

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
FILOZOFICKÁ FAKULTA
ÚSTAV ROMANISTIKY

BAKALÁŘSKÁ PRÁCE

ÚSPĚŠNOST STROJOVÉHO PŘEKLADU Z FRANCOUZŠTINY DO ČEŠTINY
V ZÁVISLOSTI NA TYPU TEXTU

Vedoucí práce: doc. PhDr. Jan Radimský, Ph.D.

Autor práce: Aneta Kupková

Studijní obor: Francouzský jazyk pro evropský a mezinárodní obchod

Ročník: třetí

2020

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně, pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb., v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledky obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 11. května 2020

Aneta Kupková

PODĚKOVÁNÍ

Tímto bych ráda poděkovala panu doc. PhDr. Janu Radinskému, Ph.D. za odborné vedení mé práce, za jeho stálou ochotu, cenné rady a čas věnovaný konzultacím.

ANOTACE

Cílem této bakalářské práce je sledovat, jak se mění kvalita strojového překladu z francouzštiny do češtiny v závislosti na funkčním stylu textu. Práce je rozdělena do třech hlavních částí.

První teoretická část je zaměřena na vznik, historii strojového překladu a jeho využití, poté budou charakterizovány přístupy strojového překladu, vznik a historii online překladačů, představení překladačů Google Translate a Microsoft Bing Translator. Dále jsou uvedeny ruční a automatické metody hodnocení úspěšnosti strojového překladu. Nakonec je krátce popsán původ a současné rozdělení funkčních stylů v českém jazyce.

Druhá metodologická část obsahuje hypotézy očekávané úspěšnosti překladu. Dále uvádí deset modelových textů vybraných pro analýzu a jejich stručná charakteristika. Na závěr této kapitoly je předveden model hodnocení, podle něhož se následně postupuje v praktické části.

Praktická část zahrnuje analýzu úspěšnosti překladu deseti vybraných textů různého funkčního stylu pomocí hodnotícího systému BLEU.

ABSTRACT

The aim of this bachelor thesis is to test and to evaluate the quality of machine translation from French to Czech depending on the type of the text, using free translators – Microsoft, Bing Translator and Google Translate. The work contains three parts.

The first, theoretical part is divided into several chapters which are focused on the history and use of the machine translation, introduction to basic strategies of machine translation (Rule-Based Machine Translation, Statistical Machine Translation, Hybrid Machine Translation, Neural machine translation), development and history of online machine translation, presentation of two online translators - Microsoft Bing Translator and Google Translate. Furthermore, first part proposes an introduction to the MT evaluation, which is divided into human evaluation methods and automatic evaluation metrics. Finally, theoretical part includes the development and division of different functions of text in the Czech language.

The second, methodological part contains the hypothesis of success of the machine translation. Furthermore, this part presents ten texts, which are analysed in the third part and their brief characterisation. Finally, there is a model of evaluation for the practical part.

The last, practical part includes the evaluation of Google 's and Bing's translations using the automatic evaluation system - BLEU.

SEZNAM ZKRATEK

ALPAC – Automatic Language Processing Advisory Committee

BT – Microsoft Bing Translator

CAT – Computer-Assisted Translation

GT – Google Translate

MT – machine translation

NMT – neuronové sítě

RBMT – pravidlový strojový překlad

SMT – statistický strojový překlad

SP – strojový překlad

Obsah

1	Úvod.....	1
2	Teoretická východiska	3
2.1	Strojový překlad	3
2.1.1	Úvodní definice	3
2.1.2	Historie strojového překladu	4
2.1.3	Využití strojového překladu	8
2.2	Rozdělení systémů SP	9
2.2.1	Pravidlový strojový překlad (RBMT)	9
2.2.1.1	Přímý přístup.....	9
2.2.1.2	Nepřímý přístup	10
2.2.2	Statický strojový překlad.....	12
2.2.2.1	World-based statistical translation model.....	13
2.2.2.2	Phrase-based statistical translation model	13
2.2.3	Hybridní strojový překlad	13
2.2.4	Neuronové sítě	14
2.3	Veřejně dostupné překladače.....	15
2.3.1	Online překladače.....	15
2.3.2	Google Translate	16
2.3.3	Microsoft Bing Translator.....	17
2.4	Metody hodnocení strojového překladu	18
2.4.1	Ruční metody hodnocení.....	18
2.4.2	Automatické měření překladu	18
2.5	Funkční styly	22
3	Metodologický základ.....	23
3.1	Výchozí hypotézy	23
3.2	Vybrané texty	23
3.3	Model pro hodnocení.....	31
4	Výsledky hodnocení	33
4.1	Hodnocení překladačů	34
4.2	Porovnání výstupů z GT a BT v závislosti na typu textu.....	43
5	Závěr	45

6	Resumé.....	47
7	Seznam obrázků.....	49
8	Bibliografie	49
8.1	Elektronické dokumenty.....	49
8.2	Tištěné dokumenty	52
9	Přílohy.....	55

1 Úvod

V dnešním globalizovaném světě 21. století má překlad důležitou roli a můžeme jeho užití vidět ve všech odvětvích lidské aktivity. Umožňuje nám vytvářet mezinárodní politiku, rozvíjet obchodní spolupráci napříč státy celého světa, poskytuje nám přístup k informacím a pomáhá překonávat kulturní a jazykové bariéry (ZEHNALOVÁ, 2015). Z tohoto důvodu se překlad stal jednou ze zásadních oblastí jazykového průmyslu a jeho poptávka stále více a rychleji roste. Strojový překlad, díky své schopnosti překládat větší množství dokumentů za kratší dobu, hraje tedy čím dál tím větší roli. Za poslední desetiletí lidé vyvinuli řadu počítačových technologií, které člověku pomáhají vyhovět požadavkům vysoké poptávky po překladu. Obecně lze ke strojovému překladu přistupovat různými způsoby. Existují pravidlové přístupy, statické modely nebo systémy založené na neuronových sítích. Každé z těchto technických zařízení prokázalo v průběhu času své výhody, ale i nedostatky, které snižovaly kvalitu jejich výsledků. Právě kvalita výstupů a schopnost přenést požadovaný text ze zdrojového jazyka co nejvěrněji do jazyka cílového bývá častým problémem, protože komplexnost přirozeného jazyka, se všemi gramatickými pravidly a bohatou slovní zásobou, zůstává pro stroj stále složitým úkolem. Nicméně díky intenzivnímu rozvoji strojových systémů a umělé inteligence v posledních letech se výsledky strojového překladu zlepšují mílovými kroky a doba, kdy stroj přeloží text stejně kvalitně jako člověk, nebo kdy si dva lidé mluvící jiným jazykem budou navzájem rozumět v reálném čase, už možná za pár let nebude jen science fiction (TÝNOVSKÝ, 2007).

Cílem této práce je sledovat, jak se mění kvalita strojového překladu z francouzštiny do češtiny v závislosti na funkčním stylu textu. Budou zkoumány texty prostěsdělovacího, odborného, publicistického, uměleckého a administrativního stylu. Tento způsob komparace umožní studovat různorodější texty s odlišnými cíli. Zároveň budou všechny texty tematicky právně zaměřené. Na základě vybraných modelových textů budu poté sledovat kvalitu výstupů dvou nejvíce využívaných online překladačů: Google Translator od firmy Google a Bing Translator od společnosti Microsoft. Práce je rozdělena do třech hlavních částí. V kapitole Teoretická východiska představím nejprve pojem strojový překlad, krátkou historii a využití strojového překladu, poté stručně charakterizují různé přístupy strojového překladu, které byly vytvořeny v průběhu času. Následně uvedu téma online překladačů, a souběžně s ním také představím Google

Translate a Bing Translator. Poté se zaměřím na otázky a metody hodnocení úspěšnosti překladu, a nakonec stručně vysvětlím základní charakteristiku funkčních stylů. Ve druhé kapitole, nazvané Metodologický základ, nejprve zformuluji hypotézy očekávané úspěšnosti překladu, poté představím modelové texty, které budu při analýze používat, a na závěr této kapitoly předvedu model hodnocení, podle něhož budu postupovat v poslední, praktické části. V té odprezentuji získané výsledky online překladačů. V závěru porovnáám výsledky obou zkoumaných online překladačů v závislosti na typu textu.

2 Teoretická východiska

2.1 Strojový překlad

Následující kapitola nám představí úvod do tématu strojového překladu. Nejprve popíšu význam a definici odborného termínu „Strojový překlad“ a souběžně vysvětlím několik zásadních pojmů spojených s tímto tématem. Poté se zaměřím na stručný přehled historického vývoje počítačových programů sloužících k překladu. Nakonec objasním význam a využití strojového překladu.

2.1.1 Úvodní definice

Hutchins a Somers (1992, str. 3) definují Strojový překlad (SP), anglicky Machine translation (MT), jako termín dnes běžně užívaný pro označení počítačových systémů, které slouží k překladu jednoho přirozeného jazyka do jazyka druhého, bez nebo s dopomocí lidského zásahu. Dříve se také používalo označení „mechanical translation“ či „automatic translation“, nicméně tyto termíny se v anglofonních zemích objevují jen zřídka. Zato jejich ekvivalenty v jiných jazycích, například francouzské „traduction automatique“, se používají běžně (HUTCHINS W. J., 1992). Alcina (2008, strana 6) uvádí několik způsobů rozdělení strojového překladu. Jednou z možných variant je rozdělit počítačové technologie užívané v překladatelském odvětví v závislosti na míře automatizace v překladatelském procesu. V tomto rozlišení jsou rozděleny do čtyř kategorií. První je plně automatizovaný vysoce kvalitní strojový překlad (Fully Automatic High Quality Translation). Ten převádí text z jazyka zdrojového do jazyka cílového bez lidského zásahu. Opakem je pak tradiční lidský překlad (Traditional Human Translation), který nevyužívá žádný stupeň automatizace. Jako další popisuje člověkem podporovaný strojový překlad (Human-Aided Machine Translation), který pro dosažení přijatelných výsledků vyžaduje pre-editaci a post-editaci lidskými překladateli. Posledním typem je strojem podporovaný lidský překlad (Machine-Aided Human Translation), fungující na základě překladové paměti (Translation Memory) [1]. Přitom dva poslední typy mohou být podle Hutchins a Somers (1992, str. 3) shrnuty i pod společným pojmem, kterým je počítačem podporovaný překlad (Computer-Aided nebo Computer-Assisted Translation).

2.1.2 Historie strojového překladu

V tomto stručném přehledu historického vývoje strojového překladu objasním hlavní etapy vzniku a evoluce snahy o vytvoření počítačového programu sloužícího k překladu přirozených jazyků. Představím nejzásadnější vědecké výzkumy a projekty, které začínají ve 40. letech 20. století s prvními mechanickými vícejazyčnými překladači, a vedou až k veřejně dostupným překladačům 21. století.

Počátky

S první myšlenkou o vytvoření strojového překladu za užití „univerzálního číselného kódu“ (universal numerical code) přišli v 17. století René Descartes a Gottfried Wilhelm von Leibniz. V polovině téhož století Cave Beck, Athanasius Kircher a Johann Becher, inspirováni hnutím „universal language“, publikovali práce o uměle vytvořeném jazyce, který by fungoval na bázi logických principů a symbolů, a tím by umožňoval komunikaci prostřednictvím jednoho jednovýznamového jazyka. Později roku 1668 John Wilkins ve svém díle „Essay towards a Real Character and a Philosophical Language“ zmínil pojem interlingua (HUTCHINS W. J., 1992). Nicméně kvůli absenci potřebných technologií myšlenky 17. století zůstaly čistě teoretické.

Představa se stala skutečností až koncem první poloviny 20. století, kdy se s příchodem počítačů se opět zvýšil zájem o vývoj strojového překladu. V roce 1933 si Francouz Georges Artsrouni podává přihlášku na patent přístroje „general-purpose machine“, který funguje i jako mechanický slovník. Ve stejném roce přichází s podobným patentem i ruský vědec Petr Trojanskij (HUTCHINS W. J., 1992). Podstatný rozvoj zažil strojový překlad na konci druhé světové války. Zvyšující se napětí mezi USA a Sovětským svazem vyústilo ve studenou válku, která rozdělila svět na dvě části. Překlad vědeckých a technických dokumentů se stává prioritou pro oba bloky. Zmíněné okolnosti a příchod prvních počítačů dopomohly k raketovému nárůstu zájmu o výzkum v oblasti strojového překladu. Prvotní myšlenka výzkumu SP pochází od britského vědce Andrewa Bootha a amerického kryptologa Warrena Weavera, kteří roku 1947 navrhují inovativní myšlenku užití nově vynalezených počítačů pro překlad přirozených jazyků. O rok později zahajují výzkum společně s Richardem H. Richensem, ve kterém zkoumají uplatnění morfologické analýzy pro vývoj mechanického slovníku. Weaverovo

memorandum „Translation” [9] pro Rockefellerovu nadaci se stalo jedním z průlomových děl na konci 50. let a dalo podnět k výzkumu strojového překladu v USA. Tehdejší ředitel v Rockefellerově nadaci využíval hlavně svých kryptologických znalostí a přirovnával strojový překlad k dešifrování kódů (HUTCHINS W. J., *Machine translation: a concise history*, 2006). Ve své práci nastínil vizi do budoucna a přiblížil metody využívající vědní disciplíny, jako jsou kryptografie, statistika, teorie informace, logika a jazykové univerzálie pro vyřešení problému s mnohoznačností (HUTCHINS W. J., *Machine translation: a concise history*, 2006).

Georgetown experiment

Roku 1954 proběhla první veřejná ukázka strojového překladu. Leo Dostert z Georgetownské univerzity ve spolupráci s IBM (International Business Machines), mezinárodní technologickou společností sídlící ve Spojených státech amerických [8], předvedli program založený na šesti gramatických pravidlech a slovní zásobě 250 slov, který zahrnoval překlad vybraných padesáti vět z ruštiny do angličtiny. Pozitivní výsledky této ukázky zaujal americký tisk a dal impuls mnohým výzkumům nejen na akademické půdě slavných univerzit v USA, ale po celém světě, a především v tehdejší Sovětské svazu. Potenciál strojového překladu poté přilákal četné sponzory, mezi které patřila i vláda Spojených států amerických (HUTCHINS W. J., 1992).

V 50. a 60. letech byly dále zformulovány tři základní metody pravidlového strojového překladu. První metodou a první generací strojového překladu je přímý přístup, který je založen na základních gramatických pravidlech. Tento model dominoval především v období před zveřejněním zprávy ALPAC, tedy v letech 1956–1966. Druhá generace SP představovala pravidlový přístup nepřímý, tedy metody využívající hlubší lingvistickou analýzu. V principu tohoto přístupu vznikly dva způsoby - transferový přístup řídicí se morfologickou a syntaktickou analýzou a interlingvální metoda, která převáděla zdrojový jazyk do uměle vytvořeného neutrálního jazyka „interlingua“, a z něj do cílového jazyka (KOEHN, 2010). Obě metody nepřímého přístupu se vyvíjely zejména až v „post alpatském“ období (HUTCHINS W. J., *Machine translation: History of Research and Applications*, 2015).

Zpráva ALPAC

Georgetownský experiment a optimismus 50. let nasadily vysoké očekávání. Nicméně vidina plně automatického překladu funkčního v rámci několika let byl spíše překladatelský sen, než realita 60. let. Komplexnost přirozeného jazyka působila pravidlovým systémům mnoho potíží na sémantické i syntaktické rovině, a navzdory vysokým nákladům systémy nevykazovaly očekávanou kvalitu výstupů v porovnání s lidskými překladateli, a proto začal být strojový překlad kritizován čím dál tím více (HUTCHINS W. J., Machine translation: a concise history, 2006). Zmíněné okolnosti vedly k sestavení vládní komise „Automatic Language Processing Advisory Committee“ neboli ALPAC, aby zhodnotila následující vyhlídky strojového překladu (HUTCHINS W. J., Machine translation: History of Research and Applications. , 2015). Ta roku 1966 přináší pro budoucnost strojového překladu nevalné výsledky. Zpráva popisuje strojový překlad jako nevýhodný a doporučuje se do budoucna zaměřit spíše na počítačové programy, které by ulehčily práci lidským překladatelům, než investice do plně automatických systémů. Studie také mimo jiné poukázala na fakt, že strojový překlad není rychlejší, ale ani levnější v porovnání s lidskými překladateli. Kvůli neúspěchu vláda téměř okamžitě stáhla veškeré financování. Způsobený nedostatek finančních prostředků na výzkum v oblasti strojového překladu poté zapříčinilo úpad na budoucí desetiletí (KOEHN, 2010).

První systémy pro komerční účely 70.–80. léta

70. a 80. léta zůstala ve znamení nepřímé metody překladu. Přes menší úpadek zájmu o strojový překlad na koci 60. let, který byl zapříčiněn kritickou zprávou ALPAC, se na začátku následující dekády začaly opět objevovat překladatelské systémy, tentokrát spíše pro komerční účely. Jeden z nejznámějších systémů vznikl v roce 1976 v rámci projektu TAUM (Traduction Automatique de l'Université de Montréal) na univerzitě v kanadském Montrealu. Météo se využíval na překlad počasí a fungoval na bázi transferového přístupu. Tento systém představoval první plně funkční překladový systém (HUTCHINS W. J., Machine translation: History of Research and Applications. , 2015). Mezi další úspěšné programy můžeme zařadit také americký systém pro překlad angličtina/ruština. Sytran byl používán od roku 1970 letectvem Spojených států, o 6 let později ho odkoupila i Evropská komise v jazykové kombinaci francouzština/angličtina.

V následujících 80. letech bylo rozvinuto mnoho dalších systémů s komerčními účely jako Logos, METAL nebo japonský překladatelský systém pro jazykovou dvojici japonština/angličtina (KOEHN, 2010).

90. léta

Převrat ve strojovém překladu nastal v průběhu 90. let. Po dlouholeté převaze pravidlových systémů se začaly, zásluhou rozsáhlého rozvoje počítačových technologií, objevovat korpusové metody. Největší rozmach v 90. letech zažil statický překlad. Jeho první model vzešel v roce 1988 od americké mezinárodní společnosti IBM z projektu Candide. Definitivní podoba tohoto programu byla spuštěna o pět let později. Systém, založený na překladu jednotlivých slov (world-based statistical translation model), sklídl úspěch, a tak se statický překlad stal hlavním předmětem výzkumu v následujících letech. V průběhu času se přešlo z překladu po slovech na překlad po frázích, který více přihlížel ke kontextu překládané věty. Následně vznikl hybridní přístup, jehož cílem bylo kombinovat to nejlepší ze statických a pravidlových modelů (HUTCHINS W. J., Machine translation: a concise history, 2006). Díky rozmachu osobních počítačů během 90. let byly rozvíjeny také programy na podporu lidských překladatelů (KOEHN, 2010), tedy softwaru fungující na základě překladatelské paměti. Na trh se dostalo hned několik systémů, například od firmy Trados nebo SLD. Období 90. let bylo dobou rozkvětu strojového překladu. Nejenže se zrodily a vyvíjely nové přístupy plně automatického překladu, ale také byl zahájen výzkum překladu mluvené řeči, nebo vznikaly online překladače, z nichž první uvedla společnost CompuServe v roce 1995 (HUTCHINS W. J., Machine translation: History of Research and Applications. , 2015).

