bez správného volání a natřásání by to pro samici bylo jako <b>hledat jehlu v kupce sena</b>.</i>
<b>GMT:</b>	<i>Without the correct call and shaking it for the female was like <b>finding a needle in a haystack</b>.</i>
<b>BMT:</b>	<i>Without the correct call and shaking it for the female was like <b>looking for a needle in a haystack</b>.</i>

Výskyt 1

Komentář: Pokud má idiom v češtině i angličtině stejné znění, překladače ho převádějí vcelku správně.

<b>OT:</b>	<i>Jaké jsou ptačí „<b>balicí techniky</b>“?</i>
<b>GMT:</b>	<i>What can „<b>packaging technology</b>“?</i>
<b>BMT:</b>	<i>What are the bird „<b>wrapping</b>“?</i>

Výskyt 2

Komentář: Pokud se objeví ustálené přenesené spojení, jehož komunikativní protějšek má v angličtině zcela jinou podobu, strojový překlad obvykle není schopen k této

odlišné variantě dospět a význam idiomu v cílovém textu zaniká, protože je přeložen doslova.

<b>OT:</b>	<i>Přestože současná <b>situace</b> Lotyšska <b>není</b> pro obyvatele <b>růžová</b>, mohou se utěšovat „<b>světlem na konci tunelu</b>“.</i>
<b>GMT:</b>	<i>While the current <b>situation</b> of Latvia, residents <b>are not pink</b>, they can take comfort “<b>light at the end of the tunnel</b>”.</i>
<b>BMT:</b>	<i>Although the current <b>situation is not rosy</b> for the residents of Latvia, may find the “<b>light at the end of the tunnel</b>”.</i>

Výskyt 3

Komentář: V této větě se vyskytují oba dva předchozí případy.

<b>OT:</b>	<i><b>Když ptáčka lapají, pěkně mu zpívají.</b></i>
<b>GMT:</b>	<i><b>If a bird Lapa, nice sing him</b></i>
<b>BMT:</b>	<i><b>When the bird catching, he sing</b></i>

Výskyt 4

<b>OT:</b>	<i><b>Podle nosa poznáš kosa</b></i>
<b>GMT:</b>	<i><b>By nosa poznas kosa</b></i>
<b>BMT:</b>	<i><b>According to nosa you spit</b></i>

Výskyt 5

Komentář: Často však dochází k úplnému rozpadu idiomu na slova doslovně přeložená a nepřeložitelná.

<b>OT:</b>	<i><b>Láska prochází žaludkem</b></i>
<b>GMT:</b>	<i><b>Love goes through the stomach</b></i>
<b>BMT:</b>	<i><b>A taste of love</b></i>

Výskyt 6

Komentář: Za zmínku stojí strojový překlad českého přísloví *láska prochází žaludkem*, které se objevilo v textu č. 5 z oblasti přírodních věd (populárně-naučný článek z časopisu 21. století). Za dvěma odlišnými variantami stojí dva různé překladatelské postupy. Zatímco Google se drží doslovného překladu, který příjemce textu může a nemusí navést na anglické přísloví *The way to a man's heart is through his stomach.*, Bing předkládá rovnou upravenou verzi, která vystihuje zamýšlený význam, ale původní podoby přísloví se nedrží.

<b>OT:</b>	<i><b>Pak už to šlo ráz na ráz.</b></i>
<b>GMT:</b>	<i><b>Then <b>we did it happen in nature.</b></b></i>
<b>BMT:</b>	<i><b>Then <b>it was led to another.</b></b></i>

Výskyt 7

Komentář: Najdou se ale i naprosto nezvládnuté a nesmyslné překlady idiomů.

### 6.1.5 Rovina textová

Jev č. 20: Koheze

Popis: Strojový překlad kohezních vztahů

<b>OT:</b>	<i>...za porušení <b>jeho</b> povinností při řešení <b>jeho</b> bytových potřeb.</i>
<b>GMT:</b>	<i>...for breach of <b>its</b> obligations in addressing <b>its</b> housing needs.</i>
<b>BMT:</b>	<i>...for a breach of <b>his</b> obligations in dealing with <b>its</b> housing needs.</i>

Výskyt 1

<b>OT:</b>	<i>Tímto směrem se ve <b>své</b> práci zaměřil francouzský psycholog (...), <b>který</b> se v roce 2007 jal zkoumat...</i>
<b>GMT:</b>	<i>This is the direction in <b>his</b> work focused French psychologist (...), <b>who</b> in 2007 began to study ...</i>
<b>BMT:</b>	<i>In this direction in <b>its</b> work focused French psychologist (...), <b>which</b> in 2007 began to examine...</i>

Výskyt 2

Komentář: Automatické překladače postrádají jakoukoli „znalost skutečného světa“ a zároveň nejsou schopny brát v potaz kontext. Lze tedy říci, že správné vyjádření kohezních vztahů je pro ně velkým problémem. Osobní, přivlastňovací či vztažná zájmena jsou typicky přeložena v jiné osobě či v jiném rodě.

Jev č. 21: AČV

Popis: Strojový překlad a reprezentace funkční větné perspektivy.

<b>OT:</b>	<i><b><u>Dermatoskopie umožňuje identifikaci živého parazita nebo jednotlivých hnid.</u></b></i>
<b>GMT:</b>	<i><b><u>Dermatoscopy allows identification of the living parasite or individual nits.</u></b></i>
<b>BMT:</b>	<i><b><u>Dermatoscopy allows the identification of live parasite or individual nits.</u></b></i>

Výskyt 1

<b>OT:</b>	<i><b><u>Stavební spořitelna není povinna zdůvodnit odmítnutí žádosti.</u></b></i>
<b>GMT:</b>	<i><b><u>Construction Savings Bank is not obliged to give reasons for refusing the request.</u></b></i>
<b>BMT:</b>	<i><b><u>The building society is not obliged to justify the denial of the request.</u></b></i>

Výskyt 2

Komentář: Neutrální, nepříznakové uspořádání větných členů v češtině, které navíc víceméně odpovídá pevnému slovosledu přeložené anglické věty (S-V-O), vyústí ve velmi zdařilé, téměř dokonalé věty.

<b>OT:</b>	<i><b><u>A my jejich léčivé účinky intuitivně využíváme.</u></b></i>
<b>GMT:</b>	<i><b><u>And we their medicinal effects intuitively use.</u></b></i>
<b>BMT:</b>	<i><b><u>And we intuitively use their healing effects.</u></b></i>

Výskyt 3

Komentář: Stylisticky příznakový slovosled mají překladače tendenci zachovat přesně tak, jak vypadá ve zdrojovém textu. Výsledkem je potom anglická věta, která vypadá,

jako by někdo její slova libovolně zpřeházel. S největší pravděpodobností má Bing v tomto aspektu díky přidanému lingvistickému modulu výhodu, jelikož někdy překvapuje velmi zdařilou úpravou původního uspořádání, která vede k dobře formulovaným anglickým větám.

<b>OT:</b>	<i>Rozhodčí deleguje řídicí orgán soutěže...</i> [předmět-přísudek-podmět]
<b>GMT:</b>	<i>Judge delegates governing body competition...</i> [podmět-přísudek-předmět]
<b>BMT:</b>	<i>Judge delegates management authority...</i> [podmět-přísudek-předmět]

Výskyt 4

<b>OT:</b>	<i>Kolaps HDP v prvním pololetí 2009 zavinila krize.</i> [předmět-další větné členy-přísudek-podmět]
<b>GMT:</b>	<i>The collapse of GDP in the first half of 2009 caused the crisis.</i> [podmět-další větné členy-přísudek-předmět]
<b>BMT:</b>	<i>The collapse in GDP in the first half of 2009 caused the crisis.</i> [podmět-další větné členy-přísudek-předmět]

Výskyt 5

Komentář: K zásadnímu nedorozumění však dochází typicky u vět, které jsou v češtině příznakové a jejich uspořádání odpovídá O-V-S. Při překladu zákonitě dojde k zachování původního slovosledu a anglická věta pak vypadá takto: O=S-V-S=O.

### 6.1.6 Rovina pragmatická

*Jev č. 22: Formální stránka*

Popis: Vliv umístění interpunkčních znamének, ale i podoby faktických údajů na strojový překlad.

<b>OT:</b>	<i>První pomoc při úrazech <u>elektřinou.</u></i>
<b>GMT:</b>	<i>First aid for shock. [protějšek zcela chybí]</i>
<b>BMT:</b>	<i>First aid for injuries. [protějšek zcela chybí]</i>

Výskyt 1

<b>OT:</b>	<i><u>První pomoc při úrazech elektřinou</u></i>
<b>GMT:</b>	<i><u>First aid for electrical shock</u></i>
<b>BMT:</b>	<i><u>First aid for injuries by electricity</u></i>

Kontrolní zdrojový text (bez tečky) a jeho strojový překlad

Komentář: V některých případech na formální stránce zdrojového textu záleží. Ve výše uvedeném Výskytu 1 určuje omylem napsaná tečka na konci nadpisu do velké míry srozumitelnost překladu.

<b>OT:</b>	<i>Je zdrojem i antinutriční látky kyseliny <u>Šťavelové</u>, která blokuje v organismu využitelnost vápníku.</i>
<b>GMT:</b>	<i>It is also a source of anti-nutritional substances <u>Šťavelové acid</u>, which blocks the utilization of calcium in the body.</i>
<b>BMT:</b>	<i>It is the source and substance of antinutriční <u>happy avelové</u>, which blocks the body's use of calcium.</i>

Výskyt 2

<b>OT:</b>	<i>Je zdrojem i antinutriční látky <u>kyseliny Šťavelové</u>, která blokuje v organismu využitelnost vápníku.</i>
<b>GMT:</b>	<i>It is also a source of anti-nutritional substances <u>oxalic acid</u>, which blocks the utilization of calcium in the body.</i>
<b>BMT:</b>	<i>Is the source of antinutriční and <u>oxalic acid</u>, a substance that blocks the body's use of calcium.</i>

Kontrolní zdrojový text (s ť) a jeho strojový překlad

Komentář: Nepřeložitelná jsou pro překladače také slova, kde se místo písmen s přimknutým háčkem vyskytuje písmeno následované apostrofem.