Současnost

V dnešním světě technologií a umělé inteligence můžeme v oboru strojového překladu sledovat obrovský krok kupředu. Od roku 2016 se většina výzkumů studující toto odvětví zaměřila především na neuronové sítě, tedy systémy inspirované biologickým neuronem lidského mozku, fungující na bázi výpočetních modelů umělé inteligence [12]. Na tyto systémy přešla většina online překladačů, protože všeobecně přinášejí kvalitnější výsledky, než předchozí systémy. Avšak ani neuronové sítě nejsou

prozatím schopné překonat kulturní a společenské odlišnosti lidské komunikace, rozeznat ironii nebo pochopit humor.

2.1.3 Využití strojového překladu

Využití můžeme rozdělit do tří skupin. Zaprvé informativní překlad neboli „gisting“, jehož hlavním účelem je v krátkém časovém úseku poskytnout hrubý překlad informace ze zdrojového jazyka (ZEHNALOVÁ, 2015). Díky službám online překladačů si každý může přeložit konkrétní výraz, celé věty, dokumenty, ale dokonce i celé webové stránky [23]. Nejužívanější online překladač, Google Translate od společnosti Google, umožňuje překlad mezi 109 jazyky, své služby poskytuje zdarma na internetových stránkách nebo v aplikaci. Google Překladač je využíván až 500 milióny uživatelů po celém světě a přeloží zhruba kolem 100 bilionů slov za den [15]. Druhým způsobem využití jsou překlady s následnou post editací, kdy strojový překlad vyhotoví meziproduct, který je pak určený na následnou redakci (ZEHNALOVÁ, 2015). Do této skupiny bychom mohli také zařadit počítačové programy podporující lidský překlad CAT, fungující na bázi překladové paměti, která zefektivňuje práci a šetří překladateli čas. CAT je často využívaným nástrojem pro firmy, které se snaží snížit náklady na překlad. Příkladem může být německá automobilová společnost BMW, která používá programy CAT už 25 let pro překlad uživatelských a servisních manuálů do 27 jazyků, a každoročně tak ušetří statisíce eur [22]. Třetí možností je tzv. „subdomain translation“, užití strojového překladu pro komunikaci v konkrétní dílčí disciplíně, jako například předpověď počasí (ZEHNALOVÁ, 2015).

2.2 Rozdělení systémů SP

V této podkapitole stručně charakterizují základní metody, jak lze přistupovat ke strojovému překladu. Nejprve se budu věnovat způsobu pravidlového strojového překladu, anglicky Rule-Based Machine Translation (RBMT). V této části vysvětlím rozdělení přímého přístupu (Direct Approach) a nepřímého přístupu (Indirect Approach). Dále představím metodu strojového překladu neboli Statistical Machine Translation (SMT) a jeho dvě hlavní metody: world-based statistical translation model a phrase-based statistical translation model. Následovat bude hybridní překlad (HMT), a nakonec přiblížím nejnovější přístup strojového překladu, který má v současnosti na trhu největší potenciál - překlad za pomoci neuronových sítí (NMT).

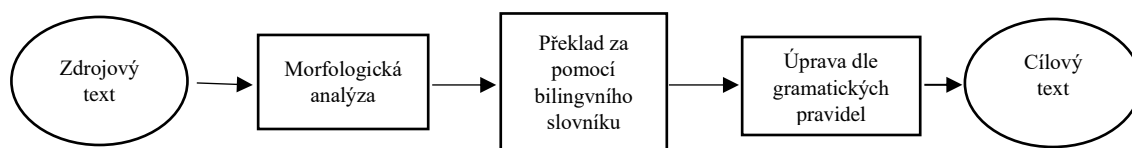
2.2.1 Pravidlový strojový překlad (RBMT)

Pravidlový strojový překlad funguje na základě bilingvních slovníků a souhrnu lingvistických pravidel (TÝNOVSKÝ, 2007). RBMT se dále dělí na přístup přímý (direct approach) a přístup nepřímý (indirect approach). Druhý zmíněný využívá dva modely - transferový model (transfer model) a interlingvální model (interlingua model) (HUTCHINS W. J., 1992). Na vývoj RBMT systémů je zapotřebí dlouhé časové období a tým lingvistů, kteří vytvoří souhrn pravidel pro jednotlivé fáze. Právě dlouhotrvající vývoj je hlavní nevýhodou této metody. Naopak výhodou těchto systémů je možnost doplňování gramatických pravidel na základě zjištěných nedostatků (TÝNOVSKÝ, 2007).

2.2.1.1 Přímý přístup

Přímý přístup patří do tzv. první generace strojového překladu. Jedná se o systémy, které byly vyvíjeny hlavně v 50. letech a na počátku 60. let 20. století, a tím se řadí mezi nejstarší systémy SP. Překlad přímý probíhá pouze na základě morfologické analýzy, při níž se každému slovnímu tvaru v textu přiřadí jejich reprezentativní slovníková podoba [16]. Během analýzy systém rozpozná a odstraní všechny koncovky slov, a tím je převede do jejich základních tvarů, které jsou poté, za pomoci bilingvních

slovníků a vložených gramatických pravidel, přeloženy do cílového jazyka (HUTCHINS W. J., 1992).



Obrázek 1 - Znárodnění přímého přístupu (HUTCHINS W. J., 1992)

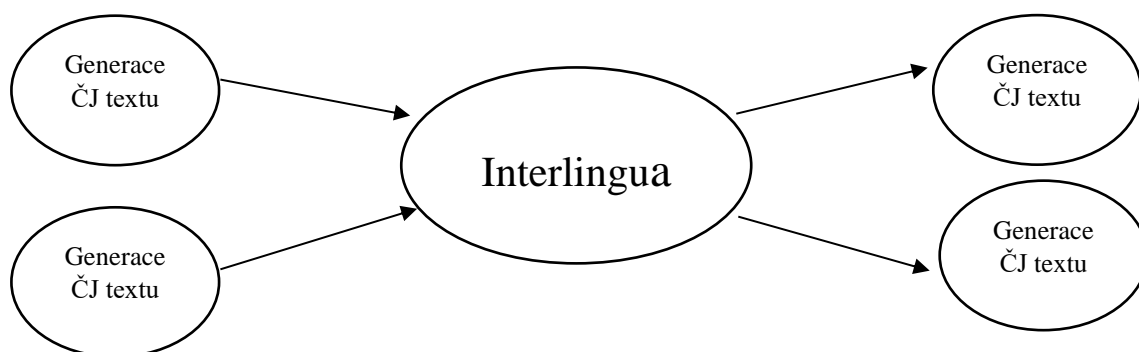
Tato metoda nicméně nebere v potaz syntaktickou analýzu textu ani sémantické vztahy ve větě, a proto často dochází k chybnému překladu na významové rovině a špatné adaptaci gramatické a sémantické stavby vět v cílovém jazyce (HUTCHINS W. J., 1992).

2.2.1.2 Nepřímý přístup

Neúspěch prvních systémů vedl k dalšímu vývoji a zlepšení pravidlového překladu. Tento přístup se více zaměřoval na analýzu textu ve zdrojovém jazyce a přenesení správného významu za pomoci určitého mezičlánku, který by provedl syntaktický a sémantický rozbor překládaného textu. Přístup nepřímý má v zásadě dvě možné varianty - interlingvální model a transferový model (HUTCHINS W. J., 1992).

2.2.1.2.1 Interlingvální model

Historiograficky první systém nepřímého překladu, interlingvální model, funguje na bázi uměle vytvořeného neutrálního jazyka „interlingua“. Zdrojový text je analyzován a převeden do abstraktního zastoupení nezávislého na jazyce. Díky této fázi získá všechny potřebné informace a následně vygeneruje překlad do cílového jazyka. Výhodou interlingvální neboli mezi jazykové metody je fakt, že neutrální zastoupení je nezávislé na jazyce a schopné tudíž překládat mezi více jazyky najednou. Cílový jazyk nemá v podstatě žádný dopad na průběh analýzy, jejímž hlavním cílem je pouze převést zdrojový text do čistě neutrálního a jazykově nezávislého jazyka „interlingua“, z kterého lze poté přeložit text do více jazyků současně. Z tohoto důvodu je tento model vhodný hlavně pro vícejazyčné strojové systémy. Na druhou stranu přesné vymezení interlingvy zůstává obtížné a definovat ji přesně nelze ani mezi jazyky sobě velmi podobnými, jako například románské jazyky (HUTCHINS W. J., 1992).



Obrázek 2 - Interlingvní model se 2 jazykovými páry (HUTCHINS W. J., 1992)

2.2.1.2.2 Transferový model

Druhá forma nepřímého přístupu RBMT se nazývá transferová metoda. Tyto systémy užívají dvojjazyčné transferové moduly (HUTCHINS W. J., 1992). Vstupní věta je na příslušnou rovinu popisu převedena na základě morfologické a syntaktické analýzy pomocí ručně implementované gramatiky. Po transferu je přeložena opět podle gramatických pravidel cílového jazyka prostřednictvím morfologické a syntaktické analýzy, a následně je vygenerován překlad vstupní věty v cílovém jazyce (BOJAR, 2012) [13]. Jedná se tedy o „transfer“ neboli přenos konceptů zdrojového jazyka do konceptů jazyka cílového. Na rozdíl od interlingválního modelu, zastoupení v podobě „transferu“, je jazykově vázaná a specifická pro každý jednotlivý jazyk, a hodí se tak spíše pro vývoj dvojjazyčných systémů (HUTCHINS W. J., 1992).



Obrázek 3 - Transferový model se 2 jazykovými páry (HUTCHINS W. J., 1992)

2.2.2 Statický strojový překlad

Až do 80. let 20. století dominovaly na trhu strojovému překladu zejména pravidlové metody, avšak rozsáhlý rozmach počítačů během 80. a 90. let umožnil vývoj nových metod na poli strojového překladu a vznik statického strojového překladu (HUTCHINS W. J., *Machine translation: a concise history*, 2006). Statická metoda funguje na bázi statických výpočtů a paralelních vícejazyčných korpusů, ve kterých najdeme originální texty v jednom jazyce a jejich překlady v jazyce druhém, zarovnané do stejných segmentů (odstavce, souvětí, fráze nebo i na slova). Tyto korpusy můžeme rozdělit zaprvé podle počtu jazyků na dvojjazyčné a vícejazyčné paralelní korpusy, a zadruhé v závislosti na směru překladu – jednosměrný, který obsahuje například francouzské originály a jejich české překlady nebo obousměrný paralelní korpus, který zahrnuje jak francouzské, tak české originály s jejich překlady [21]. Kvalita a přesnost překládaného textu pomocí SMT závisí především na dostupném množství paralelních textů a jazykové kombinaci. Například kombinace jazyků francouzština-angličtina bude vykazovat nepochybně lepší výsledky v porovnání s překladem angličtiny do čínštiny (KOEHN, 2010). Hlavní výhodou statického překladu je čas ušetřený na vývoji systémů, jelikož paralelní korpusy získávají data z volně dostupných zdrojů na internetu, a tedy není nutný, oproti pravidlovému překladu, několikaletý vývoj těchto systémů (KOEHN, 2010). SMT vytváří překlad na základě dvou modelů – prvním je model překladový, který pracuje s dvojjazyčným korpusem a zprostředkovává překlad slova nebo fráze ze zdrojového jazyka do cílového. Druhý je model cílového jazyka, který využívá pouze jednojazyčný korpus a již přeložená slova či věty upravuje do nejpravděpodobnější interpretace v cílovém jazyce bez ohledu na vstupní text zdrojového jazyka. Například upravuje slovosled věty (TÝNOVSKÝ, 2007). Překladová část SMT spočívá ve statistickém výpočtu pravděpodobnosti pomocí algoritmu, který ve všech dostupných (KISHORE, 2002) paralelních textech přeložených člověkem vyhledá různé možnosti překladu daného slova (fráze). Následně vypočítá míru jejich výskytu, tedy pravděpodobnost, jaká varianta z nabízených řešení je nejvhodnější pro překlad našeho slova (fráze) (KOEHN, 2010).

Statický překlad se dále člení na dva přístupy: world-based statistical translation, model založený na překladu jednotlivých slov a phrase-based statistical translation, který překládá zdrojový text po frázích (KOEHN, 2010).

2.2.2.1 World-based statistical translation model

Prvotní model statického strojového překladu založený na jednotlivých slovech vzešel z projektu Candide americké mezinárodní společnosti IBM (International Business Machines) [8] v letech 1980–1990. Tato metoda se sice již nevyužívá, nicméně principy, na kterých byla postavena, uplatňují dodnes její nástupci.

Model pracuje na základě překladu slova jako samostatné jednotky izolované od zbytku věty. Vyhledá slovo v paralelním korpusu, spočítá četnost výskytu slova, a poté vybere nejvíce užívanou slovní jednotku pro překlad do cílového textu (KOEHN, 2010). Nicméně fakt, že tento model nezohledňuje slova v jejich kontextu a narušuje tak různá slovní spojení, způsobuje často významové odchýlení překládaného textu.

2.2.2.2 Phrase-based statistical translation model

Doposud popisovaný model založený na slovech však nezpracovává postavení slov v kontextu věty, a nebere v potaz ani sousloví nebo ustálená slovní spojení. Výsledný překlad v cílovém jazyce se tedy může značně významově odlišovat od původního textu. Statické systémy překládající současně celé fráze tento zásadní nedostatek vylepšily. Frázový překlad nejprve rozdělí větu na menší překladové jednotky (fráze). Tyto překladové jednotky mohou pojmout více vstupních slov a následně i vyprodukovat větší počet slov. Poté jsou fráze na základě paralelních korpusů a výpočtu pravděpodobnosti přeloženy do cílového jazyka, a nakonec dochází k úpravám slovosledu (KOEHN, 2010). Překlad pomocí frází vykazuje lepší výsledky díky tomu, že slova jsou soustředěna do frází, slovního spojení více slov, které spolu gramaticky nebo sémanticky souvisí [3]. Dopomáhá tak systému lépe se vypořádat s více významy nebo homonymy.

2.2.3 Hybridní strojový překlad

Oba již zmíněné systémy mají své výhody, ale také nedostatky. Hybridní strojový překlad propojuje oba tradiční přístupy - pravidlový a statický překlad. Hybridní překlad umožňuje uplatnění gramatických pravidel před převodem ze zdrojového jazyka do cílového a statické metody například volbu slovního obsazení. Obecně tato kombinace dopomáhá k lepším výsledkům, než vykazují jednotlivé systémy samy o sobě [14].

2.2.4 Neuronové sítě

V posledních letech převzaly hlavní úlohu strojového překladu především neuronové systémy. Tento nejnovější přístup funguje na základě umělých neuronových sítí (Neural Machine Translation NMT), jejichž základní výpočetní jednotkou jsou umělé neurony, které jsou vzájemně propojeny a předávají si signály mezi sebou. Tyto umělé neurony mají vždy libovolné množství vstupů, ale jenom jeden výstup [6]. S první teoretickou myšlenkou o umělých neuronových sítích přišli v roce 1943 američtí vědci Warren McCulloch a Walter Pitts, nicméně jejich úvahu bylo možné zrealizovat až s výpočetní technikou 21. století [23] (WARREN, 1943). Neuronové systémy fungují na bázi výpočetních modelů umělé inteligence a jsou inspirovány biologickým neuronem lidského mozku [6]. Pro neuronový překlad je důležité mít k dispozici velké množství dat k trénování sítě. Tato data lze získat z wikipedie, elektronické literatury, legislativy EU a dalších volně dostupných zdrojů na internetu, kde ke každé zdrojové větě existuje její překlad v cílovém jazyce. Neuronový systém můžeme označit jako „chytřejší jazykový model“, který na rozdíl od jazykového modelu užitého ve statickém přístupu zná celou historii výstupové věty a umí uhodnout pravděpodobnost následujícího slova, a pravděpodobnost celé věty definuje jako součin všech přidaných slov.

Oproti statické metodě, která nebere v potaz souvislost mezi jednotlivými překladovými jednotkami (frázemi), vykazuje NMT lepší výsledky v různých ohledech, například vytváří plynulejší text nebo vyhodnocuje více odpovídající lexikální obsah [12]. Na druhou stranu neuronové sítě mají prozatím i své mezery. Mezi hlavní nedostatky patří například časově náročné trénování sítí, problém s překladem řídce užívaných slov nebo neúspěch překladu některých slov ve větě, a tedy jejich úplné vynechání [25].

2.3 Veřejně dostupné překladače

Úkolem této bakalářské práce je sledovat kvalitu strojového překladu pomocí veřejně dostupných překladačů Google překladač a Microsoft Bing, proto se v této kapitole zaměřím právě na tyto systémy. Nejprve krátce obecně popíši online překladače a jejich historický vývoj. Poté představím jednotlivé překladače, které jsou předmětem mého výzkumu. Jako první uvedu společnost Google a jeho online překládací službu Google Překladač, anglicky Google Translate. Jako druhý představím online překladač Bing od firmy Microsoft.

2.3.1 Online překladače

Veřejně dostupné online překladače začaly vznikat během rozkvětu strojového překladu v 90. letech. Tento rozvoj nastal zásluhou rozsáhlého rozmachu počítačových technologií, a tedy i rozšířením vlastnictví osobních počítačů v široké veřejnosti. První online překládatelská služba byla nabízena v roce 1988 od společností Systran. Služba byla poskytována ve třech jazykových kombinacích: francouzština-angličtina, angličtina-francouzština a němčina-angličtina. Uživatel odeslal text na překlad ze svého osobního počítače a výsledky mu byly zaslány zpět. Systém, fungující pouze ve Francii, zvládl přeložit 22 řádků (okolo 260 slov) za minutu a zákazníkovi bylo účtováno 1,20 \$ (přibližně tedy 26 Kč) za jednu stránku překladu. První celosvětový online překladač přivedla na trh firma CompuServe v roce 1992. Systém překládal mezi angličtinou, francouzštinou, němčinou a španělštinou v rychlosti 3 000 slov za minutu a byl omezen pouze pro registrované odběratele webové stránky CompuServe. Kvalita výstupů překladače byla velice nízká, a proto až 25 % zákazníků se po prvním použití k překladači už nevrátilo. Firma si tedy udržela pouze zákazníky, kterým záleželo více na rychlosti překladu, než na kvalitě výstupů. Scéna online překladačů se značně změnila roku 1997 při spuštění prvního bezplatného online překladače pro širokou veřejnost. Babel Fish vznikl díky spolupráci společnosti Systran a internetového vyhledávače AltaVista. Umožňoval přístup k překladu každému uživateli internetu a překládal mezi pěti hlavními evropskými jazyky. Online překladače, tedy plně automatický strojový překlad, se těšily čím dál tím větší oblibě ze strany zákazníků, a to hlavně díky rychlosti překladu a bezplatnému přístupu všem uživatelům. Od roku 2000 se objevilo na trhu hned několik

nových online překladatelských systémů, které nabízely uživatelům své služby. Mimo Babel Fish a CompuServe to byly například FreeTranslation, Gist-inTime, MT Ave, My Translator, PARS, ProMT a Reverso. Dohromady bylo spuštěno až 30 nových online překladačů, které nabízely různé jazykové kombinace jazyků z celého světa (GASPARI, 2007).