<b>OT:</b>	<i>Mateřský oddíl může v průběhu hracího roku hostování ukončit <u>30.6.</u> a vrátit hráče zpět do mateřského oddílu.</i>
<b>GMT:</b>	<i>The parent partition can during the playing of hosting terminate <u>06.30</u> and return the player back to the parent partition.</i>
<b>BMT:</b>	<i>The parent partition can in the course of the year to stop hosting <u>19.0.</u> and return the player back to the parent section.</i>

Výskyt 3

<b>OT:</b>	<i>Mateřský oddíl může v průběhu hracího roku hostování ukončit <u>30. 6.</u> a vrátit hráče zpět do mateřského oddílu.</i>
<b>GMT:</b>	<i>The parent partition can during the playing of hosting end the <u>30th 6th</u> and return the player back to the parent partition.</i>
<b>BMT:</b>	<i>The parent partition can in the course of the year, hosting the exit <u>30. 6.</u> and return the player back to the parent section.</i>

Kontrolní zdrojový text (s mezerou za tečkou po číslici 30) a jeho strojový překlad

Komentář: Rozdílná norma psaní dat v češtině a angličtině způsobuje velké potíže, nepřesnosti a chybné převody. Již v originálu chybně napsané datum překladače jen mate. Podoba dat v kontrolním překladu sice pořád neodpovídá anglické normě (např.: 30/6, 30<sup>th</sup> June, June, 30), nicméně cílový čtenář má větší šanci si údaj správně odvodit.

<b>OT:</b>	<i>Ve smyslu struktury ekonomiky se na tvorbě čínského HDP v roce 2011 největší měrou podílel sektor průmyslu (<u>46,6 %</u>) a služby (<u>43,3 %</u>), viz graf.</i>
<b>GMT:</b>	<i>In terms of economic structure in the formation of China's GDP in 2011 contributed the largest industrial sector (<u>46.6%</u>) and services (<u>43.3%</u>), see chart.</i>
<b>BMT:</b>	<i>In terms of the structure of the economy with the creation of China's GDP in 2011, the largest sectors of industry (<u>46.6%</u>) and services (<u>43.3%</u>), see graph.</i>

Výskyt 4

Komentář: Například s procentním znakem se ale překladače vypořádávají velmi dobře. Mezeru mezi číslem a znakem procent v češtině při převodu do angličtiny vynechávají a správně mění i interpunkční znaménko pro vyjádření desetinných čísel.

<b>OT:</b>	Z <u>šedesátky</u> oslovených žen...
<b>GMT:</b>	From the <u>sixties</u> approached women...
<b>BMT:</b>	from the <u>60s</u> polled women...

Výskyt 5

<b>OT:</b>	3 ‰
<b>GMT:</b>	3 ‰
<b>BMT:</b>	3 [protějšek chybí]

Výskyt 6

Komentář: Nesrovnalosti způsobují i další číselné údaje.



## 6.2 VÝSLEDKY ANALÝZY

Tato podkapitola je shrnutím toho, jakých výsledků v současnosti dosahují dva nejdostupnější automatické překladače v Česku z hlediska užitečnosti pro profesionálního překladatele, který MT používá v praxi.

### 6.2.1 Výsledky analýzy z hlediska cílového čtenáře

Strojový překlad nedosahuje úrovně lidského překladatele a jeho výstup je defektní. Následuje stručný popis chyb, se kterými se jak běžný cílový čtenář, tak profesionální překladatel při použití MT běžně setkává, uspořádaný do kategorií na základě jednotlivých úrovní jazykového plánu.

V rámci gramatiky na úrovni slova často dochází k záměně singuláru a plurálu (*lotyšský lat* → *the Latvian lats*; *jističe, pojistky* → *circuit breaker, fuse*), dále také k diskrepancím způsobeným absencí rodového rozlišení v CJ (*hráči a hráčky* → *the players and the players*) a především není možné se spolehnout na distribuci členů (*tři čtvrtiny tamní fiskální konsolidace* → *three-quarters of local fiscal consolidation*), s výjimkou těch, které jsou součástí ustálených spojení (*the Czech Republic, the IMF*). Na úrovni větné skladby způsobují potíže nominální konstrukce (*předání postiženého kvalifikovaným záchranářům* → *victim handover qualified responders*) a podmět nevyjádřený (*stále je sice televizní ikonou* → *although television is still an icon*). Nevhodně bývá přeložen i slovesný čas (*posuzuje* → *assessed, shall examine*), pasivum vytvořené pomocí zvrátného *se* (*sklízí se* → *harvesting, reaps*) či podmiňovací způsob (*kdybych jen udělal to a to, dopadlo by to jinak* → *I just did it and it turned out it would be otherwise*).

Na lexikální rovině nastávají problémy až u velmi odborné a morfologicky složitě utvořené slovní zásoby (*tungjázu, čtyřsetnásovek*). Překladače nemají k dispozici žádné ekvivalenty těchto výrazů a do CT je pak kopírují v původní podobě (*dále lze často diagnostikovat tzv. „vykřičníkovité“ vlasy* → *in addition, you can often diagnose a “vykřičníkovité” hair*). Z analýzy pokusných textů použitých v této případové studii vyplývá, že Google Translate disponuje násobně větší slovní zásobou než MS Bing Translator. Občas se stane, že překlad konkrétního výrazu v cílovém textu zcela chybí (*v ostatních případech platí příslušná zákonná ustanovení* → *in other cases, the relevant statutory provisions*), anebo je naopak přidáno něco navíc, co se v původním

textu vůbec nevyskytuje (*podmínky stavebního spoření* → *conditions for building savings loan*). Většinu běžných zkratk (např., *apod.*) zvládají překladače velmi dobře, nicméně problémy občas nastávají u méně obvyklých zkratk názvů (*HZS = Hasičský záchranný sbor* → *BOTH = ?*; *SŘ = soutěžní řád* → *the Slovak Republic*). **Termíny patří mezi silné stránky strojového překladu**, variabilita jejich protějšků je totiž ve srovnání s běžným lexikem poměrně nízká (*bezpečnost a ochrana zdraví při práci* → *occupational safety and health*). Pokud se v českém ZT vyskytne angličtina, překladače ji do CT převádí v nezměněné podobě (*hair diameter diversity* → *hair diameter diversity*). Za zmínku stojí také poměrně problematický strojový překlad jednopísmenných slov jako jsou spojky *a* a *i* nebo předložka *u* (*a Moravec si zjevně zahrál* → *a Moravec have obviously played*).

K sémantickým posunům dochází v různém rozsahu zcela běžně (*keré si námluvy zpestřují nádhernými duety* → *courtship punctuated by wonderful duets*). Avšak vážnou a zásadní chybu představuje ztráta větné negace, tzn., že překlad je naprostým opakem originálu. Stává se to relativně často (*výsledek pro něj nevyznívá právě lichotivě* → *the result is rather positive just flattering*). Překladové systémy si neporadí s obvyklými případy polysémie a nejsou schopny monosémantizovat kontextem (*poklepáním na rameno dodáváme odvalu* → *double-click on the shoulder*). Překlad idiomů je variabilní podle toho, zda v CJ existuje přímý protějšek, pokud ano, bývá úspěšný (*hledat jehlu v kupce sena* → *looking for a needle in a haystack*). Pokud však ekvivalent v CJ chybí či má zcela jinou podobu, strojový překlad je doslovný a význam idiomu tak zaniká (*situace není růžová* → *situation are not pink*).

Strojová reprezentace kohezních vztahů a funkční větné perspektivy na rovině textové je dalším aspektem, který je pro překladače náročný (*psycholog, který* → *psychologist, which; a my jejich léčivé účinky intuitivně využíváme* → *and we their medicinal effects intuitively use*). Pokud příznakový slovosled české věty odpovídá uspořádání předmět – přísudek – podmět, překladače ho zaměňují za podmět – přísudek – předmět (*kolaps HDP v prvním pololetí 2009 zavinila krize* → *the collapse of GDP in the first half of 2009 caused the crisis*).

Z pragmatického hlediska stojí za pozornost fakt, že přítomnost formálních náležitostí v textu (například čárek) strojový překlad ovlivňuje, stejně jako překlady a další chyby v originálu (*kyseliny Št'avelové* → *happy avelové*).

### 6.2.2 Výsledky analýzy z hlediska profesionálního překladatele

Pokud se na výsledky analýzy zaměříme z pohledu profesionálního překladatele, můžeme říci, že strojový překlad nabízí různé možnosti, jak zjednodušit a hlavně urychlit práci.

Zřejmá je pomoc při osvěžení nepoužívaných slovíček a jejich přesun z pasivní do aktivní slovní zásoby či zpětná kontrola podoby některých typů výrazů, např. frázových či prepozicionálních sloves, předložkových vazeb apod.

Některé chyby automatických překladačů se překladatelovy práce téměř nedotknou, protože se jedná o chyby, které jsou hluboko pod pokročilou úrovní jazyka, která je překladateli vlastní, anebo odporují zdravému rozumu. Příkladem může být záměna singuláru za plurál, nesprávné vyjádření vztahů mezi členy nominální konstrukce, překlad, který je opakem originálu, nebo nevhodně zvolený protějšek mnohoznačného výrazu.

Jiné chyby však mohou překladatele ovlivnit, hlavně v případech, kdy si správnou variantou sám není jistý. Využívá-li překladatel MT v podobě plug-inu či jinak, pozor by si měl dát na: distribuci členů a o všech členech, které nejsou součástí zavedených termínů, rozhodovat podle vlastních znalostí; na strojový překlad slovesných časů, pasiva a podmiňovacího způsobu, který není vždy adekvátní; na chybějící výrazy; na sémantické posuny, které by neměl kopírovat; a hlavně na aktuální členění větné a neúmyslné přebírání jeho strojové varianty.