2.3.2 Google Translate

Překladač Google od společnosti Google Inc. byl spuštěn pro veřejnost v dubnu roku 2006. O dva roky později byla přidána jazyková možnost češtiny. Jedná se o nejužívanější bezplatnou online překládací vícejazyčnou službu, která funguje od roku 2016 na bázi neuronových sítí, a v současné době (duben 2020) poskytuje překlad mezi 109 jazyky [7]. Nejpoužívanější online překladač na světě poskytuje služby na internetových stránkách <https://translate.google.cz/?hl=cs> nebo v mobilní aplikaci podporované jak pro operační systém Android, tak pro systémy iOS od společnosti Apple. Na webových stránkách nabízí překladač automatické rozpoznávání jazyka a umožňuje překlad textu o rozsahu více než 300 000 znacích (ŠVARCOVÁ, 2016). Google Translate přeloží zhruba kolem 100 bilionů slov za den, a je využíván až 500 milióny uživateli, tedy zhruba 6.5 % celé světové populace.

Google poskytuje překlad konkrétního slova, výrazu, vět, celých dokumentů i webových stránek. Výhodou tohoto překladače je možnost vložení přímo celých dokumentů, aniž by uživatel musel kopírovat a vkládat text k překladu přímo do určitého pole na internetové stránce. Další výhodou je nepochybně služba Google Translate Community, která byla připojena v roce 2015. Díky ní se mohou všichni uživatelé podílet na vylepšení překladače tím, že překlad opraví nebo doplní o nové slovo (ŠVARCOVÁ, 2016). Mobilní aplikace umožňuje konverzaci mezi dvěma jazyky, a tak si můžou dva lidé mluvící jiným jazykem navzájem rozumět v reálném čase. Vyslovená věta ve zdrojovém jazyce se nahraje, přeloží do požadovaného jazyka, a poté si ji uživatel může dokonce i přehrát v cílovém jazyce, pokud je hlasový výstup pro danou řeč k dispozici. Dále lze v aplikaci přeložit například text vygenerovaný z fotografie (aplikace Google).

2.3.3 Microsoft Bing Translator

Překladač Bing od společnosti Microsoft je bezplatná překladatelská služba, která byla spuštěna v roce 2007 jako součást internetového vyhledávače Bing. Pomocí neuronových sítí začal Bing překládat o dva roky později, než jeho hlavní konkurent Google, tedy v roce 2018 [11]. V současnosti překládá mezi 69 světovými jazyky, včetně češtiny. Své služby firma Microsoft nabízí přes internetové stránky <https://www.bing.com/translator>, ale také v mobilní aplikaci určené pro systémy Android i iOS.

Na webových stránkách nabízí překladač automatické rozpoznávání jazyka a umožňuje překlad o rozsahu maximálně 5000 znaků. Oproti Googlu překladač Bing nedisponuje podporou dokumentů, a není tedy možné vkládat celé dokumenty. Uživatel tudíž musí zkopírovat zdrojový text a vložit ho přímo do daného pole. V mobilní aplikaci společnost Microsoft poskytuje překlad jak psaného, tak mluveného slova, nebo překlad fotografií, a to i do českého jazyka [10].

2.4 Metody hodnocení strojového překladu

Hlavním předmětem této kapitoly je hodnocení strojového překladu. Stejně jako u lidského překladu, tak i u překladu strojového je nutné testovat kvalitu jeho výstupů. Různé metody hodnocení doprovázely SP systémy již od počátku jejich vzniku. Nejprve byly vyvíjeny ruční metody hodnocení, ty se však po čase ukázaly časově i finančně náročné. Proto se vědci od 90. let 20. století začali soustředit na automatické měření překladu a vývoj systémů, které by člověku urychlily práci. Tato kapitola je rozdělena na dvě části - Ruční metody hodnocení a Automatické měření strojového překladu.

2.4.1 Ruční metody hodnocení

Podle Koehna (2010, str. 218, 219) je ruční metoda způsob hodnocení strojového překladu vypracovaného člověkem. Bilingvní posuzovatelé mohou díky dobré znalosti obou jazyků posoudit kvalitu přeloženého textu na základě svých znalostí. Monolingvní hodnotitelé na kontrolu využívají referenční překlad. Většinou se při této metodě hodnotí každá věta zvlášť, ale bere se v potaz i celkový kontext textu. Ruční metoda je sice časově i finančně náročnější než automatické měření, zato je stále metodou nejspolehlivější. Tato metoda posuzuje strojový překlad na základě dvou kritérií. Tím prvním je „plynulost“ (fluency), která zahrnuje gramatickou stránku překládaného textu, výběr vhodné slovní zásoby a celkovou plynulost textu v cílovém jazyce. Druhé kritérium je „vhodnost“ (adequacy), tedy zdali překladač přenesl správně význam textu a jeho hlavní myšlenky. Obě kritéria se hodnotí na stupnici od jedné do pěti, přičemž nejvyšší evaluace je pátý stupeň. Ten označuje bezchybný text s plně přeneseným významem. Po posouzení všech vět textu a přiřazení odpovídajícího hodnocení se výsledek spočítá výpočtem geometrického průměru všech hodnocení všech ohodnocených vět (KOEHN, 2010).

2.4.2 Automatické měření překladu

Ruční metody představují větší finanční výdaje a potřebují delší časové období na zpracování posudků kvality strojového překladu. Tyto dva faktory vedly ke vzniku automatických měření kvality SP. Ty se tradičně nazývají metrikami strojového překladu, anglicky MT evaluation metrics (BOJAR, 2012). Metriky SP se dostaly na trh během 90.

let. Doby, kdy se naplno rozvíjel SMT a vědci potřebovali ohodnotit své novorozené systémy i několikrát denně, tudíž ruční metody, se pro ně staly značně nevýhodné. Upřednostňovali spíše systémy, které jim kvalitu jejich překladačů ohodnotily hned a za malé náklady (KOEHN, 2010).

Automatické metriky používají k evaluaci kvality jeden nebo více referenčních textů. To jsou předlohy překlady od lidského překladatele, se kterými porovnají kandidátský text konkrétního překladačského systému. Většina metod poté měří shodu a počet slov mezi referenčním a kandidátským překladem. Na základě svých výpočtů nakonec vyhodnotí kvalitu testovaného překladače. Výsledky tedy podají v několika vteřinách a nejsou tak finančně náročné. Nicméně co se týká kvality a přesnosti jejich výstupů, automatické metriky jsou stále v porovnání s ručními metody méně spolehlivé. Proto byly vytvořeny i metody na kontrolu kvality těchto automatických systémů, jako například Pearsonův korelační koeficient (KOEHN, 2010). Metrik pro měření kvality překladu existuje mnoho, a proto zde popíšeme ty nejužívanější, přičemž nejvíce se zaměříme na metodu BLEU, která bude používána následně v praktické části pro ohodnocení kvality výstupů překladačů Google a Bing.

- BLEU

Metoda BLEU (BiLingual Evaluation Understudy) byla vytvořena v roce 2001 firmou IBM, a stala se jednou ze zásadních metod pro měření kvality strojového překladu. Tato metoda značně dopomohla k urychlení vývoje mnoha překladačských systémů a k vylepšení jejich jednotlivých verzí (BOJAR, 2012). Při hodnocení porovnává n-gramy kandidátského překladu s n-gramy jednoho nebo více referenčních překladů lidského překladatele a vypočítává míru jejich shody. Na výpočet celkového hodnocení využívá dvě veličiny: modifikovanou n-gramovou přesnost a pokutu za krátkost. Kandidátský překlad může získat od BLEU výsledné hodnocení na škále 0 až 1, ale je možné ho zapisovat i jako 0 až 100 (BOJAR, 2012). Čím blíže k 1 (100), tím lepší hodnocení strojového překladu. Je vždy lepší porovnávat kandidátský text s více referenčními překlady. Ty se totiž mohou lišit ve slovním obsazení i pořadí slov, a tak čím více referenčních textů máme k dispozici, tím lepší je hodnocení BLEU. K hodnocení lze používat mimo unigramů, hodnocení na základě jednoho slova, i bigramy (dvojice slov), trigramy (trojice slov), n-gramy (ŠTANCEL, 2007). Díky tomu, že při posuzování překladů používá systém BLEU n-gramy, dokáže správně ohodnotit plynulost kandidátského překladu (MACHÁČEK, 2012).

Modifikovaná n-gramová přesnost, anglicky Modified Unigram Precision, funguje na základě unigramů, kdy každé slovo zastupuje jeden unigram. Tuto metodu si představíme na základě příkladu (příklad 1) převzatého z práce BLEU: a Method for Automatic Evaluation of Machine Translation. Nejprve určíme unigramovou přesnost. Ta se spočítá na základě porovnání celkového počtu slov kandidátského překladu a referenčních textů. Když se počet slov shoduje, udělí BLEU nejvyšší hodnocení v našem příkladu 7/7. Modifikovaná n-gramová přesnost bere v potaz i pořadí slov, a tak by kandidátský text v našem příkladu získal pouze 2/7. Tato metoda se tedy soustředí na porovnání počtu a pořadí n-gramů strojového překladu a n-gramů referenčních překladů. Strojové překladače mají tendenci používat nadbytečné množství slov v porovnání s lidskými překladateli. Modifikovaná n-gramová přesnost penalizuje nadbytečná slova kandidátského překladu (ŠTANCEL, 2007) (KISHORE, 2002).

Příklad 1

Kandidátský překlad: the the the the the the the.

Referenční překlad 1: The cat is on the mat.

Referenční překlad 2: There is a cat on the mat.

BLEU užívá také pokutu za stručnost, anglicky Brevity Penalty, pro zhodnocení, zdali je délka kandidátského překladu přiměřená k délce překladu referenčního. Metoda porovnává vždy délku strojového překladu vzhledem k optimální délce referenčního textu, tedy k referenčnímu překladu, který je nejbližší počtem slov ke kandidátskému překladu. V případě, že se obě délky rovnají, kandidátský překlad získá penalizací 1. Pokud je text kratší, obdrží hodnocení nižší než jedna v závislosti na výši rozdílu obou textů (ŠTANCEL, 2007). Tato penalizace BLEU snižuje hodnocení překladačům za nepřeloženou informaci (MACHÁČEK, 2012).

Nedostatek metriky BLEU je práce s n-gramy, například pokud budeme hodnotit strojový překlad na bázi 4-gramů, přeložené věty obsahující méně než 4 slova budou výrazně podhodnocené (ŠTANCEL, 2007). Dále Bojar (2012, str. 111,112) uvádí jako slabinu systému BLEU nezapočtení slov podobných a synonym. Systém vyžaduje přesnou shodu slovních tvarů při porovnání referenčního a kandidátského textu, a tudíž

jazyky s bohatou morfologií, jako je například čeština, budou podhodnoceny, protože systém neudělí body správně přeloženému slovu, pokud je v jiném pádě a neshoduje se stoprocentně s referencí. BLEU neposkytuje překladači body ani za užitá synonyma k referenčnímu textu.

- WER

Word error Rate (WER) udává míru chybných slov kandidátského překladu. WER se se nejprve využíval jako metoda hodnocení rozpoznávání hlasu, ale současně se uplatňuje i pro evaluaci strojových překladačských systémů. Počítá se na základě Levenshteinovy vzdálenosti, která definuje počet nutných úprav výstupu překladače v porovnání s referenčním překladem. Úpravy se dělí do tří kategorií – výměna slova (substitution), vložení chybějícího slova (insertion) a smazání slova nadbytečného (deletion). Co se týká samotného vyhodnocování výsledků, systém každému slovu kandidátského překladu přiřadí určitý počet bodů na základě změn nebo shody (za shodu slov obdrží nulovou penalizaci) s referenčním textem. Nakonec sečte všechny obdržené body za každé slovo a vydělí je počtem slov referenčního překladu. Čím méně bodů překlad získá, tím lépe (KOEHN, 2010).

- METEOR

Tento systém vznikl roku 2005 jako reakce na metodu BLEU a funguje na podobné bázi. Autoři chtěli vyvinout takovou metriku na hodnocení strojového překladu, která by byla vylepšená o nedostatky předlohového systému, tedy nezohlednění sémanticky podobných slov a synonym kandidátského překladu při porovnání s referenčním textem. Tím se snažili přiblížit k lidskému posudku překladů. METEOR posuzuje výstupy překladačů na základě jednoho nebo více referenčních textů. Nejprve systém zarovná oba překlady, následně zanalyzuje všechna možná propojení mezi párem překladu, a nakonec posuzuje slovo po slově tak, že porovnává jednotlivé propojené dvojice unigramů kandidátského a referenčního textu. Výhodou METEORU oproti metodě BLEU je, že zahrnuje do správného řešení překladu i synonyma a slova podobná, tím vylepšuje skóre hodnocených textů a umožňuje tak kvalitnější posudek i na základě menšího počtu referenčních textů (ŠTANČEL, 2007).

2.5 Funkční styly

V praktické části budeme analyzovat úspěšnost strojového překladu v závislosti na funkčním stylu textu. Proto v této kapitole krátce vysvětlím původ a současné rozdělení funkčních stylů v českém jazyce. Vycházet budu z knihy *Stylistika současné češtiny* od M. Čechové.

Funkční stylová diferenciacie spisovného jazyka odráží veškeré funkce jazyka ve společnosti. V průběhu času se jazyk vyvíjel v závislosti na vnějších okolnostech. Čechová M. (1997, str. 28) uvádí například vliv společenské situace, ekonomických a politických podmínek nebo závislost jazyka na vyspělosti kultury. S vývojem řeči se mění i vymezení funkčních stylů textu. Dále vysvětluje, že styl každého projevu je při vzniku ovlivňován mnoha faktory, především výběrem a uspořádáním výrazových prostředků, které jsou spojeny s konkrétním funkčním stylem. V české lingvistice se objevily první teorie stylové diferenciacie ve 20. a 30. letech 20. století v pracích představitelů Pražského lingvistického kroužku. Nicméně první rozdělení funkčních stylů definoval ve svých pracích B. Havránek. Ve svých dílech vymezil funkční styly pomocí jazykových prostředků a definoval čtyři základní funkce textu – komunikativní, prakticky odbornou, teoreticky odbornou a estetickou (ČECHOVÁ, 2008).

Dle M. Čechové (1997, str. 31, 33) v současné lingvistice vymezení stylové diferenciacie probíhá především na základě funkce projevu. Rozdělují se na styly sdělovací (sdělené) a na styly esteticky sdělné. Hlavní funkcí textů sdělovacích je předání informací, tedy věcná stránka sdělení. Do této skupiny řadí prostěsdělovací styl, odborný styl, administrativní styl, publicistický styl a styl řečnický, které se liší na základě funkce sdělení. Styly esteticky sdělné neboli styly umělecké literatury představují texty lyriky (poezie), epiky (próza) a dramatiky (drama). Každý funkční styl používá příslušný funkční jazyk (hovorový, úřední, odborný...), náležitý slohový postup a jazykové prostředky.

3 Metodologický základ

3.1 Výchozí hypotézy

- 1) Vzhledem k počtu uživatelů, jazykové rozšířenosti a hodnocení online překladačů lze předpokládat, že Google překladač bude vykazovat kvalitnější výstupy, než překladač Bing, a tudíž metodou BLEU získá vyšší hodnocení.
- 2) S ohledem na rozsáhlost tréninkových dat pro evropské právo, z nichž systémy MT vycházejí, lze očekávat, že právě tyto texty budou vykazovat kvalitnější překlad, a získají tak nejvyšší hodnocení.
- 3) Co se týká kvality překladu vůči funkčnímu stylu, lze předpokládat vzhledem k charakteristickým rysům funkčních stylů, že překlad uměleckého textu bude nejméně vydařený.

3.2 Vybrané texty

V následující části představím deset vybraných textů, které budu v praktické části používat, a představím stručnou charakteristiku funkčních stylů těchto textů. Pro každý funkční styl jsem vybrala dvě ukázky textů ve francouzštině. Všechny zvolené texty jsou zaměřené na právní tematiku. Charakteristika bude vycházet z učebnice Český jazyk v kostce od M. Sochrové.

- **Prostěsdělovací styl**

Hlavní funkcí prostěsdělovacího stylu je výměna a předávání běžných informací v neoficiální a každodenní komunikaci. Pro tento funkční styl je charakteristické užívání hovorové mluvy, nářečí, slangu, profesní mluvy, nespisovných výrazů nebo jednoduché větné stavby. Používá převážně informační, popisný a vyprávěcí slohový postup.

Vybrané texty:

Pro prostěsdělovací styl jsem vybrala dva texty, oba představují přepis videa, jejichž hlavním cílem je vysvětlit jednoduše a humorným způsobem právní tematiku. Text 1 je přepisem videa „Le droit c’est quoi ?“ určeného pro děti od 6 do 11 let. Video

se snaží dětem jednoduchou formou vysvětlit, co je to právo. V textu jsou používány krátké věty, spisovný a hovorový jazyk, autor nepoužívá složité odborné označení, aby byl text co nejvíce pochopitelný pro cílovou skupinu posluchačů. Text 2 je přepisem videa „Les lois belges les plus absurdes“, kde autor předává odlehčenou formou informace o belgických zákonech. Je užíván především informační a popisný slohový postup. Autor používá hovorový jazyk. Vzhledem k zaměření všech textů na právní tematiku, tedy odborné téma, je obtížné najít texty, které by odpovídaly čistě prostěsdělovacímu stylu. Proto výše popsané neobsahují plně prvky charakteristické pro tento funkční styl, jako například slang nebo nespisovné výrazy, ale jsou na pomezí odborného a prostěsdělovacího stylu.

Prostěsdělovací styl
Text 1 – Přepis videa „Le droit c’est quoi ?“ určeného pro děti od 6 do 11 let
<p>Imagine une ville où rien ne serait interdit où l’on pourrait faire tout ce qu’on veut comme on veut et il n’y aurait pas de règles, il n’y aurait pas de code de la route, ce serait dangereux non ? Dans la vie, les consignes que l’on doit respecter sont des contraintes, mais elles nous donnent aussi des libertés et des droits, celui de traverser en sécurité par exemple. Chaque pays a des règles qui s’appliquent à tous, ce sont les lois. L’ensemble des lois constituent le droit d’un pays. Le droit donne des obligations et des responsabilités et il organise les rapports entre les personnes en disant ce qui est obligatoire, autorisé ou interdit, le droit évite les conflits. Il oblige ceux qui n’obéissent pas aux lois à réparer les dégâts causés par leurs actes, il protège les plus faibles parce qu’il encadre les accords que les personnes passent d’entre elles. Ces accords s’appellent des contrats. Les contrats sont partout, entre le garagiste et son client, entre le patron et son employé. Toutes ces relations ont en commun de devoir respecter le droit [5].</p>
Text 2 – Přepis videa „Les lois belges les plus absurdes“
<p>Il n’y a pas mal de lois absurdes l’on a déjà vu, mais il y a une loi exceptionnelle qui concerne les moules. Une loi vieille de 1901 qui a abouti à cette conséquence : si vous prenez les moules sur une brise-lame sans autorisation, vous êtes passible d’un emprisonnement qui peut aller d’un jour à 7 jours. Il est bien évident que cette loi est complètement désuète et n’est plus appliquée aujourd’hui Ces badauds en pleine pêche aux moules illégale pourraient donc bien finir sous les verrous. En Belgique la loi protège les moules, c’est la moindre des choses [14].</p>

- **Odborný styl**

Hlavní cílem odborného stylu je poučit a předat přesné informace. Texty tedy plní dvě funkce: odborně sdělnou a vzdělávací. Mezi charakteristické rysy tohoto stylu patří věcná správnost, obsahová úplnost, přehlednost, jednoznačné a přesné vyjádření, spisovný jazyk, užívání odborných termínů a názvů, výrazy bez citového zabarvení. Používá převážně popisný, výkladový a úvahový slohový postup. M. Sochrová (2009, str. 178) dále rozlišuje odborný styl na styl populárně naučný, styl prakticky odborný, styl vědecký a styl esejistický.

Vybrané texty:

K odbornému funkčnímu stylu jsem vybrala dva texty. První je definice Assemblée nationale a druhý popisuje fáze řízení o nesplnění povinnosti členské země Evropské unie. Text 3 je přesná definice Národního shromáždění, francouzsky Assemblée nationale“. Definice užívá spisovný jazyk, přesné a jednoznačné výrazy i odborné termíny (le pouvoir législatif). Text 4 plní funkci odborně sdělnou. Evropská komise může podniknout právní kroky proti zemi, která nesplňuje evropské právo. Tento text je oficiálně vydán Evropskou komisí na jejich internetových stránkách. Plní charakteristické rysy stylu odborného jako věcná správnost, obsahová úplnost, přehlednost, jednoznačné a přesné vyjádření. Vyznačuje se dále spisovným jazykem, užíváním odborných právních termínů (une procédure formelle d’infraction, la violation présumée, être déferés à la Cour...).

Odborný styl
Text 3 – Definice Assemblée nationale
L’Assemblée nationale fait partie du Parlement et partage le pouvoir législatif, c’est-à-dire le pouvoir de proposer et de voter des lois, avec le Sénat. Elle est composée 577 député·e·s élu·e·s pour 5 ans [2].
Text 4 – Fáze řízení o nesplnění povinnosti
La Commission peut engager une procédure formelle d’infraction lorsque le pays de l’UE concerné ne communique pas les mesures destinées à transposer

intégralement les dispositions d'une directive ou ne remédie pas à la violation présumée du droit de l'UE. La procédure suit un certain nombre d'étapes définies dans les traités de l'UE, chacune se soldant par une décision formelle:

- 1) La Commission envoie une lettre de mise en demeure exigeant de plus amples informations du pays concerné, qui doit envoyer une réponse détaillée dans un délai déterminé, qui est généralement de 2 mois.
- 2) Si la Commission conclut que le pays ne s'acquitte pas de ses obligations en vertu du droit européen, elle peut lui adresser un avis motivé: une demande formelle de se conformer au droit de l'Union. Elle explique pourquoi elle considère que le pays ne respecte pas le droit de l'Union. Elle demande également que le pays de l'UE l'informe des mesures prises, dans un délai déterminé, qui est généralement de 2 mois.
- 3) Si le pays de l'UE ne respecte toujours pas ses obligations, la Commission peut décider de saisir la Cour de justice. La plupart des cas sont réglés avant d'être déférés à la Cour.
- 4) Si un pays de l'UE ne communique pas les mesures destinées à mettre en œuvre les dispositions d'une directive en temps voulu, la Commission peut demander à la Cour de lui infliger des sanctions.
- 5) Si la Cour estime qu'un pays a enfreint la législation de l'Union, les autorités nationales doivent prendre des mesures pour se conformer à l'arrêt de la Cour. [17]

- **Administrativní styl**

Jedná se o styl úředního jednání a komunikace s úřady. Hlavní funkce administrativního textu jsou tyto: řídicí, odborně sdělná a správní. Pro tento funkční styl je charakteristická věcná výstižnost a správnost, grafická přehlednost, užívá výhradně spisovného jazyka, odborných termínů, zkratek, ustálených formulací (např. vejde v platnost) a používá převážně informační, popisný, charakterizační a výkladový slohový postup.

Vybrané texty:

Oba vybrané texty pro administrativní styl pochází z evropského práva. Jedná se o oficiálně vydané dokumenty od Evropské unie, z nichž jeden je výňatek z usnesení Evropského parlamentu a druhý představuje nařízení Evropského parlamentu a Rady.

Oba texty plní funkci řídicí a používají spisovný odborný jazyk. V textu 5 najdeme ustálené formulace jako například „considérant que“, příznačné pro přijaté právní předpisy nebo rozsudky. V textu 6 můžeme vidět užití zkratk (FEDER, FCE, FS, FEAMP).

Administrativní styl
Text 5 – Výňatek z Usnesení Evropského parlamentu ze dne 15. ledna 2020 o Zelené dohodě pro Evropu
considérant que la présente résolution constitue la première réaction générale du Parlement à la communication de la Commission sur le pacte vert pour l’Europe; que le Parlement reviendra, au fur et à mesure de l’avancement des travaux sur le pacte vert, avec des positions plus détaillées sur des mesures et des actions concrètes et fera usage de ses pleins pouvoirs législatifs pour réexaminer et modifier les propositions de la Commission afin de garantir qu’elles soutiennent l’ensemble des objectifs du pacte vert [18]
Text 6 – Nařízení Evropského parlamentu a Rady (EU) 2020, 460
(1) Les conséquences de la propagation du COVID-19 qui ont touché les États membres sont sans équivalent. La crise de santé publique actuelle entrave la croissance dans les États membres, ce qui accentue les graves pénuries de liquidités consécutives à l’augmentation significative et soudaine des investissements publics nécessaires dans leurs systèmes de santé et dans d’autres secteurs de leurs économies. Il en résulte une situation exceptionnelle à laquelle il convient de remédier par des mesures spécifiques.
(2) Il est essentiel que le manque de liquidités et de fonds publics dans les États membres ne freine pas les investissements au titre des programmes soutenus par le Fonds européen de développement régional (FEDER), le Fonds social européen (FSE) et le Fonds de cohésion (FC) (ci-après dénommés collectivement «Fonds») ainsi que par le Fonds européen pour les affaires maritimes et la pêche (FEAMP), qui sont nécessaires à la lutte contre la propagation du COVID-19 [19].

- **Publicistický styl**

Publicistický styl si klade za cíl rychle a výstižně informovat o aktuálních událostech z domova i ze světa, ovlivňovat názory a přesvědčovat své publikum. Plní tím tedy funkci agitační, sdělnou a rekreační (poskytnutí zábavy). Jedná se především o styl

hromadných sdělovacích prostředků, jako jsou noviny, rozhlas nebo televize. Publicistický styl se vyznačuje pravdivostí informací, aktuálností témat a obsahovou objektivitou autora. Texty jsou výstižné, přesvědčivé, snaží se zapůsobit na čtenáře/diváka svou originalitou a používají spisovný jazyk, obrazné vyjadřování, citově zabarvené výrazy, ironii nebo také řečnické otázky. Tento funkční styl používá především informační, úvahový a výkladový slohový postup.

Vybrané texty:

Vybrané texty pro publicistický styl pochází ze dvou nejznámějších francouzských deníků Le Monde a Le Figaro. První text publicistického stylu představuje úryvek z článku o důchodové reformě a druhý text je úryvkem z rozhovoru s viceprezidentem Evropské komise ohledně finanční situace států ve spojitosti s Covid-19. Oba texty informují o aktuálních událostech, plní tak funkci sdělnou a agitační, a jsou psané ve spisovném jazyce. V prvním textu (text 7) autor popisuje vyjádření premiéra ohledně důchodové reformy. Autor zde udává objektivní fakta jako datum, místo, citace premiéra. Text ale také obsahuje citové zabarvení, kdy autor používá například ironii (Tout va très bien entre le gouvernement et la majorité).

Publicistický styl

Text 7 – Úryvek z článku „Réforme des retraites : Edouard Philippe se « réjouit » des questions des députés LRM sur le « coût » de la réforme“ z francouzského deníku Le Monde

Tout va très bien entre le gouvernement et la majorité – c'est en tout cas ce que tient à afficher Edouard Philippe. Le premier ministre l'a assuré, jeudi 13 février, en disant « se réjouir » que « les députés de la majorité (...) partagent son attention permanente aux questions d'équilibre financier » du système de retraites, après que deux d'entre eux lui ont écrit pour avoir des précisions sur l'impact budgétaire de la réforme [24].

Text 8 – Úryvek z INTERVIEW - Le vice-président de la Commission européenne Valdis Dombrovskis salue les réformes de la France, mais la met en garde sur ses finances publiques z francouzského deníku Le Figaro

LE FIGARO. - Vous présentez ce mercredi les rapports analysant les performances économiques de chaque État membre. Quelle est votre évaluation ?

Valdis DOMBROVSKIS. - Dans l'ensemble, il est inquiétant de constater que la dette se maintient à un niveau élevé dans certains pays. De plus, les États membres ont encore du mal à revenir aux niveaux d'investissement public d'avant crise... bien qu'en parallèle, le niveau des dépenses est plus élevé qu'avant la crise. La croissance de la productivité reste également un défi. Face à cela, nous devons investir davantage dans les nouvelles technologies [20].

- **Umělecký styl (styl umělecké literatury)**

Styl umělecké literatury představuje texty lyriky (poezie), epiky (próza) a dramatiky (drama). Plní funkci především esteticky sdělnou, ale také formativní a informativní (vzdělávací, zábavná, výchovná). Hlavním cílem je kultivovat osobnost čtenáře prostřednictvím emocionálních a racionálních prožitků. Pro umělecký styl je charakteristické užívání individuálního stylu práce každého autora, mnohoznačností, využití všech jazykových prostředků, kompozičních postupů a rozmanitých obrazných pojmenování. Texty zahrnují všechny útvary národního jazyka, a tak v uměleckém stylu můžeme pozorovat jak spisovný jazyk, tak nespisovná vyjádření, profesní mluvu nebo slang. Umělecká díla používají hlavně vyprávěcí, popisný a úvahový slohový postup, a často se různé styly prolínají.

Vybrané texty:

Vybrané texty uměleckého textu tvoří dvě ukázky z policejního románu *Temná událost* od francouzského spisovatele Honorého de Balsaca. Román s kriminálním

zaměřením jsem vybrala proto, abych zachovala celkové zaměření textů na právní tematiku. Jedná se tedy o prozaické texty, které obsahují trestněprávní terminologii. Úryvky užívají spisovný jazyk, archaické vyjadřování, složitější a delší věty/souvětí.

Umělecký styl

Text 9 – Úryvek z knihy Temná událost - Honoré de Balsac

Le Code de Brumaire an IV réservait au directeur du Jury du Département la poursuite immédiate du délit commis à Gondreville. Remarquez, en passant, que la Convention avait rayé de la langue judiciaire le mot crime. Elle n'admettait que des délits contre la loi, délits emportant des amendes, l'emprisonnement, des peines infamantes ou afflictives. Néanmoins, la peine afflictive de la mort devait être supprimée à la paix, et remplacée par vingtquatre années de travaux forcés. Ainsi la Convention estimait que vingt-quatre années de travaux forcés égalaient la peine de mort. Que dire du Code pénal qui inflige les travaux forcés à perpétuité ? L'organisation alors préparée par le Conseil d'État de Napoléon supprimait la magistrature des directeurs du Jury qui réunissaient, en effet, des pouvoirs énormes. Relativement à la poursuite des délits et à la mise en accusation, le directeur du Jury était en quelque sorte à la fois agent de police judiciaire, procureur du roi, juge d'instruction et Cour royale. (Balsac)

Text 10 – Úryvek z knihy Temná událost - Honoré de Balsac

Avant de partir, il s'était entendu naturellement avec le préfet, qui, pris par la nuit, ne put se servir du télégraphe. On expédia sur Paris une estafette afin de prévenir le ministre de la Police Générale, le Grand-Juge et l'Empereur de ce crime inouï. Lechesneau trouva dans le salon de Gondreville mesdames Marion et Grévin, Violette, le valet de chambre du sénateur, et le juge de paix assisté de son greffier. Déjà des perquisitions avaient été pratiquées dans le château. Le juge de paix, aidé par Grévin, recueillait soigneusement les premiers éléments de l'instruction. Le magistrat fut tout d'abord frappé des combinaisons profondes que révélaient et le choix du jour et celui de l'heure. L'heure empêchait de chercher immédiatement des indices et des preuves. Dans cette saison, à cinq heures et demie, moment où

Violette avait pu poursuivre les délinquants, il faisait presque nuit ; et, pour les malfaiteurs, la nuit est souvent l'impunité. Choisir un jour de réjouissances où tout le monde irait voir la mascarade d'Arcis, et où le sénateur devait se trouver seul chez lui, n'était-ce pas éviter les témoins ? (Balsac)

3.3 Model pro hodnocení

Úspěšnost strojového překladu v závislosti na funkčním stylu textu budeme hodnotit na základě metody BLEU, která byla již představena v kapitole Metody hodnocení strojového překladu. Systém BLEU porovnává výstupy překladačů s referenčním textem a ohodnotí překladatelský systém na škále 0 až 1 s tím, že čím vyšší hodnocení je, tím kvalitnější kandidátský překlad je. Nejlepší hodnocení, které lze obdržet, je tedy 1. Užívá se také hodnocení 0 až 100. Pro evaluaci je nutné mít referenční překlad ke každému textu. Texty 4, 5 a 6 pocházejí z legislativy Evropské unie a existují ve všech 24 úředních jazycích, referenční překlady jsem získala z oficiálních stránek EU. Texty 9 a 10 představují ukázky z francouzského vydání knihy *Temná událost*, která byla přeložena do českého jazyka. Jako referenční překlad tedy použiji překlad knihy do českého jazyka, vydání z roku 1923, přeložené od A. Muřovského. Ke zbylým textům 1, 2, 3, 7 a 8 neexistuje oficiální překlad, a proto jsem pro účel této práce vytvořila referenční překlady těchto textů. S korekcí mého překladu mi pomohla Mgr. Kateřina Drsková z katedry romanistiky. Vybraných deset textů si nejprve přeložime v překladači Google a poté v překladači Bing. Výstupy z obou překladatelských systémů uložíme každý zvlášť v textovém dokumentu. V tomto typu dokumentu je důležité uložit i zdrojový text a referenční překlad, abychom následně mohli všechny čtyři dokumenty vložit do hodnotícího systému BLEU na internetové stránce <https://www.letsmt.eu/Bleu.aspx>. Všechny texty vkládané do systému BLEU, tedy zdrojový, referenční, a oba kandidátské překlady ke všech vybraným textům lze najít v přílohách této práce.

Dále budou texty vkládány do hodnotícího systému BLEU, nejprve v neupravené formě v textovém dokumentu (v původní podobě výstupů z obou překladačů). Poté formu textového dokumentu upravím a zarovnam analyzovaný text do stejného počtu řádků ve všech používaných dokumentech, tedy ve zdrojovém, v referenčním, a v obou

kandidátských překladech. Díky této metodě zjistím, jaká je nejvhodnější forma zarovnání textu pro obdržení co možná nejlepších výsledků systému BLEU.

4 Výsledky hodnocení

V následující kapitole se budu věnovat analýze výsledků kvality překladačů: Google Translate (dále Google, GT) a Microsoft Bing Translator (dále Bing, BT). Pomocí hodnocení systému BLEU budu porovnávat kvalitu výstupů jednotlivých textů obou překladačů.

Pro každý vybraný text jsem vytvořila tabulku, ve které můžeme pozorovat hodnocení BLEU pro jednotlivé překladače. Po tabulce bude následovat vždy můj komentář ohledně kvality výstupů překladačů a jejich porovnání v závislosti na typu analyzovaného textu. Překlady jsou číselně seřazené od 1 do 10. Nejprve budu v textu 1 a 2 rozebírat prostěsdělovací styl. Následně v textech 3 a 4 porovnáám výsledné hodnocení výstupů obou překladačů odborného stylu. Poté představím evaluaci BLEU administrativních textů 5 a 6. Nato uvedu výsledky a porovnání publicistického stylu, textů 7 a 8. Na závěr zanalyzuji poslední dva texty 9 a 10, zastupující styl umělecký.

Výsledné hodnocení všech překladů bude uváděno v unigramech a čtyřgramech, tedy hodnocení na základě porovnání jednoho slova kandidátských a referenčních překladů a hodnocení, kdy systém porovnává čtyři slova jako celek. Metoda BLEU má tendence podhodnocovat posuzované texty při využívání více gramových porovnání, proto budou představovány výsledky unigramového (1-gram) i čtyřgramového (4-gramy) hodnocení. Výsledná hodnota bude zapisována jako 0 až 100.