Relativně spolehnout se dá na zavedené termíny a idiomy vyskytující se v obou jazycích. V těchto oblastech může strojový překlad navést na správný ekvivalent, či ho přímo poskytnout a překladateli pak zbývá si ho už jen ověřit. Zároveň poskytuje strojový překlad informaci o tom, jak je daný výraz nejčastěji překládán, a s tím potom úzce souvisí jeho užitelnost. To může být směrodatné například u ustálených frází.

Výstup Googlu Translate se dá označit za trochu doslovnější než výstup Bing Translatoru. Google, tedy čistě statistický překladač, se podoby zdrojového textu drží do větší míry než Bing doplněný lingvistickými prvky (*láska prochází žaludkem* → *Google: love goes through the stomach x Bing: a taste of love*).

Analýza dále ukazuje, že MS Bing Translator není dostatečně vybaven co se týče velmi odborné slovní zásoby. Proto ho coby pomocný nástroj profesionálního překladatele při

překladu vysoce odborných textů jako je například text č. 3 z oblasti medicíny nelze doporučit (*Dále pak vidáme žluté tečky odpovídající degenerovaným keratinocytům vlasového folikulu, které spolu s mazem rozšiřují jednotlivá ústí. → Bing: Then we see the yellow dots corresponding to the degenerovaným keratinocytům of the hair follicle, which together with the smear extend various mouth. x Google: Furthermore, we see yellow dots corresponding degenerated hair follicle keratinocytes, which together with sebum expand each estuary.).*

Pokud se podíváme na tematické okruhy, nejlepší výkon MT podává při překladu jednodušších, věcně-sdělných a terminologicky méně náročných textů, kde převládají oznamovací věty s neutrálním AČV. Z pokusných textů této případové studie jsou to články z časopisu Ekonom (Lotyšsko se proškrtalo k Euru. Řecko dál tápe; České firmy platí faktury nejpozději v Evropě; Za pár let největší ekonomika světa), články z časopisu Psychologie dnes (Chvála doteků; Smutek), články z časopisu 21. století (Jak opeřenci svádějí samičky?; Po stopách počasí značených křídou), bakalářská diplomová práce (Vliv stresu během konsektivního tlumočení) či formulář, který seznamuje zaměstnance s pokyny BOZP a PO. Naopak výzvou je pro strojový překlad náročnější gastronomická terminologie studijního materiálu pro studenty hotelové školy (Potraviny a výživa I) a velmi náročná terminologie odborného článku z medicínského časopisu Československá dermatologie (Dermatoskopické vyšetření v neobvyklých indikacích). Kvalita MT se dále snižuje také u stylisticky náročných Všeobecných obchodních podmínek stavební spořitelny, kde se vyskytují dlouhá a složitá souvětí, kromě toho často v kondicionálu. Při překladu soutěžního řádu nohejbalu naráží automatický překlad na oba dva výše zmíněné problémy – méně obvyklou terminologii i stylisticky náročná souvětí. Zhruba uprostřed na pomyslné linii kvality překladu se pohybuje květnatý, stylisticky zajímavý komentář z časopisu Reflex (Dvě hodiny nudy). Toto hodnocení z hlediska tematických okruhů a typů textů zhruba nastiňuje, u jakých textů může strojový překlad profesionálovi ulehčit práci a u jakých naopak nemůžeme čekat příliš velkou výpomoc, což platí především pro MS Bing Translator z důvodů uvedených výše.

Z výsledků analýzy také vyplývá, že pro profesionálního překladatele je v současnosti výhodnější používat Google Translate. Sice lze říci, že Bing dosahuje mírně lepších výstupů z hlediska syntaxe, která lépe odpovídá úzu CJ a dělá tak jeho strojový překlad čtivějším pro běžného cílového příjemce, avšak profesionální překladatel ocení zejména

rozsáhlou slovní zásobu Googlu a propracovanější překlad termínů, protože po syntaktické stránce příliš velkou výpomoc nepotřebuje.

## DISKUZE

Dalším krokem při stanovování toho, do jaké míry může být strojový překlad profesionálnímu překladateli užitečný, by mohl být výzkum zaměřený na konkrétní experiment s překladateli, kteří by byli rozděleni na dvě skupiny. Jedna by pracovala na překladu s pomocí MT, druhá nikoliv a poté by jejich překlady byly analyzovány a srovnány. V podobném experimentu by také jedna skupina překladatelů mohla využít Google Translate a druhá MS Bing Translator. Výsledky by opět byly podrobeny analýze a překladatelé by také vyplnili dotazník zaměřený na to, do jaké míry a v čem konkrétně jim strojový překlad byl nápomocen.

Rovněž by bylo zajímavé přesněji stanovit kvantitativní parametry překladu s MT, tedy například procento strojovým překladem ušetřeného času nebo průměrnou rychlost překladu např. normostrany s MT a bez něj.

Případová studie, která je základem této práce, byla provedena na 10 různorodých pokusných textech. Ty do různé míry odkryly přednosti a nedostatky strojového překladu. Pokud by podobný výzkum měl proběhnout znovu, doporučila bych vybrat jiný typ textu místo textu č. 10, bakalářské diplomové práce, jelikož tento text nebyl pro MT dostatečnou výzvou a ve srovnání s ostatními pokusnými texty vzešlo z analýzy jeho strojového překladu relativně málo poznatků.

## ZÁVĚR

Strojový překlad je v současnosti zdarma dostupný všem internetovým uživatelům. Čím dál častěji ho využívají také profesionální překladatelé, kteří díky němu mohou ušetřit určitou část práce, času a energie, kterou překladu věnují. Cílem této bakalářské práce bylo pomocí případové studie zjistit, co může moderní strojový překlad profesionálnímu překladateli nabídnout, jaké má přednosti a nedostatky, na co se profesionální překladatel může relativně spolehnout a na co si dát pozor.

Práce nejdříve přibližuje teoretické poznatky. Zabývá se vznikem a vývojem strojového překladu (kapitola 1), lingvisticky orientovaným, statistickým a hybridním strojovým překladem (kapitoly 2 a 3) a dvěma současnými, v Česku nejrozšířenějšími překladači, Googlem Translate a Microsoft Bing Translatorem (kapitola 4).

Práce dále pokračuje případovou studií, pro jejíž účely bylo vybráno 10 rozmanitých věcně-sdělných pokusných textů. Tyto texty patří do tematických okruhů ekonomie, gastronomie, medicína, právo, přírodní vědy, psychologie, publicistika, sport, technika a tlumočnictví. Studie byla zaměřena na jeden směr překladu, a to z češtiny do angličtiny. Texty byly strojově přeloženy Googlem Translate a MS Bing Translatorem a podrobeny analýze, která byla provedena na základě jednotlivých úrovní jazykového plánu (kapitola 6.1). Dle jednotlivých jazykových rovin byly výsledky analýzy také shrnuty (kapitola 6.2).

Výsledky analýzy v kapitole 6.2 ukazují, že pokud se na výstup strojového překladu podíváme z hlediska běžného příjemce (tedy i profesionálního překladatele), je možné vysledovat následující chyby a nesrovnalosti:

V rámci gramatiky jsou to chyby, které spočívají v záměně singuláru a plurálu, přechylování ženského rodu a distribuci členů, potíže překladačům způsobují také nominální konstrukce a podmět nevyjádřený. Překladové systémy také někdy tápou v překladu slovesných časů, pasiva či podmiňovacího způsobu.

Na rovině lexikální dochází k vynechávkám či přidání slov navíc. Trochu překvapivě způsobují strojovému překladu problémy jednopísmenná slova, jako jsou *a*, *i*, *u*, apod. Slovní zásoba překladačů je poměrně široká, problémy nastávají až u velmi odborných

nebo morfologicky složitě utvořených výrazů (*tungiázu, čtyřsetnásobek*). Pokud překladač nemá k dispozici žádný ekvivalent, kopíruje neznámá slova do CT v původní podobě. Kopírováním řeší překladače také výskyt jiného než zdrojového jazyka ve zdrojovém textu, např. angličtinu v českém ZT. **Vůbec nejužitečným aspektem a výhodou MT je překlad termínů.** Variabilita jejich protějšků je totiž ve srovnání s běžným lexikem poměrně nízká.

Analýza pokusných textů na rovině sémantické ukazuje, že běžné jsou menší či větší sémantické posuny. Překladače si také neumějí poradit s homonymií či polysémií. Strojový překlad má problém s monosémantizací kontextem i v rámci jedné věty. Za relativně častý rušivý element se dá považovat překlad, který nezachovává negaci a je tedy naprostým opakem originálu. Idiomy, které v podobném nebo přímo stejném znění existují i v CJ, zvládají překladače poměrně dobře. Idiomy, které se svou podobou v CJ liší, jsou přeloženy doslova a jejich přenesený význam v CT zaniká.

S komplexnějšími úrovněmi jazykového plánu naráží strojový překlad na stále více problémů. Na rovině textové nastávají potíže s kohezí a AČV. Je-li slovosled české věty příznakový ve formátu předmět – přísudek – podmět, přeložený ekvivalent takové věty odporuje logice, jelikož předmět a podmět nejsou během překladového procesu navzájem zaměněny tak, aby uspořádání věty odpovídalo anglické pevné větne struktuře.

Z pragmatiky stojí za zmínku formální nedostatky textu, které mohou kvalitu strojového překladu značně snížit.

Srovnání výkonu Googlu Translate a MS Bing Translatoru na pokusných textech ukazuje, že **Google disponuje násobně větší slovní zásobou** než Bing. Překlad Googlem vyvinutého statistického systému je ale doslovnější, málokdy se vypořádá s příznakovým slovosledem a nečekejme žádné překladatelské modifikace a transformace. **Výstup Bingu je díky lingvistickému modulu přirozenější a čtivější** než výstup Googlu, především syntax zde lépe odpovídá úzu cílového jazyka, ve větší míře můžeme očekávat správný překlad příznakového slovosledu a přenesených významů. Na druhou stranu se ale v překladu Bingu často vyskytují nepřeložená česká slova anebo výrazy navíc, někdy i zcela nesmyslné.