4.1 Hodnocení překladačů

- Prosdědělovací styl

Text 1		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	59,86	55,64
Hodnocení BLEU 4-gram	26,11	25,33

Text 1 byl vložen nejprve v neupravené formě, a získal hodnocení Google 36,32 a Bing 22,62. Po zarovnání do vět ve všech užitých dokumentech se evaluace BLEU zvýšila v průměru přibližně o 28 % na 59,86 pro Google, a 55,64 pro Bing. V kandidátských překladech obou systémů se vyskytovaly nejvíce gramatické chyby jako například chybně vyčasované sloveso „Imagine“, kdy systémy přeložily do 2. osoby množného čísla jednotného „Představte si“, namísto do 2. osoby jednotného čísla „Představ si“. Dále lze ve výstupech z překladačů vyzorovat chybný překlad slov, ustálených výrazů nebo právních termínů. Výraz „Code de la route“ nebyl správně přeložen ani jedním zkoumaným překladačem. GT tento výraz přeložil jako „dopravní zákony“ a Bing vygeneroval překlad „dopravní kód“. Dále pro „l'ensemble des lois“ oba překladače použily výraz „všechny zákony“ namísto „souhrn zákonů“ nebo u slova „le garagiste“ uvedly „majitel garáže“ místo „majitel autoservisu“, který je pro daný kontext vhodnější. Výstupy se vyznačovaly chybnou syntaxí, především překladač Bing měl v druhé větě potíže se slovosledem. Překladatelské systémy měly dále také potíže se změnou významu oproti zdrojovému textu. Například GT přeložil „le droit évite les conflits“ jako „zákon se vyhýbá konfliktům“ namísto „zákon předchází sporům“. Překlad GT dosáhl lepšího výsledku, protože byl obecněji lépe pochopitelný, než tomu bylo u výstupu z překladače Bing, který vykazoval větší počet gramatických i překladatelských chyb, kupříkladu chybnou shodu podmětu s přísudkem: ce sont les lois - je to zákony, chybné překlady: des obligations et des responsabilités - povinnosti a povinnosti.

Text 2		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	55,23	56.57
Hodnocení BLEU 4-gram	29,51	29.78

V původní neupravené podobě výstup Googlu obdržel 32,58, a překladač Bing získal 46,74. Po zarovnání jednotlivých vět do řádků vzniklo zhruba 16% vylepšení. V textu 2 měly překladače potíže nejprve s překladem francouzské negace v první větě „Il n’y a pas mal de lois absurdes“, jejíž význam v češtině je „Už jsme viděli spoustu nesmyslných zákonů“. Oba zkoumané systémy význam „Il n’y a pas mal de“ převrátily. Google jej přeložil jako „Existuje několik absurdních zákonů“, Bing vygeneroval překlad „Existuje poměrně málo absurdních zákonů“. Dále systémy až příliš věrně kopírovaly skladbu věty zdrojového jazyka, který je zcela odlišný od jazyka českého. Například v prvním souvětí oba překladače nechaly stejné pořadí vět jako ve francouzském originálu, a tak je souvětí pochopitelné, ale v češtině nezní úplně přirozeně. Stejný problém můžeme pozorovat i v druhé větě. Poté si můžeme ve výstupech strojového překladače všimnout vynechání překladu francouzských determinantů, konkrétně členů neurčitých. Tuto chybu vidíme například ve větě „mais il y a une loi exceptionnelle“, kde ani GT ani překladač Bing nepřeložily neurčitý člen „une“ a nabídly překlad „ale existuje výjimečný zákon“. Dále oba systémy udělaly několik gramatických chyb. GT uvedl slovo „vlnolam“ ve špatném pádu, kdy místo 4. pádu „na vlnolamu“ použil 1. pád „vlnolam“. Ani překladač Bing se nevyhnul gramatickým chybám. Například namísto „odnětí svobody, které“ vygeneroval špatnou shodu podstatného jména „odnětí“ se zájmenem vztažným „který“ „odnětí svobody, který“.

- Odborný styl

Text 3		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	61,14	63,80
Hodnocení BLEU 4-gram	33,90	36,37

U překladu definice Assemblée nationale vyhodnotil BLEU jako zdařilejší překladač Bing. Google získal evaluaci o 2,66 nižší než Bing. Nejprve byly analyzovány texty zarovnané z textového dokumentu do jednoho řádku, a získaly hodnocení Google 49,97 a Bing 56,22. Po úpravě formy a rozdělení textu do dvou řádků (rozdělení po větách) se úspěšnost překladů zvýšila v průměru zhruba o 9 %. Většina chyb v textu 3 spočívala v mylném překladu odborných termínů, například „le pouvoir législatif“ přeložily oba překladatelské systémy jako „zákonodárnou pravomoc“ místo „zákonodárné moci“. Poté se objevovaly chyby v psaní velkých počátečních písmen, především u Google překladače. GT uvedl chybně výraz „parlament“ s malým počátečním písmenem. Naopak slovo „Pravomoc“ uvedl s počátečním velkým písmenem. Tuto chybu lze vysvětlit tím, že slovu předchází zkratka, po které následuje tečka „tj.“, a tak systém pravděpodobně špatně vyhodnotil konec věty, a tedy použil velké písmeno jako by začínala další věta. Překladače měly také potíže se slovosledem v první větě, který až moc připomíná francouzskou syntax, například umístění slovního spojení „se Sénátem“, které bylo u obou systémů až na konci. Dále ani jeden z překladačů nezachoval ženský rod pro překlad „député·e·s élu·e·s“, který je ve francouzštině vyjádřen pomocí koncovky „e“. Oba systémy vygenerovaly překlad pouze v mužském rodu „poslanců“. Přidané koncovky tedy překladače zcela vynechaly.

Text 4 – Řízení o nesplnění povinnosti		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	58,48	61,89
Hodnocení BLEU 4-gram	25,41	28,07

Řízení o nesplnění povinnosti je text, který je oficiálně přeložen na stránkách Evropské komise, tudíž dle hypotézy o evropském právu a administrativě by právě tento text měl vykazovat vysoké hodnocení BLEU. Google dosáhl unigramového hodnocení 58,48. Bing získal o 3,41 evaluaci vyšší. BT celkově lépe odpovídal počtem slov v porovnání s referenčním textem, a tak získal lepší hodnocení v pokutě za stručnost. V té získal zdařilejší evaluaci ve čtyřech větách ze šesti zkoumaných. Text byl vložen do systému BLEU dvakrát. Poprvé nebyla forma textu upravována a podruhé byl text zarovnán po úsecích do šesti řádků testového dokumentu. Po zarovnání do řádků se hodnocení výrazně zlepšilo, zhruba o 30 %. Původně Google obdržel 31,70 a Bing 28,16 dle unigramového hodnocení. Mezi zásadní chyby výstupů strojového překladu patří špatná syntax, chybné nebo nepřesné překlady odborných termínů. Například „pays concerné“ přeložily oba analyzované systémy jako „dotčená země“ namísto „příslušná země“, „le pays de l'UE“ jako „země EU“ místo oficiálně užívaného spojení „členský stát“. Překladače se až příliš drží doslovného překladu a skladby vět, a přejímají tak francouzskou skladbu věty. Můžeme to například vidět ve větě „la Commission peut décider de saisir la Cour de justice“, která byla GT vygenerována jako „může se Komise rozhodnout postoupit záležitost Soudnímu dvoru“. GT se drží příliš francouzské větné stavby u výrazu „rozhodnout postoupit záležitost“, v českém jazyce nelze toto spojení přeložit pouze za pomoci infinitivu, ale je vhodnější užití vedlejší věty, jak tomu je v referenčním překladu, tedy „může Komise rozhodnout, že věc předloží Soudnímu dvoru“. Bing přeložil tuto větu správně vedlejší větou. Gramatické chyby se objevují méně než v předchozím prostědělovacím stylu, ale několik jich tam najdeme, kupříkladu slovo „předtím“ oba překladatelské systémy uvedly v pravopisně nesprávné podobě před tím, BT dále chyboval ve shodě podmětu s přísudkem ve spojení „opatření vyhověly“, namísto „opatření vyhověla“.

- **Administrativní styl**

Text 5		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	57,72	50,59
Hodnocení BLEU 4-gram	39,80	24,49

Text 5 byl vložen nejprve v neupravené formě, kdy Google získal hodnocení 40,98 a Bing 33,87. Poté jsem výňatek z Usnesení Evropského parlamentu zarovнала podle vět do dvou řádků a evaluace systému BLEU se vylepšila o 16 %. S výňatkem z Usnesení Evropského parlamentu měly systémy potíže především kvůli druhému dlouhému souvětí. Celkově se s textem lépe vypořádal GT, ten přeložil první větu téměř bezchybně. Pouze výraz „le pacte vert“ přeložil namísto „Zelená dohoda“ „Zelený pakt“. Avšak oba systémy správně uvedly velké písmeno „Z“, i když ve zdrojovém textu je malé písmeno. Druhé souvětí bylo pro oba překladače více komplikovanější, a tak systémy měly potíže se slovosledem, který kopírovaly z francouzského originálu, překládaly doslova, a tudíž text nebyl českému jazyku příliš přizpůsobený. GT uvedl více slov přesněji, a tak získal vyšší hodnocení. Bing přeložil chybně například výrazy „des mesures et des actions concrètes“ jako „konkrétní opatřením a opatření“. Na druhou stranu lze říci, že hodnocení Bingu je podhodnocené, jelikož systém BLEU neuznává slova v jiném pádě, a tak systému neuznal například správně přeložené slovo „reakcí“, které v referenčním textu bylo uvedeno v jiném pádě „reakci“, nebo mnohé další.

Text 6		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	60,80	66,95
Hodnocení BLEU 4-gram	36,78	41,00

Text 6 dosáhl po zarovnání do řádku 17% zlepšení. Výňatek z Nařízení Evropského parlamentu a Rady získal nejvyšší hodnocení ze všech textů 66,95, které obdržel překladač Bing. Druhá hypotéza ohledně kvality překladu evropského práva a jeho hodnocení systémem BLEU se tedy vyplnila. Oba překladače obdržely nejlepší hodnocení v metodě pokuta za stručnost. Lze tedy říci, že počtem slov odpovídaly nejvíce referenčnímu textu. I přes poměrně slušný překlad měly překladače potíže se syntaxí, především s uspořádáním vět v delším souvětí. To lze pozorovat hned na první větě textu, která evidentně kopíruje zdrojový text. U GT se dále objevovaly gramatické chyby jako špatné skloňování podstatných jmen, například „dalších odvětví“ namísto „dalších odvětvích“. Bing volil přesnější slovní zásobu a překládal dobře i delší úseky, proto obdržel také nejvyšší čtyřgramové hodnocení (41) ze všech textů. Kupříkladu u výrazu „une situation exceptionnelle à laquelle il convient de remédier“ předložil zcela stejný překlad jako referenční text „výjimečná situace, kterou je třeba řešit“. Celkově oba systémy překládaly lépe delší úseky textu.

- **Publicistický styl**

Text 7		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	43,73	48,57
Hodnocení BLEU 4-gram	12,85	8,64

První text publicistického stylu obdržel po patřičné úpravě formy zlepšené hodnocení BLEU o 13 %. Úryvek z článku o důchodové reformě získal po textech uměleckého stylu nejnižší evaluaci ze zbylých textů. Systémy měly potíže s porozuměním ironie a sarkasmu v první větě. V druhém delším souvětí lze například u obou překladačů pozorovat špatný překlad francouzského zájmena „le“ ve výrazu „Le premier ministre l’a assuré“, který přeložily jako „ho“ namísto „o tom“. Dále překladač Bing chyboval v uvedení uvozovek, které zaprvé v textu špatně umístil a zadruhé uvedl anglické uvozovky namísto českého značení uvozovek dole na začátku výrazu, a uvozovek nahoře na jeho konci. Tuto chybu lze vidět v druhé větě, kde Bing uvedl nesprávně všechny uvedené uvozovky – Premiér ho ve čtvrtek (13.února) ujistil , že se " raduje " , že " členové většiny (...) sdílet jeho neustálou pozornost na" finanční rovnováhu". Nepatrný rozdíl mezi překladači lze pravděpodobně vysvětlit tím, že BT lépe odpovídal počtem slov referenčnímu překladu, a tak získal lepší hodnocení v pokutě za přesnost, které se odrazilo v celkovém hodnocení BLEU.

Text 8		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	56,25	51,22
Hodnocení BLEU 4-gram	30,49	23,99

Hodnocení systému BLEU se u textu 8 zlepšily přibližně o 7 % po úpravě formy v textovém dokumentu. Druhý publicistický text - úryvek z rozhovoru z deníku Figaro obdržel zdařilejší hodnocení, než první úryvek z článku. O něco lépe byl hodnocen Google Překladač, který vygeneroval přesnější překlad některých slov, například slovo „un défi“ uvedl správně jako „výzva“, na rozdíl od překladu BT, který uvedl „problém“, nebo výraz „évaluation“, který zdařile přeložil jako „hodnocení“. Avšak je vidět, že i GT měl potíže s překladem. Kupříkladu můžeme pozorovat nevhodný slovosled v první větě, kdy příslovečné určení času „ve středu“ umístil stejně jako francouzský originál namísto na začátek věty. U překladače Bing vidíme vliv anglického jazyka, který je prostředníkem mezi českým a francouzským překladem. Název deníku „Le Figaro“ přeložil jako „The Figaro“. Bing dále uvedl chybný nebo nepřesný překlad. Mimo již zmíněné slovo „défi“ a „évaluation“ dále špatně vygeneroval například výraz „les rapports analysant“, u kterého uvedl jako český ekvivalent „zprávy, které rozhodují“ namísto „zprávy, které zkoumají (popřípadě analyzují)“. Oba systémy, GT i BT, zachovaly grafickou úpravu textu zdrojového jazyka. Například namísto české dvojtečky po uvedení člověka, jenž pronesl dále psaný text „FIGARO:“ zanechaly tečku a odrážku „LE FIGARO. –“.

U publicistického stylu by především výrazně zlepšilo hodnocení, kdyby kandidátské překlady byly porovnány s více referenčními překlady. Některé věty jsou správně, jenom jinak řečené. Publicistický styl umožňuje větší kreativitu autora, oproti například stylu administrativnímu nebo odbornému. Například poslední věta „Face à cela, nous devons investir davantage dans les nouvelles technologies.“ byla přeložena od GT jako „V této souvislosti musíme více investovat do nových technologií.“. I když je věta správně přenesena do cílového jazyka, po porovnání s referenčním překladem získala hodnocení 23,77.

- Umělecký styl

Text 9		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	39,13	34,72
Hodnocení BLEU 4-gram	8,89	7,81

Hodnocení BLEU uměleckého textu se po zarovnání do řádků zlepšilo o 16 %. Překlady krásné literatury dopadly v hodnocení BLEU dle očekávání nejhůře. Překladačské systémy obecně nejsou přizpůsobeny k překladu krásné literatury, mají problém s dlouhými souvětími a užitým jazykem, a proto jsou výstupy místy až nepochopitelné. Překladače měly mnohonásobně větší potíže přizpůsobit překládaný text obou výňatků z knihy *Temná událost* od Honorého de Balsaca cílovému jazyku, a často zanechávaly stejnou syntax zdrojového jazyka. To společně se špatně přeloženými slovy utvořilo často nesmyslné výstupy obou zkoumaných překladačů. Příkladem může být hned první věta výstupu BT: „Brumaire iV kód vyhrazena pro ředitele poroty oddělení okamžité stíhání trestného činu spáchaného v Gondreville.“. Dále systémy chybovaly v odborné terminologii trestního práva, například slovo „délits“ v první větě přeložily oba systémy jako „trestný čin“ místo „přečin“. GT i BT dále generovaly chybné překlady slov. Kupříkladu slovo „Département“ uvedly jako „oddělení“ namísto slova „okres“. Na této ukázce můžeme vidět jednu z hlavních slabín strojového překladu – vícevýznamová slova. Další chybné překlady byly například u GT „la magistrature“ - „magistrace“, u překladače Bing poté spojení „Le Code de Brumaire“, které bylo přeloženo „Brumaire kód“ namísto „Zákonník z brummairu“. Dále můžeme pozorovat chybu jako např. chybějící překlad, kupříkladu ve spojení „le directeur du Jury“, které GT vygeneroval jako „ředitel jury“. Zanechal tedy francouzské slovo „jury“ a vůbec ho nepřeložil.

Text 10		
Překladač	Google	Bing
Hodnocení BLEU 1-gram	35,06	38,71
Hodnocení BLEU 4-gram	4,54	4,46

Druhý výňatek z knihy *Temná událost* dopadl výsledkově podobně jako první umělecký text, přičemž v textu 10 byl hodnocen jako zdařilejší překlad Bing Translator. Text vykazoval chyby podobného charakteru jako předchozí ukázka. Objevovaly se chyby syntaktické, chybějící překlad například slovo „une estafette“ neboli česky „posel“, které překladač Bing zanechal ve francouzském znění „Estafette“, jelikož ho pravděpodobně vyhodnotil jako ženské jméno. Dále v textu můžeme vidět gramatické chyby, kupříkladu špatné skloňování podstatných jmen i jmen vlastních - „senátorův komorník“ od BT, který měl být dle kontextu překladu ve čtvrtém pádě. Výstupy obsahovaly také špatné překlady jednotlivých spojení například „le juge de paix“ (smírčí soudce) GT uvedl jako „spravedlnost za míru“. V textu 10 jsem objevila jedno zajímavé zjištění a to, že překladač Bing má slabinu v překladu hodin. Pokud se do systému nevloží celé hodiny, ale například „à cinq heures et demie“, Bing vyhodnotí překlad „v pět a půl“ namísto v půl páté. Tuto chybu uvedl jak v textu 10, tak i v dalších pokusech při vložení pouze daného výrazu. Při překladu z češtiny do francouzštiny je překlad správný a při zadání „v půl páté“ překladač automaticky vygeneruje překlad „à cinq heures et demie“.

4.2 Porovnání výstupů z GT a BT v závislosti na typu textu

Systém BLEU celkově ohodnotil jako kvalitnější výstupy ty od Překladače Microsoft Bing, který získal vyšší hodnocení u šesti textů z deseti vybraných modelů. Překladač Google přeložil lépe čtyři texty. Výsledky obou překladačů byly velice vyvážené a žádné hodnocení nevykazovalo větší než sedmibodový rozdíl mezi systémy. V příloze 1 (porovnání GT a BT v závislosti na typu texty, str. 56) lze najít grafické zpracování

výsledků obou zkoumaných systémů v závislosti na funkčním stylu textu. Jedná se o porovnání průměru hodnocení dvou textů, zastupující konkrétní funkční styl. Výsledné hodnoty GT a BT jsou poté porovnány vedle sebe. První dva texty zastupovaly prostěsdělovací styl. Text 1 přeložil lépe překladač Bing. Naopak v textu 2 představil kvalitnější výsledky Google Translate. V celkovém průměru obdržel vyšší hodnocení pro prostěsdělovací styl Google Translate. Jeho překlad byl obecněji lépe pochopitelný (především text 1), než tomu bylo u výstupu z překladače Bing, který vykazoval větší počet gramatických i překladatelských chyb. Odborný text byl u BT nejlépe hodnocený funkční styl. U GT se umístil na druhém místě. To lze pravděpodobně vysvětlit tím, že jeden z textů byl výrazně kratší, než všechny ostatní, a druhý text pocházel z evropského práva. Všeobecně vygeneroval lepší překlady systémem Bing Translator, který získal od systému BLEU vyšší evaluaci u obou modelových textů. Překlad BT více odpovídal počtem slov vůči referenčnímu textu, a tak obdržel lepší hodnocení v pokutě za stručnost. Například v textu 4 byly touto metodou lépe hodnoceny čtyři věty ze šesti zkoumaných. Texty 5 a 6, zastupující styl administrativní, byly v průměru o 0,49 lépe hodnoceny Google Translate, než překladačem Bing, to je ovšem téměř zanedbatelný rozdíl. Lze tedy hodnotit, že evaluace byly na velice podobné úrovni. Text 6 získal poté nejlepší hodnocení ze všech textů, to lze přisoudit k faktu, že se jedná o text oficiálně vydaný Evropskou unií, k němuž byl vypracován překlad do všech úředních jazyků EU, tedy i do češtiny. Tyto překlady jsou volně dostupné na stránkách EU, a překladače mají tudíž dostupná data na trénink neuronových sítí. Publicistický styl, který zastupovaly úryvky článků z nejznámějších francouzských deníků Le Monde a Le Figaro, byl druhý nejhůře hodnocený funkční styl. V překladu textu 7 uspěl lépe BT, konkrétně získal o 4,84 lepší hodnocení. To lze vysvětlit lepším hodnocením pokutou za stručnost. Naopak GT obdržel vyšší skóre o 5,7 u textu 8, ve kterém přeložil správně více slov než Bing. Při porovnání průměru z obou textů, lepší výsledky v publicistickém stylu získal Google Translate. Poslední zkoumaný funkční styl, tedy styl umělecký, představoval dva výňatky z knihy Temná událost od francouzského spisovatele Honorého de Balsaca. Tento funkční styl dopadl dle předpokladů nejhůře. V prvním výňatku byl podle systému BLEU zdařilejší GT. Avšak v druhém výňatku uspěl lépe BT. V celkovém průměru obdržel zdařilejší výsledky Google Translate. Nicméně ani jeden ze systémů strojového překladu se nezvládl zdaleka přiblížit lidskému překladu krásné literatury. Největší potíže jim působila delší souvětí a užitý styl jazyka.