**Pro profesionálního překladatele se proto hodí spíše Google Translate, který mu nabízí pomoc v oblasti, ve které to nejvíc potřebuje: v terminologii.** Profesionál však ocení i další drobnosti, které mu MT může nabídnout, jako je osvěžení vlastní slovní zásoby či zpětná kontrola ustálených frází či předložkových vazeb apod.

Překladatel profesionál, který se rozhodne MT využít ve vlastní praxi, především mírně či středně pokročilý, by si měl dát pozor na distribuci členů a o všech členech, které nejsou součástí zavedených termínů, rozhodovat podle vlastních znalostí; na strojový překlad slovesných časů, pasiva a podmiňovacího způsobu, který není vždy adekvátní; na chybějící výrazy; na sémantické posuny, které by neměl kopírovat; a zejména na aktuální členění větné a neúmyslné přebírání jeho strojové varianty.

**Výpomoc strojového překladu se vyplatí při překladu jednodušších, věcně-sdělných a terminologicky nenáročných textů, kde převládají oznamovací věty s neutrálním AČV,** tedy například článků z novin a odbornějších časopisů, diplomových prací, formulářů, pokynů, návodů, oznámení atd. Se stoupající syntaktickou složitostí a odbornější terminologií kvalita strojového překladu klesá a s ní i využitelnost pro překladatele. U složitějších textů jako jsou odborné studie a právní dokumenty rozhodně nelze doporučit MS Bing Translator, u něhož je diskutabilní, zda příležitostné nabídnutí termínu vyváží čas věnovaný pročtení strojového výstupu a orientaci v něm. Google Translate lze díky širší slovní zásobě doporučit i u složitějších textů (*Dále pak vidáme žluté tečky odpovídající degenerovaným keratinocytům vlasového folikulu, které spolu s mazem rozšiřují jednotlivá ústí.* → Bing: *Then we see the yellow dots corresponding to the degenerovaným keratinocytům of the hair follicle, which together with the smear extend various mouth.* x Google: *Furthermore, we see yellow dots corresponding degenerated hair follicle keratinocytes, which together with sebum expand each estuary).*

## SUMMARY

The contemporary, relatively acceptable quality of machine translation output has opened new possibilities for its use. Even professional translators can choose to implement MT in their everyday practice to save some time and effort. However, before starting to consult its results, professional translators should be aware of its strong points and drawbacks, of its capabilities and limits. The goal of this bachelor thesis is to provide an overview of what MT can and cannot offer to the professional translators.

The first part of the thesis is composed of literature review providing background information on the historical development of machine translation, the two main approaches to it – linguistic and statistical, available information about the two most frequently employed MT systems in the Czech Republic – Google Translate and Microsoft Bing Translator, and the latest trend in MT development – hybrid machine translation.

The thesis continues with a case study. Ten various rather long texts in Czech were chosen to be automatically translated by Google Translate and MS Bing Translator into English. The output was analyzed and assessed. The results of the analysis present advantages and disadvantages of Czech to English machine translation, provide a comparison between Google and Bing, and assess which system would be more helpful for a professional.

Machine translation is by no means perfect, mistakes can be found on all linguistic levels of the language scheme including for example flaws in the translation of verb tenses, passive voice and conditionals, translations being the opposite of the original, the literal translation of idioms or problems in cohesion and functional sentence perspective. On the other hand, automatic translating systems have a considerably large vocabulary, especially of terms, and besides, instantly offer set phrases, phrases of courtesy, prepositional phrases, etc.

Whereas Bing Translator (thanks to its linguistically informed statistical approach) generally provides a better translation from the syntactic point of view, Google Translate has a broader vocabulary, especially in terms of terminology. Since terminology is a very important issue in translation, Google Translate is recommended to professional Czech to English translators.

The results of the analysis have further shown that machine translation provides the most effective help with relatively uncomplicated, pragmatic and terminologically undemanding texts consisting of indicative sentences with unmarked word order.

## BIBLIOGRAFIE

ALPAC (Automatic Language Processing Advisory Committee), 1966. *Language and machines: Computers in Translation and Linguistics – A Report by the Automatic Language Processing Advisory Committee*. Washington, D.C.: National Academy of Sciences. 138 s.

ARNOLD, Douglas et al., 1994. *Machine Translation: An Introductory Guide*. Manchester: NCC Blackwell. 240 s. ISBN 1855542463.

BROWN, Peter F. et al., 1990. A Statistical Approach to Machine Translation. *Computational Linguistics*. **16**(2), 79-85. ISSN 0891-2017.

GASPARI, Federico a HUTCHINS, W. John, 2007. Online and free! Ten years of online machine translation: origins, developments, current use and future prospects. In: *Proceedings of the Machine Translation Summit XI: September 10-14, 2007, Copenhagen, Denmark*. [Allschwil]: EAMT (The European Association for Machine Translation), s. 199-206. ISBN 9788790708160 8790708164.