5 Závěr

Cílem této bakalářské práce je sledovat, jak se mění kvalita strojového překladu z francouzštiny do češtiny v závislosti na funkčním stylu textu. Pro svou analýzu jsem vybrala deset různých vzorových textů, zastupující prostěsdělovací, odborný, administrativní, publicistický a umělecký styl. Pro hodnocení úspěšnosti jsem použila automatickou metodu měření pro strojový překlad – hodnotící systém BLEU.

Práce je rozdělena do třech hlavních částí. V části Teoretická východiska je nejprve představen pojem strojový překlad, krátká historie a využití strojového překladu. Poté byly stručně charakterizovány odlišné přístupy strojového překladu, které byly vytvořeny v průběhu času. Následně je uvedeno téma online překladačů, a souběžně s ním je také představen Google překladač a překladač Bing. Dále jsou uvedeny ruční a automatické metody hodnocení úspěšnosti strojového překladu. Nakonec je krátce popsán původ a současné rozdělení funkčních stylů v českém jazyce.

V další části, nazvané Metodologický základ, jsou zformulovány hypotézy očekávané úspěšnosti. Poté je představeno deset vybraných modelových textů, které jsou použity v praktické části. Zároveň je popsána jejich stručná charakteristika. Na závěr této kapitoly je předveden model hodnocení, podle něhož se následně postupuje v poslední, tedy praktické části.

Poslední část zahrnuje praktickou analýzu textů. Výsledky unigramového a čtyřgramového hodnocení systému BLEU jsou uspořádané do tabulek, přičemž při celkovém porovnání obou zkoumaných překladačů byly užity unigramové výsledky, poněvadž čtyřgramová evaluace může překlady podhodnocovat. Dále lze pozorovat, jak důležitou roli představuje zarovnání textu v textovém dokumentu před vložením do systému BLEU. U všech modelových textů se prokázalo výrazné zlepšení po zarovnání textů do řádků podle vět nebo úseků textu. Největší vylepšení hodnocení BLEU po úpravě formy bylo zaznamenáno u druhého textu odborného stylu - Fáze řízení o nesplnění povinnosti, jehož bodování se zvedlo o 30 % po zarovnání. V průměru texty po zarovnání do nejvhodnější formy získaly zlepšení o 16,5 %.

V praktické části mé práce lze pozorovat, že funkční styl textu má velký vliv na úspěšnost systémů strojového překladu. Dále bylo zjištěno, že důležitou roli hraje i tematické zaměření textů. Úspěšnost se také odvíjí od množství volně dostupných tréninkových dat na internetu, na kterých překladače mohou trénovat neuronové sítě. Jako

nejvíce zdařilý překlad se prokázal text 6 od BT, který spadal pod administrativní styl a představoval text EU, tudíž druhá hypotéza ohledně úspěšnosti překladu evropského práva se ukázala být pravdivá. Nejméně zdařilý výstup vygeneroval také překladač Bing, kdy se jednalo o text 9 uměleckého stylu, který byl obecně nejhůře překládaný. První hypotéza v metodologické části přisuzovala větší úspěšnost překladači Google. Tato domněnka se ukázala jako mylná. V celkovém součtu bodů všech analyzovaných textů by získal nepatrně lepší hodnocení překladač Bing, konkrétně obdržel pouze o 0,6 vyšší hodnocení. Lze tedy konstatovat, že výsledky obou překladačů byly opravdu vyrovnané. Bing uvedl nejzdařilejší výsledky v odborném stylu, který následoval styl administrativní, po něm se umístil prostěsdělovací styl, na čtvrtém místě skončil publicistický styl a nejhůře dopadl umělecký styl. Google vygeneroval nejlepší překlad v administrativním stylu. Těsně za ním skončil odborný styl, poté prostěsdělovací, publicistický a umělecký.

V hodnocení BLEU se odrazily i zásadní nedostatky tohoto systému. Ve většině analyzovaných textů se objevilo minimálně jednou, v některých i vícekrát, že texty byly podhodnocené. Systém BLEU neuznává slova v jiném tvaru, než je uvedeno v referenčním textu, a tak slova vhodně užitá v textu, avšak ve špatném pádě, nepřijal. Dále hodnotící systém penalizoval i správný překlad, který užil zcela odpovídající synonymum ke slovu referenčního překladu. Tato slabina by mohla být také lépe vyplněna při použití více preferenčních překladů, které by dopomohly k lepšímu hodnocení obou překladačů.

6 Résumé

Ce mémoire de licence a pour objectif de tester et d'évaluer le succès de la traduction automatique de la langue française en tchèque en fonction du style du texte à l'aide de la méthode d'évaluation automatique BLEU. Ils ont utilisé les deux traducteurs gratuits disponibles - Microsoft Bing Translator et Google Translate. Le travail est divisé en trois grandes parties – la partie théorique, la partie méthodologique et finalement la partie pratique. Chaque partie est par la suite divisée en plusieurs chapitres.

La première partie théorique est tout d'abord consacrée à l'histoire de la traduction automatique et des possibilités de son utilisation. Ensuite il y a une description des stratégies de base de la traduction automatique. Premièrement, il y a les règles de base de la traduction (approche directe, approche par transfert et approche interlangue). Deuxièmement la traduction automatique statistique qui utilise les méthodes statistiques pour générer la traduction. Troisièmement, le système hybride qui combine les deux systèmes précédents. Quatrièmement, les systèmes fonctionnant en vertu des réseaux de neurones artificiels. Puis il y a une introduction aux traducteurs automatiques où on explique brièvement leurs histoires. Par la suite, le chapitre propose aussi des informations fondamentales sur les deux traducteurs disponibles gratuitement analysés dans ce mémoire – traducteur Microsoft Bing et Google Translate. Ensuite il y a des méthodes d'évaluations de la traduction automatique. Elles se distinguent des méthodes d'évaluation faites par des personnes et en méthodes automatique (BLEU, WER, METEOR). Enfin, on montre des fonctions différentes du style de la langue tchèque où on distingue les cinq fonctions principales du texte en tchèque. En premier lieu des textes du langage quotidien, en deuxième lieu des textes spécialisés, en troisième lieu des textes administratifs, en quatrième lieu des textes journalistiques et en cinquième lieu des textes artistiques.

La deuxième partie est consacrée à la méthodologie de la partie pratique. Tout d'abord on indique les hypothèses du succès de la traduction. Ensuite on a présenté des textes utilisés dans la partie pratique, il s'agit de dix textes avec des fonctions différentes. Enfin on a expliqué la méthode d'évaluation dans la partie pratique.

La troisième partie représente les résultats de l'évaluation des traducteurs Bing et Google par le système BLEU. Les résultats sont présentés en nombre de 0 à 100. Plus la note de l'évaluation est élevée, plus le succès de la traduction automatique l'est

également. Les résultats de Google et de Bing sont représentés individuellement pour chaque texte en tableau avec un commentaire de l'évaluation qualitative des deux traducteurs. Enfin il y a une comparaison des traducteurs Google et Bing.

En conclusion, les résultats de l'analyse ont montré que Microsoft Bing a en général mieux réussi à traduire des textes. Concrètement, il a obtenu la meilleure évaluation pour six des textes sur dix documents analysés. Google Translate a eu plus de succès en traduction sur 4 des textes. Néanmoins les résultats sont pratiquement équilibrés. Le traducteur Microsoft Bing a mieux traduit les textes spécialisés, tandis que le traducteur Bing a généré une meilleure traduction des textes administratifs. Les textes artistiques, dont concrètement les deux extraits du roman policier « Une ténébreuse affaire » de Honoré de Balzac, ont obtenu tous les deux du système BLEU les moins bonnes évaluations.

7 Seznam obrázků

Obrázek 1 - Znázornění přímého přístupu (HUTCHINS W. J., 1992).....	10
Obrázek 2 - Interlingvní model se 2 jazykovými páry (HUTCHINS W. J., 1992)	11
Obrázek 3 - Transferový model se 2 jazykovými páry (HUTCHINS W. J., 1992)	11

8 Bibliografie

8.1 Elektronické dokumenty

[1] ALCINA, Amparo. Translation technologies Scope, tools and resources. In: ResearchGate [online]. 2008, [cit. 2020-01-25]. Dostupné z: https://www.researchgate.net/publication/233657002_Translation_technologies_Scope_tools_and_resources

[2] AUTOR NEZNÁMÝ. Définition Assemblée nationale. In: EDUCA DROITS: Éducation des enfants et des jeunes au(x) droit(s) [online]. [cit. 2020-04-24]. Dostupné z: <https://educadroit.fr/parcours/1/episode/1>

[3] AUTOR NEZNÁMÝ. Kolokace. In: Wikipedia: the free encyclopedia [online]. Poslední aktualizace 12. 10. 2019 [cit. 2020-02-28]. Dostupné z: <https://cs.wikipedia.org/wiki/Kolokace>

[4] AUTOR NEZNÁMÝ. Les lois belges les plus absurdes. In: Youtube [online]. 27.10.2016, [cit. 2020-04-24]. Dostupné z: <https://www.youtube.com/watch?v=NIW5BMn6LDw>. Kanál uživatele qceuppens.

[5] AUTOR NEZNÁMÝ. Module 1 : le droit, c'est quoi ? In: Youtube [online]. 14.2.2017, [cit. 2020-04-24]. Dostupné z: https://www.youtube.com/watch?v=bf6MFnr-OTY&feature=emb_logo. Kanál uživatele Défenseurs des droits.

- [6] AUTOR NEZNÁMÝ. Neuronové sítě. In: Wikipedia: the free encyclopedia [online]. Poslední aktualizace 11. 9. 2019 [cit. 2020-03-02]. Dostupné z: https://cs.wikipedia.org/wiki/Um%C4%9B1%C3%A1_neuronov%C3%A1_s%C3%AD%C5%A5
- [7] AUTOR NEZNÁMÝ. Překladač Google. In: Wikipedia: the free encyclopedia [online]. Poslední aktualizace 5. 4. 2020 [cit. 2020-04-09]. Dostupné z: https://cs.wikipedia.org/wiki/P%C5%99eklada%C4%8D_Google
- [8] AUTOR NEZNÁMÝ. Společnost IBM. In: Wikipedia: the free encyclopedia [online]. 2020, [cit. 2020-02-28]. Dostupné z: <https://cs.wikipedia.org/wiki/IBM>
- [9] AUTOR NEZNÁMÝ. Warren Weaver. In: Wikipedia: the free encyclopedia [online]. [cit. 2020-03-04]. Dostupné na: https://cs.wikipedia.org/wiki/Warren_Weaver
- [10] AUTOR NEZNÁMÝ. 20 nejlepších překladačů v roce 2019. 2019, [cit. 2020-04-09]. Dostupné z: <https://www.onlineprekladac.eu/blog/20-nejlepsich-online-prekladacu-v-roce-2019/>
- [11] AUTOR NEZNÁMÝ. 20 nejlepších překladačů v roce 2020. 9.1.2020, [cit. 2020-04-09]. Dostupné z: <https://www.onlineprekladac.eu/blog/nejlepsi-prekladace-v-roce-2020-online-prekladace-prekladace-pro-mobily-i-prohlizece/>
- [12] BOJAR, Ondřej. Hluboká revoluce ve strojovém překladu. In: Youtube [online]. 14.09.2017 [cit. 2020-03-02]. Dostupné z: <https://www.youtube.com/watch?v=pgU8gIFZduo>. Kanál uživatele Machine Learning Meetups.
- [13] BOJAR, Ondřej. Introduction - Sociable tanks | MT talks #1. In: Youtube [online]. 16.12.2014, [cit. 2020-02-25]. Dostupné z: https://www.youtube.com/watch?v=kOY_F1UTySs. Kanál uživatele Matematicko-fyzikální fakulta UK.

- [14] BOJAR, Ondřej. Strojový překlad. In: CzechEncy: Nový encyklopedický slovník češtiny [online]. 2017, [cit. 2020-0-01]. Dostupné z: <https://www.czechency.org/slovník/STROJOV%C3%9D%20P%C5%98EKLAD>
- [15] BROOKS, Richard. 11 Google Translate Facts You Should Know [Online]. 2.5.2016, [cit. 2020-01-22]. Dostupné z: <https://k-international.com/blog/google-translate-facts/>
- [16] CVRČEK, Václav. Morfologická analýza. In: WIKI Český národní korpus [online]. Poslední aktualizace 13. 9. 2013 [cit. 2020-02-25]. Dostupné z: https://wiki.korpus.cz/doku.php/pojmy:morfologicka_analyza
- [17] EVROPSKÁ KOMISE. Infringement procedure. In: Commission Européenne : Appliquer la législation européenne [online]. [cit. 2020-04-14]. Dostupné z: https://ec.europa.eu/info/law/law-making-process/applying-eu-law/infringement-procedure_fr
- [18] EVROPSKÝ PARLAMENT. Résolution du Parlement européen du 15 janvier 2020 sur le pacte vert pour l'Europe. In: Parlement européen [online]. 24.4.2020, [cit. 2020-04-24]. Dostupné z: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0005_FR.html
- [19] EVROPSKÝ PARLAMENT. Règlement (UE) 2020/460 du Parlement européen et du Conseil. In: EUR-Lex [online]. 30.3.2020, [cit. 2020-04-24]. Dostupné z: https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv:OJ.L_.2020.099.01.0005.01.FRA
- [20] HUBAUT, Leonor. Bruxelles prêt à de la souplesse sur les dépenses liées au coronavirus. In: Le Figaro [online]. 26.2.2020, [cit. 2020-04-24]. Dostupné z: <https://www.lefigaro.fr/conjoncture/la-dette-de-la-france-devrait-legerement-augmenter-selon-bruxelles-20200226>

[21] CHLUMSKÁ, Lucie. Paralelní korpusy. In: WIKI Český národní korpus [online]. Poslední aktualizace 21.8.2019 [cit. 2020-02-28]. Dostupné z: <https://wiki.korpus.cz/doku.php/pojmy:paralelni>

[22] JANKOVICS, László. Umělá inteligence: úspory nákladů překladu a tvorby technické dokumentace [Online]. 27.12.2019, [cit. 2020-01-22]. Dostupné z: https://www.technickytydenik.cz/rubriky/denni-zpravodajstvi/umela-inteligence-uspory-nakladu-prekladu-a-tvorby-technicke-dokumentace_48829.html

[23] KASÍK, Pavel. Google skokově vylepšil překlad. Umělá inteligence se zakousla do češtiny [Online]. 2017, Poslední aktualizace 20.4.2020 [cit. 2020-01-22]. Dostupné z: https://www.idnes.cz/technet/internet/google-translate-neural-networks.A170418_224649_sw_internet_pka

[24] LE MONDE AVEC AFPR. Réforme des retraites : Edouard Philippe se « réjouit » des questions des députés LRM sur le « coût » de la réforme. In: Le Monde [online]. 14.2.2020, [cit. 2020-04-24]. Dostupné z: https://www.lemonde.fr/politique/article/2020/02/13/retraites-edouard-philippe-se-rejouit-des-questions-des-deputes-lrm-sur-le-cout-de-la-reforme_6029463_823448.html

[25] YONGHUI, Wu a spol. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. In: Cornelle University [online]. Poslední aktualizace 8.10.2016, [cit. 2020-03-02]. Dostupné na: <https://arxiv.org/pdf/1609.08144.pdf>.

8.2 Tištěné dokumenty

BOJAR, Ondřej. Čeština a strojový překlad - Strojový překlad našincům a našinci strojovému překladu. Praha: Institute of Formal and Applied Linguistics, 2012. ISBN 978-80-904571-4-0.

ČECHOVÁ, M., M. KŘEMOVÁ a E. MINÁŘOVÁ. Stylistika současné češtiny. Jihlava: Lidové noviny, 2008. ISBN 978-80-7106-961-4

DE BALSAC, Honoré. Une ténébreuse affaire. 1841 Dostupné z: <https://onemorelibrary.com/index.php/fr/livres/litterature/book/litterature-francaise-180/une-tenebreuse-affaire-1778>

DE BALSAC, Honoré. Temná událost. Praha: Čsl. Podniky, 1923. Dostupné z: Digitální knihovny Kramerius od Národní knihovny České republiky

GASPARI, F., a J. HUTCHINS. 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). ISBN 9788790708160 8790708164. Dostupné z: www.hutchinsweb.me.uk/MTS-2007.pdf

HUTCHINS, W. J. a H. L. SOMERS. An Introduction to Machine Translation Somers. London: Academic Press, 1992. ISBN: 0-12-362830-X. Dostupné z: <http://www.hutchinsweb.me.uk/IntroMT-TOC.htm>

HUTCHINS, W. John. Machine translation: A concise history. Computer aided translation: Theory and practice. China: Chinese University of Hong Kong. Dostupné z: <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>

HUTCHINS, W. John. Machine translation: History of Research and Applications. Routledge Encyclopedia of Translation Technology, ed. Chan Sin-wai. 2015. Dostupné z: <http://www.hutchinsweb.me.uk/Routledge-2014.pdf>

KISHORE, P., S. ROUKOS, T. WARD, and W. ZHU. BLEU: a Method for Automatic Evaluation of Machine Translation. IBM T. J. Watson Research Center, Philadelphia, 2002.