Google Inc. *The Official Google Translate Blog*, 2009-2014 [online]. Dostupné z: <http://googletranslate.blogspot.cz/>

HUTCHINS, W. John, 2010. Machine translation: a concise history. *Journal of Translation Studies*. **13**(1-2), 29-70. ISSN 1027-7978.

HUTCHINS, W. John a SOMERS, Harold L., 1992. *An Introduction to Machine Translation*. London: Academic Press. 362 s. ISBN 0-12-362830-X.

CHURCH, K. W. a HOVY, E., 1993. Good applications for crummy machine translation. *Machine translation*. **8**(4), 239-258. ISSN 0922-6567.

KARLÍK, Petr, NEKULA, Marek a PLESKALOVÁ, Jana, 2002. *Encyklopedický slovník češtiny*. 1. vyd. Praha: Nakladatelství Lidové noviny. 604 s. ISBN 80-7106-484-X.

KNIGHT, Kevin, 1997. Automating Knowledge Acquisition for Machine Translation. *IA Magazine*. **18**(4), 81-96. ISSN 0738-4602.

KOEHN, Philipp, 2010. *Statistical Machine Translation*. 1st ed. Cambridge New York: Cambridge University Press. 446 s. ISBN-13 978-0-511-69132-4.

KOEHN, Philipp, 2013. Open Problems In Machine Translation. In: *YouTube* [online]. 25.03.2013. Dostupné z: <http://www.youtube.com/watch?v=6UVgFjJeFGY>. Kanál uživatele The University of Edinburgh.

Microsoft Corporation. *About Bing Translator*. [Online]. ©2014. Dostupné z: <http://www.bing.com/translator/help/#Languages>

Microsoft Corporation. *Microsoft Translator (and Bing Translator) Official Team Blog, 2007-2014* [online]. Dostupné z <http://blogs.msdn.com/b/translation/>

SCHWARTZ, Barry, 2007. Google Translate Drops Systran For Home Brewed Translation. In: *Searchengineland.com* [online]. 27. 10. 2007. [Cit. 14. 3. 2014]. Dostupné z: <http://searchengineland.com/google-translate-drops-systran-for-home-brewed-translation-12502>

Ústav pro jazyk český AV ČR. *Internetová jazyková příručka* [online]. Dostupné z: <http://prirucka.ujc.cas.cz/>

WILKS, Yorick, 2009. *Machine Translation: Its Scope and Limits*. New York: Spinger. 254 s. ISBN 978-0-387-72773-8.

## ANOTACE

Autor:	Běhálková Veronika
Katedra:	Katedra anglistiky a amerikanistiky, FF UPOL
Název česky:	Evaluace a srovnání nástrojů pro strojový překlad
Název anglicky:	Evaluation and Comparison of Machine Translation Tools
Vedoucí práce:	PhDr. Pavel Král
Počet stran s externími přílohami:	366
Počet stran bez příloh:	53
Počet znaků s externími přílohami:	1 057 221
Počet znaků bez příloh:	68 324
Počet externích příloh (CD):	2
Počet titulů použité literatury:	17

Cílem bakalářské práce je zjistit, co může moderní, volně dostupný strojový překlad v podobě Googlu Translate a Microsoft Bing Translatoru nabídnout profesionálnímu překladateli, který překládá ve směru z češtiny do angličtiny.

Práce se skládá z obecného teoretického přehledu a případové studie. Teoretická část přibližuje historii a vývoj strojového překladu, popisuje starší lingvistické a mladší statistické a hybridní metody, jak strojového překladu dosáhnout, a podává dostupné informace o Googlu Translate a Microsoft Bing Translatoru.

V rámci případové studie byla provedena analýza strojově přeloženého materiálu, která profesionálnímu překladateli poskytuje přehled předností a nedostatků strojového překladu, pro praxi doporučuje jeden překladač a dále hodnotí, u jakých typů textů může být strojový překlad profesionálovi nejvíce nápomocen.

Klíčová slova česky:

strojový překlad, statistický strojový překlad, evaluace strojového překladu, Google Translate, Microsoft Bing Translator

## **ABSTRACT**

The goal of the Bachelor thesis is to determine what the current free machine translation tools (Google Translate and Microsoft Bing Translator) can offer to a professional translator translating from Czech to English.

The thesis is composed of literature review and a case study. The theoretical part provides information on the history and development of machine translation, describes the chronologically older linguistic approach and the current statistical and hybrid approaches, and summarizes available information on Google Translate and Microsoft Bing Translator.

The case study consists of an analysis of a selected material translated by Google Translate and MS Bing Translator. The study provides an overview of the advantages and disadvantages of using MT from professional translators' point of view, recommends one of the systems for use in practice, and evaluates with what types of texts machine translation can assist professionals most effectively.

Key words:

machine translation, MT, statistical machine translation, SMT, machine translation evaluation, Google Translate, Microsoft Bing Translator