KOEHN, Philipp. Statistical Machine Translation. Cambridge: Cambridge University Press, 2010. ISBN 978-521-87415-1

SOCHROVÁ, Marie. Český jazyk v kostce pro SŠ. Fragment, 2009. ISBN 978-80-253-0950-6.

ŠTANCEL, Radovan. Metódy hodnotenia kvality strojového prekladu. Brno, 2007. Diplomová práce. Masarykova univerzita v Brně, Fakulta informatiky.

ŠVARCOVÁ, Zora. Hodnocení automatického překladu SMT systémů (Google, Translate, Bing) z francouzštiny do češtiny: Kolokace z oblasti „Bezpečnost potravin“. České Budějovice, 2016. Bakalářská práce. Jihočeská univerzita v Českých Budějovicích, Filozofická fakulta, Ústav romanistiky.

TÝNOVSKÝ, Miroslav. Využití lingvistických informací při EBMT. Praha, 2007. Diplomová práce. Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky.

WARREN, McCulloch a Walter PITTS. Logical calculus of the ideas immanent in nervous activity. Springer, Berlin, Heidelberg, 1943. ISBN 978-3-642-70913-5

ZEHNALOVÁ, Jitka a spol. Kvalita a hodnocení překladu: Modely aplikace. Olomouc, Univerzita Palackého, 2015. ISBN 978-80-244-4792-6.

9 Přílohy

Grafy

Příloha č. 1 – Porovnání úspěšnosti výsledného hodnocení GT a BT v závislosti na funkčním stylu textu

Tabulky obsahující vždy:

- zdrojový text ve francouzském jazyce
- referenční překlad v českém jazyce
- český překlad pomocí Google Translate
- český překlad jednotlivých textů pomocí Microsoft Bing Translator

Příloha č. 2 - Přepis videa „Le droit c’est quoi ?“ určeného pro děti od 6 do 11 let

Příloha č. 3 - Přepis videa „Les lois belges les plus absurdes“

Příloha č. 4 - Definice Assemblée nationale

Příloha č. 5 - Fáze řízení o nesplnění povinnosti

Příloha č. 6 - Výňatek z Usnesení Evropského parlamentu ze dne 15. ledna 2020 o Zelené dohodě pro Evropu

Příloha č. 7 - Výňatek z Nařízení Evropského parlamentu a Rady (EU) 2020,460

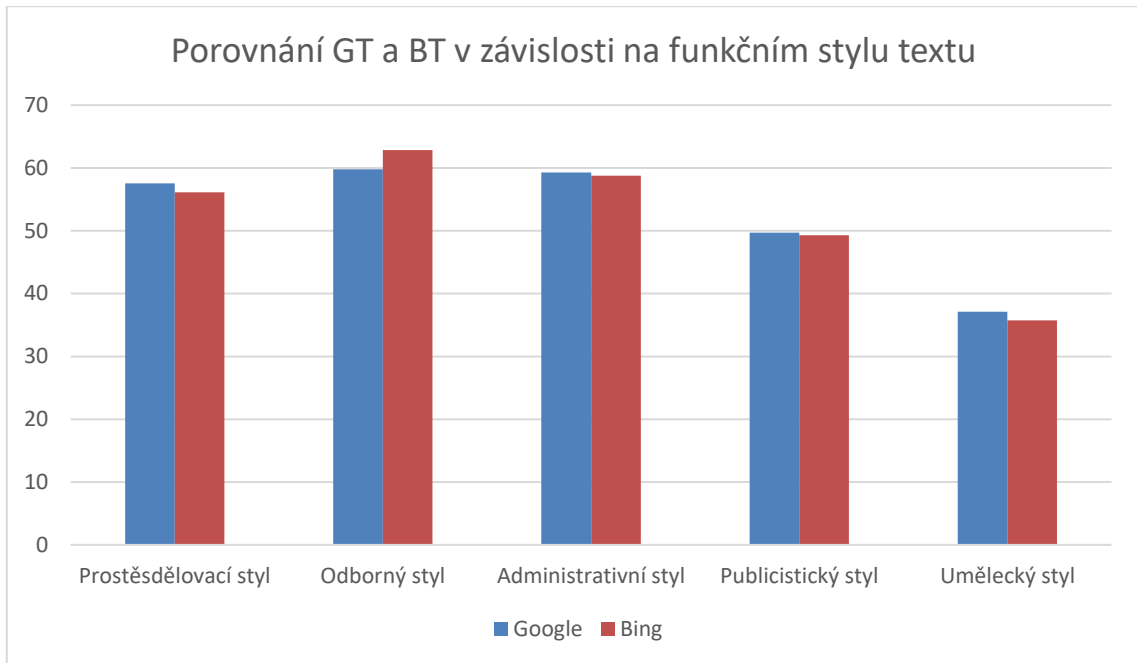
Příloha č. 8 - Úryvek z článku „Réforme des retraites : Edouard Philippe se « réjouit » des questions des députés LRM sur le « coût » de la réforme“ z francouzského deníku Le Monde

Příloha č. 9 - Úryvek z INTERVIEW - Le vice-président de la Commission européenne Valdis Dombrovskis salue les réformes de la France, mais la met en garde sur ses finances publiques z francouzského deníku Le Figaro

Příloha č. 10 – Úryvek 1 z knihy Temná událost- Honoré de Balsac

Příloha č. 11 Úryvek 2 z knihy Temná událost- Honoré de Balsac

Příloha č. 1



Příloha č. 2

Text 1	
Zdrojový text	<p>Imagine une ville où rien ne serait interdit où l'on pourrait faire tout ce qu'on veut comme on veut et il n'y aurait pas de règles, il n'y aurait pas de code de la route, ce serait dangereux non ? Dans la vie, les consignes que l'on doit respecter sont des contraintes, mais elles nous donnent aussi des libertés et des droits, celui de traverser en sécurité par exemple. Chaque pays a des règles qui s'appliquent à tous, ce sont les lois. L'ensemble des lois constituent le droit d'un pays. Le droit donne des obligations et des responsabilités et il organise les rapports entre les personnes en disant ce qui est obligatoire, autorisé ou interdit, le droit évite les conflits. Il oblige ceux qui n'obéissent pas aux lois à réparer les dégâts causés par leurs actes, il protège les plus faibles parce qu'il encadre les accords que les personnes passent d'entre elles. Ces accords s'appellent des contrats. Les contrats sont partout, entre le garagiste et son client, entre le patron et son employé.</p>

<p>Referenční text</p>	<p>Představ si město, kde by nebylo nic zakázané, kde bychom si mohli dělat všechno, co bychom chtěli, jak bychom chtěli, a neexistovala by žádná pravidla, žádná pravidla silničního provozu, nebylo by to nebezpečné? V každodenním životě musíme dodržovat pokyny, které nás sice omezují, ale také nám dávají svobodu a práva, jako například bezpečně přejít silnici. Každá země má svá pravidla, která jsou platná pro všechny občany, říkáme jim zákony. Souhrn zákonů tvoří právní řád té dané země. Právní řád určuje povinnosti, odpovědnosti a utváří vztah mezi lidmi tím, že říká, co je povinné, povolené, zakázané, a tak předchází sporům. Těm, co zákony nedodržují, přikazuje nahradit veškeré škody způsobené jejich jednáním. Ochraňuje nejslabší tím, že určuje pravidla dohod sjednaných mezi lidmi. Těmto dohodám se říká smlouvy. Smlouvy najdeme všude, třeba mezi majitelem autoservisu a jeho zákazníkem nebo mezi nadřízeným a jeho zaměstnancem.</p>
<p>Překlad Google</p>	<p>Představ si město, kde by nebylo nic zakázané, kde bychom si mohli dělat všechno, co bychom chtěli, jak bychom chtěli, a neexistovala by žádná pravidla, žádná pravidla silničního provozu, nebylo by to nebezpečné? V každodenním životě musíme dodržovat pokyny, které nás sice omezují, ale také nám dávají svobodu a práva, jako například bezpečně přejít silnici. Každá země má svá pravidla, která jsou platná pro všechny občany, říkáme jim zákony. Souhrn zákonů tvoří právní řád té dané země. Právní řád určuje povinnosti, odpovědnosti a utváří vztah mezi lidmi tím, že říká, co je povinné, povolené, zakázané, a tak předchází sporům. Těm, co zákony nedodržují, přikazuje nahradit veškeré škody způsobené jejich jednáním. Ochraňuje nejslabší tím, že určuje pravidla dohod sjednaných mezi lidmi. Těmto dohodám se říká smlouvy. Smlouvy najdeme všude, třeba mezi majitelem autoservisu a jeho zákazníkem nebo mezi nadřízeným a jeho zaměstnancem.</p>
<p>Překlad Bing</p>	<p>Představte si město, kde by nic nebylo zakázáno, kde byste mohli dělat, co chcete, a tam by žádná pravidla, tam by žádný dopravní kód, bylo by to nebezpečné v pořádku? V životě jsou pokyny, které je třeba dodržovat, omezení, ale také nám dávají svobodu a práva, například bezpečné překračování. Každá země má pravidla, která platí pro každého, je to zákony. Všechny zákony jsou právem země. Zákon dává povinnosti a povinnosti a organizuje vztahy mezi lidmi tím, že říká, co je povinné, povolené nebo zakázané,</p>

	<p>zákon zabraňuje konfliktu. Zavazuje ty, kteří nedodržují zákony, aby napravili škody způsobené jejich činy, chrání ty nejslabší, protože rámuje dohody, které o nich lidé dělají. Tyto dohody se nazývají smlouvy. Smlouvy jsou všude, mezi majitelem garáže a jeho klientem, mezi šéfem a jeho zaměstnancem.</p>
--	--

Příloha č.3

Text 2	
Zdrojový text	<p>Il n'y a pas mal de lois absurdes l'on a déjà vu, mais il y a une loi exceptionnelle qui concerne les moules. Une loi vieille de 1901 qui a abouti à cette conséquence : si vous prenez les moules sur une brise-lame sans autorisation, vous êtes passible d'un emprisonnement qui peut aller d'un jour à 7 jours. Il est bien évident que cette loi est complètement désuète et n'est plus appliquée aujourd'hui. Ces badauds en pleine pêche aux moules illégale pourraient donc bien finir sous les verrous. En Belgique la loi protège les moules, c'est la moindre des choses.</p>
Referenční text	<p>Už jsme viděli spoustu nesmyslných zákonů, ale existuje jeden vskutku výjimečný ohledně slávky jedlé (mušlí). Je to starý zákon z roku 1901, který vede v důsledku k tomu, že pokud si na pobřeží nasbíráte mušle bez příslušného povolení, hrozí vám odnětí svobody na dobu jednoho až sedmi dnů. Je zřejmé, že tento zákon je zcela zastaralý a dnes se už neuplatňuje. Takže lidé, kteří nezákonně využívají volného sběru mušlí, by mohli skončit za mřížemi. V Belgii zákon zkrátka chrání mušle, to je přece maličkost.</p>
Překlad Google	<p>Existuje několik absurdních zákonů, které jsme již viděli, ale existuje výjimečný zákon, který se týká mušlí. Starý zákon z roku 1901, který vedl k tomuto důsledku: pokud vezmete mušle na vlnolam bez povolení, můžete být uvězněni od jednoho dne do 7 dnů. Je zřejmé, že tento zákon je zcela zastaralý a dnes se již nepoužívá. Tito diváci uprostřed nezákonného lovu mušlí by proto</p>

	mohli skončit za mřížemi. V Belgii zákon chrání slávky, to je to nejméně.
Překlad Bing	Existuje poměrně málo absurdních zákonů, které jsme již viděli, ale existuje výjimečný zákon, který se zabývá slávkami. Starý zákon z roku 1901, který vyústil v tento důsledek: pokud budete mít mušle na vlnovodu bez povolení, jste náchylní k trestu odnětí svobody, který se může pohybovat od jednoho dne do 7 dnů. Je zcela jasné, že tento zákon je zcela zastaralý a dnes již není prosazován. Tito přihlížející uprostřed nezákonného lovu mušlí by mohli skončit za mřížemi. V Belgii zákon chrání mušle, to je to nejmenší.

Příloha č.4

Text 3	
Zdrojový text	L'Assemblée nationale fait partie du Parlement et partage le pouvoir législatif, c'est-à-dire le pouvoir de proposer et de voter des lois, avec le Sénat. Elle est composée 577 député·e·s élu·e·s pour 5 ans.
Referenční text	Národní shromáždění je součástí Parlamentu a podílí se na zákonodárné moci, tudíž má společně se Senátem pravomoc navrhnout zákony a hlasovat o nich. Skládá se z 577 poslanců a poslankyní volených na 5 let.
Překlad Google	Národní shromáždění je součástí parlamentu a sdílí zákonodárnou pravomoc, tj. Pravomoc navrhnout a přijímat zákony, se Senátem. Skládá se z 577 poslanců zvolených na 5 let.
Překlad Bing	Národní shromáždění je součástí Parlamentu a sdílí zákonodárnou pravomoc, tedy pravomoc navrhnout a hlasovat o právních

	předpisech se Senátem. Skládá se z 577 poslanců zvolených na 5 let.
--	---

Příloha č.5

Text 4	
Zdrojový text	<p>La Commission peut engager une procédure formelle d'infraction lorsque le pays de l'UE concerné ne communique pas les mesures destinées à transposer intégralement les dispositions d'une directive ou ne remédie pas à la violation présumée du droit de l'UE. La procédure suit un certain nombre d'étapes définies dans les traités de l'UE, chacune se soldant par une décision formelle:</p> <ol style="list-style-type: none"> 1) La Commission envoie une lettre de mise en demeure exigeant de plus amples informations du pays concerné, qui doit envoyer une réponse détaillée dans un délai déterminé, qui est généralement de 2 mois. 2) Si la Commission conclut que le pays ne s'acquitte pas de ses obligations en vertu du droit européen, elle peut lui adresser un avis motivé: une demande formelle de se conformer au droit de l'Union. Elle explique pourquoi elle considère que le pays ne respecte pas le droit de l'Union. Elle demande également que le pays de l'UE l'informe des mesures prises, dans un délai déterminé, qui est généralement de 2 mois. 3) Si le pays de l'UE ne respecte toujours pas ses obligations, la Commission peut décider de saisir la Cour de justice. La plupart des cas sont réglés avant d'être déférés à la Cour. 4) Si un pays de l'UE ne communique pas les mesures destinées à mettre en œuvre les dispositions d'une directive en temps voulu, la Commission peut demander à la Cour de lui infliger des sanctions. 5) Si la Cour estime qu'un pays a enfreint la législation de l'Union, les autorités nationales doivent prendre des mesures pour se conformer à l'arrêt de la Cour.

<p>Referenční text</p>	<p>Pokud příslušná evropská země nesdělí, kterými opatřeními plně dotčená ustanovení směrnic provádí, nebo nezjedná nápravu vytykaného porušení předpisů EU, může Komise zahájit formální řízení o nesplnění povinnosti. Toto řízení se skládá z několika kroků, které jsou stanoveny ve Smlouvách. Každý krok je zakončen formálním rozhodnutím:</p> <ol style="list-style-type: none"> 1) Komise zašle dané zemi výzvu k zaslání dalších informací a ta musí ve stanovené lhůtě (obvykle 2 měsíců) zaslat podrobnou odpověď. 2) Dospěje-li Komise k závěru, že země neplní své povinnosti vyplývající z právních předpisů EU, může zaslat odůvodněné stanovisko s formální žádostí o zajištění souladu s právem EU. V něm objasní, proč se domnívá, že tato země porušuje unijní právo. Rovněž požádá, aby daná země ve stanovené lhůtě (obvykle do 2 měsíců) informovala Komisi o přijatých opatřeních. 3) Jestliže ani poté daný členský stát nedosáhne souladu s právem EU, může Komise rozhodnout, že věc předloží Soudnímu dvoru. Většina případů je vyřešena předtím, než je věc Soudnímu dvoru postoupena. 4) Pokud daná země EU včas nesdělí, kterými opatřeními provádí ustanovení směrnice, může Komise požádat Soudní dvůr o uložení pokuty. 5) Jestliže Soudní dvůr shledá, že daná země právní předpisy EU porušila, musí její vnitrostátní orgány přijmout opatření, aby rozsudku Soudního dvora vyhověla.
<p>Překlad Google</p>	<p>Komise může zahájit formální řízení o nesplnění povinnosti, pokud dotčená země EU nesdělí opatření určená k úplnému provedení ustanovení směrnice nebo nenapraví údajné porušení právních předpisů EU. Postup probíhá v několika fázích definovaných ve smlouvách EU, z nichž každá vede k formálnímu rozhodnutí:</p> <ol style="list-style-type: none"> 1) Komise zasílá formální výzvu s žádostí o další informace od dotčené země, která musí zaslat podrobnou odpověď ve stanovené lhůtě, která je obvykle 2 měsíce. 2) Pokud Komise dojde k závěru, že země neplní své povinnosti podle evropského práva, může jí zaslat odůvodněné stanovisko: formální žádost o dodržení práva Unie. Vysvětluje, proč se domnívá, že země nedodržuje právo Unie. Rovněž požaduje, aby

	<p>země EU informovala o přijatých opatřeních ve stanovené lhůtě, která je obvykle 2 měsíce.</p> <p>3) Pokud země EU stále neplní své povinnosti, může se Komise rozhodnout postoupit záležitost Soudnímu dvoru. Většina případů je vyřešena před tím, než je předložena Soudnímu dvoru.</p> <p>4) Pokud země EU včas nesdělí opatření určená k provedení ustanovení směrnice, může Komise požádat Soudní dvůr o uložení sankcí.</p> <p>5) Pokud se Soudní dvůr domnívá, že země porušila právo Unie, musí vnitrostátní orgány přijmout opatření, aby vyhověly rozsudku Soudního dvora.</p>
<p>Překlad Bing</p>	<p>Komise může zahájit formální řízení o nesplnění povinnosti, pokud dotčená země EU nezveřejní opatření k úplnému provedení ustanovení směrnice nebo k nápravě údajného porušení práva EU. Tento postup se řídí řadou kroků vymezených ve smlouvách EU, z nichž každý vede k formálnímu rozhodnutí:</p> <p>1) Komise zašle výzvu požadující další informace od dotčené země, která musí zaslat podrobnou odpověď ve stanovené lhůtě, která je obvykle 2 měsíce.</p> <p>2) Pokud Komise dospěje k závěru, že země neplní své povinnosti podle evropského práva, může jí vydat odůvodněné stanovisko: formální žádost o dosažení souladu s právem EU. Vysvětluje, proč se domnívá, že země nerespektuje právo EU. Rovněž žádá, aby ji země EU informovala o přijatých opatřeních ve stanovené lhůtě, která činí obvykle 2 měsíce.</p> <p>3) Pokud země EU stále neplní své povinnosti, může se Komise rozhodnout, že věc postoupí Soudnímu dvoru. Většina případů je vyřešena před tím, než jsou postoupeny Soudnímu dvoru.</p> <p>4) Pokud země EU včas nesdělí opatření k provedení ustanovení směrnice, může Komise požádat Soudní dvůr, aby jí uložil sankce.</p> <p>5) Pokud Soudní dvůr shledá, že země porušila právo EU, musí vnitrostátní orgány přijmout opatření, aby rozhodnutí Soudního dvora splnily.</p>

Text 5	
Zdrojový text	<p>considérant que la présente résolution constitue la première réaction générale du Parlement à la communication de la Commission sur le pacte vert pour l'Europe;</p> <p>que le Parlement reviendra, au fur et à mesure de l'avancement des travaux sur le pacte vert, avec des positions plus détaillées sur des mesures et des actions concrètes et fera usage de ses pleins pouvoirs législatifs pour réexaminer et modifier les propositions de la Commission afin de garantir qu'elles soutiennent l'ensemble des objectifs du pacte vert</p>
Referenční text	<p>vzhledem k tomu, že toto usnesení představuje první obecnou reakci Parlamentu na sdělení Komise o Zelené dohodě pro Evropu; vzhledem k tomu, že Parlament se hodlá během uskutečňování Zelené dohody znovu vyjadřovat k této věci prostřednictvím podrobnějších postojů ke zvláštním opatřením a politickým krokům a hodlá plně využít své legislativní pravomoci k přezkumu a úpravám návrhů Komise, aby podporovaly všechny cíle Zelené dohody</p>
Překlad Google	<p>vzhledem k tomu, že toto usnesení představuje první obecnou reakci Parlamentu na sdělení Komise o Zeleném paktu pro Evropu; že Parlament se v průběhu práce na Zeleném paktu vrátí s podrobnějšími stanovisky ke konkrétním opatřením a činnostem a využije své plné legislativní pravomoci k přezkumu a úpravě návrhů Komise zajistit, aby podporovaly všechny cíle Zeleného paktu</p>
Překlad Bing	<p>vzhledem k tomu, že toto usnesení je první obecnou reakcí Parlamentu na sdělení Komise o Zeleném paktu pro Evropu; že se Parlament vrátí s postupem práce na Zeleném paktu s podrobnějšími postoji ke konkrétním opatřením a opatřením a využije svých plných legislativních pravomocí k přezkumu a změně návrhů Komise, aby zajistil, že budou podporovat všechny cíle Zeleného paktu</p>

Text 6	
Zdrojový text	<p>(1) Les conséquences de la propagation du COVID-19 qui ont touché les États membres sont sans équivalent. La crise de santé publique actuelle entrave la croissance dans les États membres, ce qui accentue les graves pénuries de liquidités consécutives à l'augmentation significative et soudaine des investissements publics nécessaires dans leurs systèmes de santé et dans d'autres secteurs de leurs économies. Il en résulte une situation exceptionnelle à laquelle il convient de remédier par des mesures spécifiques.</p> <p>(2) Il est essentiel que le manque de liquidités et de fonds publics dans les États membres ne freine pas les investissements au titre des programmes soutenus par le Fonds européen de développement régional (FEDER), le Fonds social européen (FSE) et le Fonds de cohésion (FC) (ciaprès dénommés collectivement «Fonds») ainsi que par le Fonds européen pour les affaires maritimes et la pêche (FEAMP), qui sont nécessaires à la lutte contre la propagation du COVID-19.</p>
Referenční text	<p>(1) Členské státy byly šířením onemocnění COVID-19 postiženy způsobem, jenž nemá obdoby. Tato současná krize v oblasti veřejného zdraví brání růstu v členských státech, což následně zhoršuje vážný nedostatek likvidity způsobený náhlým a významným zvýšením veřejných investic, jež jsou třeba v systémech zdravotní péče těchto států a v dalších odvětvích jejich ekonomik. Vznikla tak výjimečná situace, kterou je třeba řešit prostřednictvím zvláštních opatření.</p> <p>(2) Je mimořádně důležité, aby nedostatek likvidity a veřejných prostředků v členských státech nebránil investicím v rámci programů podporovaných Evropským fondem pro regionální rozvoj (EFRR), Evropským sociálním fondem (ESF) a Fondem soudržnosti (FS) (společně dále jen „fondy“) a Evropským námořním a rybářským fondem (ENRF), jež jsou třeba pro boj proti šíření onemocnění COVID-19.</p>

<p>Překlad Google</p>	<p>(1) Důsledky šíření COVID-19, které ovlivnily členské státy, nejsou srovnatelné. Současná krize v oblasti veřejného zdraví brzdí růst v členských státech a prohlubuje závažný nedostatek likvidity vyplývající z náhlého a významného nárůstu nezbytných veřejných investic do jejich zdravotnických systémů a dalších odvětví jejich hospodářství. Výsledkem je výjimečná situace, která musí být napravena zvláštními opatřeními.</p> <p>(2) Je nezbytné, aby nedostatek likvidity a veřejných prostředků v členských státech nebránil investicím v rámci programů podporovaných Evropským fondem pro regionální rozvoj (EFRR), Evropským sociálním fondem (ESF) a Fondem soudržnosti (dále jen „fond“), jakož i Evropským fondem pro námořní záležitosti a rybolov (EMFF), které jsou nezbytné k boji proti šíření COVID-19.</p>
<p>Překlad Bing</p>	<p>(1) Důsledky rozšíření COVID-19, které postihly členské státy, nemají obdoby. Současná krize v oblasti veřejného zdraví brzdí růst v členských státech a zhoršuje vážný nedostatek likvidity v důsledku významného a náhlého nárůstu veřejných investic potřebných v jejich zdravotnických systémech a dalších odvětvích jejich ekonomik. Výsledkem je výjimečná situace, kterou je třeba řešit konkrétními opatřeními.</p> <p>(2) Je nezbytné, aby nedostatek likvidity a veřejných prostředků v členských státech nebránil investicím do programů podporovaných Evropským fondem pro regionální rozvoj (EFRR), Evropským sociálním fondem (ESF) a Fondem soudržnosti (FC) (souhrnně označovaným jako "fondy"), jakož i Evropským fondem pro námořní záležitosti a rybolov (FEAMP) , které jsou nezbytné pro kontrolu šíření COVID-19.</p>

Příloha č.8

<p>Text 7</p>	
<p>Zdrojový text</p>	<p>Tout va très bien entre le gouvernement et la majorité – c'est en tout cas ce que tient à afficher Edouard Philippe. Le premier ministre l'a assuré, jeudi 13 février, en disant « se réjouir » que «</p>

	<p>les députés de la majorité (...) partagent son attention permanente aux questions d'équilibre financier » du système de retraites, après que deux d'entre eux lui ont écrit pour avoir des précisions sur l'impact budgétaire de la réforme.</p>
Referenční text	<p>Mezi vládou a většinou poslanců funguje všechno, jak má – to se alespoň za každou cenu snaží tvrdit Edouard Philippe. Premiér o tom ubezpečil 13. února, když řekl, že ho těší, že „většina poslanců sdílí jeho stálou pozornost vůči otázkám finanční rovnováhy důchodového systému“, poté, co se na něj písemně obrátili dva poslanci s žádostí o upřesnění dopadu reformy na rozpočet.</p>
Překlad Google	<p>Mezi vládou a většinou je vše v pořádku - to chce Edouard Philippe ukázat. Předseda vlády ho ve čtvrtek 13. února ujistil a řekl: „radovat se“, že „poslanci většiny (...) sdílejí jeho trvalou pozornost na otázky finanční rovnováhy“ důchodového systému, po dvou napsali mu podrobnosti o dopadu reformy na rozpočet.</p>
Překlad Bing	<p>Mezi vládou a většinou jde všechno velmi dobře - to je alespoň to, co chce Edouard Philippe ukázat. Premiér ho ve čtvrtek (13. února) ujistil, že se "raduje", že "členové většiny (...) sdílet jeho neustálou pozornost na "finanční rovnováhu" otázky důchodového systému, poté, co dva z nich napsal mu pro objasnění rozpočtového dopadu reformy.</p>

Příloha č. 9

Text 8	
Zdrojový text	<p>LE FIGARO. - Vous présentez ce mercredi les rapports analysant les performances économiques de chaque État membre. Quelle est votre évaluation ?</p> <p>Valdis DOMBROVSKIS. - Dans l'ensemble, il est inquiétant de constater que la dette se maintient à un niveau élevé dans certains pays. De plus, les États membres ont encore du mal à revenir aux</p>

	<p>niveaux d'investissement public d'avant crise... bien qu'en parallèle, le niveau des dépenses est plus élevé qu'avant la crise. La croissance de la productivité reste également un défi. Face à cela, nous devons investir davantage dans les nouvelles technologies.</p>
Referenční text	<p>FIGARO: Ve středu představíte zprávy, které zkoumají výkonnost jednotlivých členských států. Jaké je Vaše hodnocení ?</p> <p>Valdis DOMBROVSKIS: Celkově je znepokojivé, že se úroveň dluhu v některých zemích drží vysoko. Členské státy navíc mají problém vrátit výši veřejných investic na úroveň před krizí... i přesto, že výdaje jsou teď zároveň vyšší než před krizí. Výzvou zůstává i růst produktivity. Na tuto situaci bude třeba reagovat investicemi do nových technologií.</p>
Překlad Google	<p>FIGARO. - Předkládáte tuto středu zprávy, které analyzují hospodářskou výkonnost každého členského státu. Jaké je vaše hodnocení?</p> <p>Valdis DOMBROVSKIS. - Celkově je znepokojivé, že v některých zemích je dluh stále vysoký. Kromě toho se členské státy stále snaží navrátit k předkrizovým úrovním veřejných investic ... i když zároveň je výše výdajů vyšší než před krizí. Výzvou zůstává i růst produktivity. V této souvislosti musíme více investovat do nových technologií.</p>
Překlad Bing	<p>THE FIGARO. - Ve středu předkládáte zprávy, které rozhodují o hospodářské výkonnosti každého členského státu. Jaký je váš odhad?</p> <p>Valdis DOMBROVSKIS. - Celkově je znepokojivé, že dluh zůstává v některých zemích vysoký. Členské státy se navíc stále potýkají s problémy s návratem k předkrizovým úrovním veřejných investic... i když je úroveň výdajů vyšší než před krizí. Problémem zůstává i růst produktivity. Tváří v tvář tomu musíme více investovat do nových technologií.</p>

Text 9	
Zdrojový text	<p>Le Code de Brumaire an IV réservait au directeur du Jury du Département la poursuite immédiate du délit commis à Gondreville. Remarquez, en passant, que la Convention avait rayé de la langue judiciaire le mot crime. Elle n'admettait que des délits contre la loi, délits emportant des amendes, l'emprisonnement, des peines infamantes ou afflictives. Néanmoins, la peine afflictive de la mort devait être supprimée à la paix, et remplacée par vingtquatre années de travaux forcés. Ainsi la Convention estimait que vingt-quatre années de travaux forcés égalaient la peine de mort. Que dire du Code pénal qui inflige les travaux forcés à perpétuité? L'organisation alors préparée par le Conseil d'État de Napoléon supprimait la magistrature des directeurs du Jury qui réunissaient, en effet, des pouvoirs énormes. Relativement à la poursuite des délits et à la mise en accusation, le directeur du Jury était en quelque sorte à la fois agent de police judiciaire, procureur du roi, juge d'instruction et Cour royale.</p>
Referenční text	<p>Zákonník z brummairu r. IV. přiznal řediteli soudu tohoto okresu právo přímého vyšetřování přečinu, spáchaného v Gondreville. Všimněte si, že Konvence vyškrtla ze soudní mluvy slovo zločin. Připouštěla pouze přečiny proti zákonu, jež trestala pokutami, vězením, tresty potupnými i tělesnými. Trest smrti pak měl být zaměněn 24 letou nucenou prací v káznici. Konvence uznala, že se čtyřadvacetiletá nucená práce rovná trestu smrti. Co říci o trestním zákonníku, jenž ukládá stálou nucenou práci? Organizace tehdy připravená Státní Radou Napoleonovou zrušila úřad direktorů Soudu, jenž měl opravdu příliš mnoho moci. Vzhledem k vyšetřování zločinu a k podání žaloby, byl ředitel Soudu jaksi zároveň agentem soudní olicie, zástupcem královým, vyšetřujícím soudcem i královským Dvorem.</p>
Překlad Google	<p>Rok Brumaire Code IV vyhradil řediteli poroty oddělení okamžité stíhání trestného činu spáchaného v Gondreville. Mimochodem, všimněte si, že Úmluva vymanila slovo zločin z legálního jazyka. Připustil pouze přestupky proti zákonu, přestupky týkající se pokut, trestu odnětí svobody, hanlivého nebo urážlivého trestu. Trest smrti</p>

	<p>však měl být zrušen v míru a nahrazen čtyřiceti lety nucené práce. Konvent tak odhadl, že dvacet čtyři let nucené práce se rovnalo trestu smrti. A co trestní zákoník, který způsobuje nucenou práci na celý život? Organizace poté připravená Napoleonskou radou zrušila magistraci ředitelů poroty, kteří ve skutečnosti spojili obrovské pravomoci. Pokud jde o stíhání trestných činů a ukládání obvinění, byl ředitel jury zároveň soudním policejním úředníkem, státním zástupcem krále, vyšetřujícím soudcem a královským soudem.</p>
<p>Překlad Bing</p>	<p>Brumaire iV kód vyhrazena pro ředitele poroty oddělení okamžité stíhání trestného činu spáchaného v Gondreville. Všimněte si, mimochodem, že úmluva odstranila slovo trestný čin ze soudního jazyka. Přiznala pouze trestné činy proti zákonu, trestné činy, které vedly k pokutám, trestu odnětí svobody, nechvalně proslulým nebo svěceným trestům. Nicméně sužující trest smrti musel být v míru zrušen a nahrazen čtyřiceti lety nucené práce. Konvent tak odhadl, že pětadvacet let nucené práce se rovná trestu smrti. A co trestní zákoník, který ukládá nucenou práci na věky? Organizace, kterou pak připravila Napoleonova státní rada, odstranila soudnictví ředitelů poroty, kteří získali obrovské pravomoci. Pokud jde o stíhání trestných činů a obžalobu, ředitel poroty byl nějakým způsobem soudním policejním důstojníkem, královským prokurátorem, vyšetřujícím soudcem a královským soudem.</p>

Příloha č. 11

Text 10	
<p>Zdrojový text</p>	<p>Avant de partir, il s'était entendu naturellement avec le préfet, qui, pris par la nuit, ne put se servir du télégraphe. On expédia sur Paris une estafette afin de prévenir le ministre de la Police Générale, le Grand-Juge et l'Empereur de ce crime inouï. Lechesneau trouva dans le salon de Gondreville mesdames Marion et Grévin, Violette, le valet de chambre du sénateur, et le juge de paix assisté de son greffier. Déjà des perquisitions avaient été pratiquées dans le château. Le juge de paix, aidé par Grévin, recueillait soigneusement les premiers éléments de l'instruction. Le magistrat</p>

	<p>fut tout d'abord frappé des combinaisons profondes que révélaient et le choix du jour et celui de l'heure. L'heure empêchait de chercher immédiatement des indices et des preuves. Dans cette saison, à cinq heures et demie, moment où Violette avait pu poursuivre les délinquants, il faisait presque nuit ; et, pour les malfaiteurs, la nuit est souvent l'impunité. Choisir un jour de réjouissances où tout le monde irait voir la mascarade d'Arcis, et où le sénateur devait se trouver seul chez lui, n'était-ce pas éviter les témoins ?</p>
<p>Referenční text</p>	<p>Než se vydal na cestu, dohodl se s prefektem, jenž ale, poněvadž byla noc, nemohl užítí telegrafu.</p> <p>Byl vyslán do Paříže zvláštní posel, aby hrozný zločin oznámil policejnímu ministru, nejvyššímu soudci a Císaři. V Gondreville našel Lechesneau v salóne paní Marionovou a Grévinovou, Violetta, sluhu senátorova, smírčího soudce s drábem. V zámku provedena prohlídka. Smírčí soudce s Grévinem sebrali pečlivě prvky nutné k vyšetřování. Soudce byl překvapen zvláštní shodou okolností, již vykazovala volba dne i hodiny. Tu chvíli nebylo možno hledati známek a důkazů. Tu dobu o půl šesté, kterou chvíli Violetta mohl stíhati pachatele, je už skoro tma – a pro zločince bývá noc často beztrestnou. Vybrati den všeobecného veselí, kdy budou všichni na maškarách v Arcisu a senáter bude sám doma, to bylo patrně proto, aby nebylo svědků.</p>
<p>Překlad Google</p>	<p>Před odjezdem přirozeně vycházel s prefektem, který, když byl v noci, nemohl telegrafovat. Do Paříže byl vyslán kurýr, který varoval ministra obecné policie, velkého soudce a císaře tohoto neuvěřitelného zločinu. Lechesneau našel v salonu Gondreville dámy Marion a Grevin, Violette, komoru senátora a spravedlnost za míru, kterou mu pomáhal jeho úředník. V zámku už byly provedeny prohlídky. Spravedlnost míru za pomoci Grévina pečlivě shromáždila první prvky instrukce. Magistrát byl poprvé zasažen hlubokými kombinacemi odhalenými výběrem dne a hodiny. Čas zabránil okamžitému hledání stop a důkazů. V této sezóně, v půl páté, když Violette dokázala stíhat delikventy, byla téměř noc; a pro zločince je noc často beztrestnost. Rozhodnout se pro den oslavy, kdy by každý šel navštívit arcisovu maškarádu a kde měl být senátor sám doma, tomu nebránilo svědkům?</p>

Překlad Bing	<p>Před odjezdem se přirozeně dohodl s prefektem, který v noci nemohl telegraf použít. Estafette byla poslána do Paříže, aby varovala ministra generální policie, velkého soudce a císaře tohoto neslýchaného zločinu. Lechesneau našel v gondreville obývacím pokoji dámy Marion a Grévin, Violette, senátorův komorník, a smírčí soudce za pomoci jeho úředníka. Na hradě již byly provedeny prohlídky. Smírčí soudce za pomoci Grévina pečlivě shromáždil první prvky procesu. Soudce byl nejprve zasažen hlubokými kombinacemi odhalenými volbou dne a výběrem hodiny. Čas zabránil okamžitému hledání stop a důkazů. V této sezóně, v pět a půl hodiny, kdy Violette byla schopna stíhat pachatele, byla téměř tma; a pro zločince je noc často beztrestností. Vybrat si den radosti, kdy by se všichni šli podívat na Arcisovu šarádu, a kde senátor musel být sám doma, nemělo se vyhnout svědkům?</p>
-----------------	---